Empirical evidence on inflation expectations in the new Keynesian Phillips curve[*]

Sophocles Mavroeidis Mikkel Plagborg-Møller James H. Stock
University of Oxford   Harvard University   Harvard University

February 23, 2013

Abstract

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 used 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 data set and a flexible empirical approach, we find that researchers are faced with substantial model uncertainty, as different combinations of various a priori reasonable specification choices give rise to a vast set of point estimates. Moreover, conditional on a model, 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. Finally, we offer advice on how to overcome the identification problem and reach an empirical consensus.

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 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.[1] Samuelson and Solow (1960) extended what they called "Phillips' curve" to U.S. data and to price inflation.[2] The subsequent study of

---

[1]Mankiw (2001, p. C46) offers a prescient quote from the work of Hume (1752).

[2]Irving Fisher demonstrated a strong correlation between U.S. inflation and unemployment already in 1926, three decades before the seminal paper by Phillips, cf. Fisher (1973). Fisher interpreted the causality as running from prices to real economic activity, however.

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, known as 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. The policy relevance of this research agenda is further highlighted in two speeches by Federal reserve chairman Bernanke (2007, 2008). 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 generalized method of moments (GMM). The literature, which dates back to the seminal papers by Roberts (1995), Fuhrer and Moore (1995) and Galí and Gertler (1999), 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 micro data to study price dynamics. These papers are outside the scope of our review, which focuses only on studies that have used macro data. Interested readers are referred to the review by Nakamura and Steinsson (2013).

Despite the apparent early empirical success of the NKPC, the literature which 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 data set. 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 et al. (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 what previous individual papers have considered (in fact, we claim to estimate 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 non-trivial consequences for the model, as established by Preston (2005). Due to space constraints, we do not discuss recent movements away from the NKPC, such as state dependent pricing (Dotsey et al., 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, minimum distance and VAR-based maximum likelihood. These methods consist of estimators and confidence sets for the structural parameters, as well as specification tests for overidentifying restrictions and the stability of the parameters over time. We outline the empirical strategy of using survey data on expectations, originally implemented by Roberts (1995). Furthermore, we propose a new idea for identification using data revisions as 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. Weak instrument issues provide a unifying explanation of the sensitivity of NKPC estimates and the puzzling disagreement of 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 100 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 non-rationality 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 data sets, instrument choices, specifications, estimators and attention to the weak identification problem, no consensus has been

reached on 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 data set 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 data set. 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, conditional on a particular model, the sampling uncertainty is large. We show this by computing weak identification robust confidence sets for a range of benchmark specifications. The only type of specification that appears to be typically better identified is one that 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 pursue substantially new types of data sets, as well as estimation approaches that are tailored to handle the identification problem.

2 Economic Theory, Specifications and Policy Implications

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. The origins of the NKPC can be traced back to the late seventies, in the work of Fischer (1977) and Taylor (1979).

2.1 Economic foundations of the model

A simple derivation of the NKPC can be obtained as follows.[3] The basic ingredients 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. All variables,

---

[3]For detailed derivations, see Woodford (2003, ch. 3) or Galí (2008, ch. 3).

except prices and inflation, are expressed in percent deviations from a zero-inflation steady state. Prices are expressed in logs and inflation in percentage points. 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 θ 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:

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

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 β denote the common subjective discount factor. The optimal reset price can, in a first-order log-linearization, be expressed as a mark-up over a weighted average of current and expected future marginal costs:

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

where $mc_{i,t+j,t}^n$ 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,

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

Substituting (3) into (2) and rearranging yields

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

where $\kappa = \frac{1-\alpha}{1-\alpha+\alpha\varepsilon} \leq 1$, inflation is given by $\pi_t = p_t - p_{t-1}$, and $mc_t = mc_t^n - p_t$ denotes aggregate real marginal costs. Inserting (4) into (1), we find

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

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 (RE), 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

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

where we have, importantly, used the law of iterated expectations, $E_t(E_{t+1}) = E_t$. 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:

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

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

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

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

Non-rational expectations While we did not need to impose assumptions on firms' individual beliefs to arrive at the 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. This point has been made by Preston (2005), who concluded that, under general non-rational expectation formation, inflation must *irreducibly* depend on long-horizon average expectations of inflation and marginal cost, as in equation (5). In other words, if firms' expectations are not based on the same information set or if they are not perfectly model-consistent, the difference equation (7) is in general inconsistent with the standard New Keynesian modeling framework.[4] As we discuss in section 3.2, this has implications for empirical tests of the NKPC that use survey forecasts to proxy for the expectation term.

---

[4]See also Angeletos and La'O (2009). Adam and Padula (2011) derive a difference equation NKPC under a set of assumptions on the possibly non-rational cross-sectional expectation operator. Their "Condition 1" essentially invokes the law of iterated expectations at the aggregate level. As Preston (2005) argues, this condition will rarely hold unless expectations are, in fact, fully rational.

2.2   Extensions

It was recognized early on that the purely forward-looking NKPC (8) has difficulty fitting aggregate US 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 θ of firms that are unable to re-optimize their prices in the Calvo model index them to past inflation, see Christiano et al. (2005) and Sbordone (2005, 2006).[5]

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

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

where $\gamma\left(L\right) = 1 - \gamma_1 L - \gamma_2 L^2 - \cdots - \gamma_l L^l$ 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 γ_1 is often denoted γ_b. Equation (9) nests the pure NKPC (8) with $\gamma\left(L\right) = 1$ 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 et al. (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 $\gamma\left(L\right)$ 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., $\gamma_f = 1 - \left(\gamma_1 + \gamma_2 + \cdots + \gamma_l\right) = \gamma\left(1\right)$.

The parameters of equation (9) are often referred to as 'reduced-form' or 'semi-structural' 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 $\gamma_f = \theta/\left(\theta + \omega\right)$ and $\gamma_1 = 1 - \gamma_f$, where ω 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 semi-structural ones and thus have nontrivial implications for inference.

Trend inflation   The derivation of equation (8) follows from log-linearizing firms' optimizing conditions around a zero-inflation steady state. Allowing for non-zero steady state inflation, often

---

[5]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.

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., $\bar{\pi}_t = \lim_{T \to \infty} E_t(\pi_{t+T})$. With non-zero 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, p. 2105) show that if non-resetting firms' prices are indexed to a mixture of past inflation π_{t-1} (weight ρ) and current trend inflation $\bar{\pi}_t$ (weight $1 - \rho$), then

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

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.[6] 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 non-reset 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.

2.3   Policy implications

Ideally, estimation uncertainty in the NKPC parameters would only translate into limited ambiguity about our understanding of 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 25 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 γ_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 γ_f and λ as being variation free from the remaining structural parameters (which are calibrated as in Galí, 2008, ch. 3.4) and vary them over a set of values that is consistent with the spread of estimates reported in the literature, cf. section 4 below, namely $\gamma_f = 0.3, 0.4, \ldots, 0.8$ 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 λ, lower values of γ_f imply more sluggish adjustment and more hump-shaped dynamics in inflation. The disparity between the high-γ_f and low-γ_f dynamics increases the lower is λ. For $\lambda = 0.03$ (the solid 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 $\gamma_f = 0.8$. This sensitivity of key economic measures,

---

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

such as the welfare costs of disinflation policy, to the NKPC parameters extends beyond the simple model considered here, as demonstrated by Fuhrer (1997), Mankiw (2001) and Estrella and Fuhrer (2002).[7] If indeed the NKPC is a good approximation to actual price setting, it is therefore highly desirable from a policy perspective to obtain precise estimates of the NKPC coefficients.[8]

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.[9] For ease of exposition, we focus in this section on estimation of 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 (GIV).

2. Derive expectations from a particular reduced-form model (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 the Generalized Method of Moments (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.

GIV   This approach was originally proposed for the estimation of rational expectation (RE) models by McCallum (1976). Hansen and Singleton (1982), who studied estimation of Euler equation models, called it *Generalized Instrumental Variable* estimation (GIV). It has been popularized in

---

[7]Adding estimation uncertainty in the non-NKPC equations to the mix will of course generate even greater uncertainty about appropriate impulse responses.

[8]The degree of inflation indexation also influences the relative optimality of policies that flexibly target the price level or inflation rate (Woodford, 2003, ch. 8.2.1).

[9]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 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 π_{t+1}. This yields

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

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

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

Suppose Z_t is a vector of valid instruments, such that

$$E\left[Z_t h_t(\vartheta)\right] = 0 \tag{13}$$

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 RE 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}. There is an infinite number of predetermined variables that 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.

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 vector autoregression (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 π_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

$$Y_t = AY_{t-1} + V_t, \tag{14}$$

where Y_t, V_t are $nl \times 1$ and A is $nl \times nl$. Equation (14) implies that $E_t(\pi_{t+1}) = \zeta' Y_t$, where ζ is the row of A that corresponds to the inflation equation in the VAR. Substituting this in the NKPC (8), yields the moment conditions $E\left[(\pi_t - \beta Y_t' \zeta - \lambda x_t) Y_{t-1}\right] = 0$, which determine the structural parameters

(β, λ) given ζ, and ζ is identified by the reduced-form equation $E\left[\left(\pi_t - Y'_{t-1}\zeta\right)Y_{t-1}\right] = 0$. The VAR assumption implies that all the relevant instruments are given by Y_{t-1}, and the model can be estimated by GMM using the $2nl$ moment conditions

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

$$\widetilde{h}_t\left(\vartheta, \zeta\right) = \left(\pi_t - \beta Y'_t\zeta - \lambda x_t \; \vdots \; \pi_t - Y'_{t-1}\zeta\right)'. \tag{16}$$

Here ζ_0 is the true value of ζ. 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.[10] VAR-GMM, VAR-ML and VAR-MD are numerically identical if the model is just-identified.

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.

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 or the Federal Reserve's "Greenbooks", are used as proxies for inflation expectations in the NKPC. Let $\pi^s_{t+j|t}$ denote the j-step-ahead survey forecast of inflation at time t. The most common implementation substitutes the one-quarter-ahead forecast $\pi^s_{t+1|t}$ for $E_t\left(\pi_{t+1}\right)$ in the NKPC (8) to get

$$\pi_t = \beta \pi^s_{t+1|t} + \lambda x_t + u_t + \beta \varepsilon_t, \quad \text{where} \tag{17}$$

$$\varepsilon_t \equiv E_t\left(\pi_{t+1}\right) - \pi^s_{t+1|t}. \tag{18}$$

The survey error ε_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 ε_t as well as on the correlation between $\pi^s_{t+1|t}$ and the cost-push shock u_t. Some authors treat $\pi^s_{t+1|t}$ as exogenous, but we argue below that the assumptions underlying this are too strong. Alternatively, one may use $\pi^s_{t+1|t-1}$, which is certainly predetermined, instead of $\pi^s_{t+1|t}$ that 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.

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

---

[10]Guerrieri et al. (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.

$E_t\left(\pi_{t+1}\right)$ with $\phi E_t(\pi_{t+1}) + (1-\phi)\,\pi_{t+1}^s$ in the NKPC (8).[11] The parameter ϕ is not identified if survey expectations are rational.

Because survey *estimation* of the NKPC obviates the need for modeling inflation expectations, some authors interpret the procedure as allowing for non-rational 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 non-rational expectations formation (see Fuhrer, 2012).

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., data sets of different vintages, one possible identification 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 likely do not affect current inflation if we control for the latest-vintage data. More precisely, past-vintage variables 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 strong-instrument 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 traditional strong instrument asymptotic approximations. Strong instrument asymptotics imply that GMM estimators of ϑ are consistent and asymptotically normal under certain regularity conditions, see Newey and McFadden (1994). In section 3.3 below we will argue that the strong instrument approximation is not empirically relevant in the present context, as it cannot 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 $E_{t-1}\left[h_t\left(\vartheta_0\right)\right] = 0$, for which the theory of optimal instruments of Chamberlain (1987) provides a useful efficiency bound. $E_{t-1}\left[h_t\left(\vartheta_0\right)\right] = 0$ implies that all predetermined variables Y_{t-j}, $j \geq 1$ are admissible instruments. The optimal choice of instruments is described in section A.2.1 of the Appendix.

---

[11]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.

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}. Because their derivation requires modeling the conditional mean and variance of the data, none of the papers in this literature has attempted to use optimal instruments.[12]   Therefore, we cannot rank GIV estimators reviewed in this paper in terms of efficiency. Indeed, claims about their relative efficiency are not formally justified.

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 $\tilde{h}_t(\vartheta_0, \zeta_0) = (u_t, v_{\pi t})'$ 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 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 VAR-ML. To evaluate the likelihood, they use an algorithm by Anderson and Moore (1985) that finds a RE solution for any given NKPC parameters. For certain parameter combinations, there may be multiple stable RE solutions, a situation known as indeterminacy (see, e.g., Lubik and Schorfheide, 2004), and so the likelihood is not uniquely determined by the NKPC parameters. The Fuhrer-Moore approach to VAR-ML deals with the non-uniqueness by restricting the parameter space to the region in which indeterminacy does not occur, the so-called determinacy region. Kurmann (2007) argues that the assumption of determinacy is extraneous to the model, and he provides a technique for evaluating the likelihood without imposing determinacy. When the parameters fall in the indeterminacy region, his method effectively lets the data choose amongst the different possible VAR solutions. He demonstrates that this approach yields a very different estimate on U.S. data than the Fuhrer-Moore approach. See section A.2.3 in the Appendix for more discussion of this issue.

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 (LIML) and two stage least squares (2SLS), respectively, in the linear IV regression model (Fukač and Pagan, 2010). Based on results for the IV model, it is plausible that VAR-ML has better higher-order asymptotic properties than VAR-MD under strong-instrument asymptotics.

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 MA components. When the MA roots are large (i.e., nearly noninvertible), no finite-order VAR can produce an accurate representation of the dynamics. Infinite-order VARs may also arise for other reasons, e.g., by omitting relevant variables from a finite-order VAR.[13]

---

[12]Fuhrer and Olivei (2004) 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.

[13]See also Fernández-Villaverde et al. (2007). But note that it is possible that omission of variables from a VAR

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 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 $E\left[Y_{t-1}\left(\pi_t - \beta\pi_{t+1} - \lambda x_t\right)\right] = 0$, in that GIV uses direct projections of future inflation on predetermined variables, while VAR-GMM uses iterated multi-step forecasts from a VAR.

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-step-ahead forecasts. The former substitutes $\pi_{t+1|t}^s$ for $E_t\left(\pi_{t+1}\right)$ in (8), yielding equations (17)–(18). The second possibility is to substitute $\pi_{t+1|t-1}^s$ for $E_t\left(\pi_{t+1}\right)$ in (8) to get

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

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

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 ε_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 ε_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* π_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 $\left[E_{t-1}\left(\pi_{t+1}\right) - \pi_{t+1|t-1}^s\right]$ 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 $E_{t-1}\left(\varepsilon_t\right) = 0$ in (18). Therefore, moment conditions for (17) are the same as for the GIV approach (13) where π_{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.

As argued in section 2.1, the difference equation NKPC (8) is 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) or (19) are only microfounded

does not necessarily cause misspecification (Fukač and Pagan, 2007).

if the inflation forecasts firms rely on are rational and identical across firms. However, a common empirical finding is 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 et al. (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 non-rational 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.

3.3   Weak identification

Weak identification arises when ϑ is almost unidentified, i.e., when the moment conditions are close to being satisfied for all parameters ϑ in a non-vanishing neighborhood of the true value ϑ_0. When identification is weak, conventional strong-instrument asymptotics provide poor approximations to the sampling distribution of GMM estimators and tests, even in large samples. The result is non-normally distributed estimators that are biased toward their OLS (or NLS) counterparts and conventional confidence sets that 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 which 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 completely unidentified because no relevant pre-determined instruments exist. Therefore, the model predicts that identification will become arbitrarily weak as the slope of the NKPC λ 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 completely unidentified. In other words, good economic policy is bad for econometric analysis (cf. Mavroeidis, 2010; Cochrane, 2011).

When identification is weak, we show below that different 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 ϑ around the true value ϑ_0. Recent advances in econometrics have made it possible to do inference that is fully robust to weak identification. These weak identification robust methods have the property that they are always reliable irrespective of the quality of the instruments. Unlike the standard non-robust methods, confidence intervals are only tight when parameters are well-identified. Thorough surveys on the consequences of weak identification and methods of inference that are robust to it include Stock et al. (2002), Dufour (2003) and Andrews and Stock (2005).

Simulation studies Mavroeidis (2005) illustrates the above-mentioned points 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 finite-sample 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 here 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

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

with accompanying VAR(2) reduced-form dynamics

$$x_t = \xi' Y_{t-1} + \varepsilon_{xt} \tag{22}$$

for the forcing variable. Here $Y_t = (\pi_t, x_t, \pi_{t-1}, x_{t-1}, 1)'$, and $\xi = (\xi_{\pi 1}, \xi_{x 1}, \xi_{\pi 2}, \xi_{x 2}, c_x)'$ are the reduced-form coefficients. The GIV estimator that we consider estimates (γ_f, λ, c) by linear GMM, with $\Delta \pi_t$ as the independent variable, $(\pi_{t+1} - \pi_{t-1})$ and x_t (and a constant) as regressors, and Y_{t-1} as instruments.[14] The three VAR-based methods exploit the VAR(2) reduced form for (π_t, x_t) to estimate the same three parameters, as explained in section 3.1.[15] More details on our implementation of the estimators are given in section A.3 of the Appendix.

We start off by comparing the finite-sample performance of the estimators under a weakly identified data generating process (DGP), *DGP 1*, with the performance under strong identification, *DGP 2*. Both DGPs have $\gamma_f = 0.7$, $\lambda = 0.03$ and $c = 0$ as their true structural NKPC parameters. Such parameter values represent typical estimates from the literature. For both DGPs, the true reduced-form dynamics of inflation and the forcing variable follow a VAR(2), so the identifying restrictions for the VAR methods are valid. In *DGP 1*, our choice of reduced-form VAR parameters ξ for the forcing variable are based on OLS estimates on quarterly U.S. data with the labor share as x_t. *DGP 2* has ξ set to empirically unrealistic values that lead to much stronger identification. Details about the specifications, as well as a precise quantification of the strength of identification, are given in section A.3 of the Appendix.

For each DGP, we simulate samples of 200 observations each and calculate point estimates for each of the four estimators.[16] We execute 10,000 Monte Carlo repetitions per DGP. Figures 2–3 display the densities of the sampling distributions of the γ_f estimators under *DGPs 1* and *2*, respectively. Evidently, the four estimators exhibit quite different behaviors in the weakly identified parametrization, *DGP 1*. The GIV estimates of γ_f are biased downward toward the probability limit of the OLS estimator, which is close to 0.5 for all DGPs in this paper.[17] While the sampling distribution density for GIV has rather fat tails, it is single-peaked and not far from bell-shaped.[18]

---

[14]Observe that the NKPC (21) may equivalently be written $\Delta \pi_t = c + \gamma_f E_t(\pi_{t+1} - \pi_{t-1}) + \lambda x_t + u_t$.

[15]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.

[16]Because the specifications are overidentified, the VAR estimators do not coincide.

[17]By "OLS" we mean a simple regression of $\Delta \pi_t$ on $(\pi_{t+1} - \pi_{t-1})$ and x_t.

[18]Simulations with smaller true values of λ yield a stronger bias of γ_f estimates toward 0.5.

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 clearly give spurious results. In the strongly identified parametrization, *DGP 2*, the situation is entirely different. Here the sampling densities 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 tightly concentrated around the true value $\gamma_f = 0.7$ than the GIV density.

We now wish to illustrate the effect of misspecification of the VAR assumption. We do this by constructing a modified version of *DGP 1*, in which the reduced-form error in the inflation equation $\pi_t = \zeta' Y_{t-1} + v_{\pi t}$ follows an MA(1) process $v_{\pi t} = \varepsilon_{\pi t} + b' \varepsilon_{t-1}$, where $\varepsilon_t = (\varepsilon_{\pi t}, \varepsilon_{xt})'$ is i.i.d., and the MA parameter $b \in \mathbb{R}^2$ is non-zero. This VARMA reduced form is consistent with the NKPC (21) because the choices of $(\gamma_f, \lambda) = (0.7, 0.03)$ and ξ in *DGP 1* lead to indeterminacy. Details on the DGP, which we label *DGP 3*, are given in section A.3 of the Appendix. Figure 4 depicts the sampling distribution of the four γ_f estimators under *DGP 3*. Because *DGP 3* is much more strongly identified than *DGP 1*, the sampling densities all look quite bell-shaped. However, the central tendency of the bell curves for the three VAR methods has been shifted toward $\gamma_f \approx 1.1$ due to the presence of the MA root, which none of the VAR methods is robust to.[19] Because GIV is robust to indeterminacy, it performs very well in this case.

Section A.3 in the Appendix displays additional results for analogs of *DGPs 1* and *2* for which we set $\gamma_f = 0.3$. The qualitative conclusions are unchanged. In particular, the spurious clustering of VAR estimates around $\gamma_f = 1$ in weakly identified settings appears to be robust to the choice of true value of γ_f. This is interesting, as Sbordone (2005) and Kurmann (2007) both report rather large VAR-based estimates of γ_f, and our empirical VAR-GMM estimates in section 5 similarly concentrate around 1.

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. The NKPC (21) can then be written

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

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 $\Delta \pi_t$.[20] 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 γ_f), all VAR estimators coincide, and it is easy to show that the VAR estimator of δ is $1/\hat{\rho}_1$. The implied γ_f estimator is then

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

---

[19]Note that this bias is a different phenomenon from the bimodality in Figure 2, since *DGP 3* is strongly identified.
[20]Our weakly identified *DGP 1* nearly has this reduced form, as the reduced-form coefficient on π_{t-1} in the inflation equation is approximately 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 γ_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 γ_f in the regression of $\Delta\pi_t$ on $(\pi_{t+1} - \pi_{t-1})$, which equals $1/2$.[21] We are thus able to explain both biases observed in Figure 2.

Another way to view the weak identification VAR bias is as follows. The VAR estimator of δ 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 $\Delta\pi_t = \bar{u}_t$.[22] This is trivially an AR(1) with coefficient 0. Note, however, that if the DGP for inflation really has this form, γ_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 γ_f close to 1, even though the parameter is completely 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 properly 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 maximum likelihood analysis are currently under development, cf. Andrews and Mikusheva (2011) and references therein.

3.4 Other issues

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

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 4 lags of 6 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 et al. (2008). It is well known that use of many instruments biases 2SLS estimators towards OLS. To get some

---

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

[22]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. On a side note, the fact that $\gamma_f < 1/2$ is necessary for determinacy in this model illustrates the Kurmann (2007) critique of the Fuhrer and Moore (1995) approach to VAR-ML: If one imposes the extra determinacy assumption, the parameter space is cut in half.

intuition, think about the limit case where the number of instruments is the same as the sample size. Then 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. Recent contributions by Andrews and Stock (2007) and Newey and Windmeijer (2009) demonstrate some robustness properties of the weak identification robust methods to many weak instruments. However, a key requirement is that the instruments are sufficiently informative to render the model well-identified asymptotically. These results do not apply when there are many instruments that may be arbitrarily weak. These results also ignore the complications arising from uncertainty in the estimation of variance of the moment conditions, which can be substantial.[23] Therefore, we recommend against the use of many instruments in the estimation of the NKPC.

Unit roots   Empirically, U.S. inflation appears close to non-stationary in certain subsamples. Fanelli (2008), Nymoen et al. (2010) and Boug et al. (2010) 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. In addition to the econometric concerns, Nymoen et al. (2010) argue that non-stationarity undermines the model's microfoundations, which are based on log-linearizing around a constant steady state.

When the inflation coefficients in the NKPC sum to 1, the model can be written in terms of changes in inflation. If instruments are also 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.[24]

Misspecification   The various methods described so far are not equivalent under 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 in empirical estimates reported by Fuhrer and Moore (1995), Galí and Gertler (1999) and Jondeau and Le Bihan (2005), so Jondeau and Le Bihan argue that this could be the source of disparity in the estimates. However, differences remain when the NKPC is extended to include more lags of inflation. Hence, we believe that the Kurmann (2007) critique provides a better explanation of this disparity. Mavroeidis (2005) explores the implications of omitted dynamics for the bias of GMM estimation of the hybrid NKPC of Galí and Gertler (1999).

Cagliarini et al. (2011) and Imbs et al. (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 et al. (2011) trace this bias to the presence of an additional error term in the aggregate NKPC resulting from sectoral heterogeneity.

Autocorrelated cost-push shocks   Autocorrelation of u_t in the NKPC (8) violates the common identifying assumption $E_{t-1}(u_t) = 0$. With the exception of external instruments, all other

---

[23]Uncertainty arising from the HAC estimator of the variance of the moment conditions rises sharply with the number of instruments, and is another source of problems for GMM tests in finite samples.

[24]Note that the restriction that inflation coefficient sum to 1 is not sufficient for inflation to have a unit root. This depends on the dynamics of the forcing variable.

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 $\tilde{u}_t$ include a future inflation forecast error, so they may exhibit MA(1) autocorrelation even when the structural error u_t is not autocorrelated (Jondeau and Le Bihan, 2005; Mavroeidis, 2005; Eichenbaum and Fisher, 2007). Boug et al. (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 et al. (2009) show by simulation that GMM estimates of the slope of the NKPC are biased downwards and the Hansen (1982) J test has little power against this misspecification in realistic sample sizes.

If u_t is autocorrelated, valid (internal) instruments can only be derived if we impose restrictions on the autocorrelation structure of u_t. If it is of some finite order q, then valid instruments are variables that are predetermined in period $t - q$. If u_t is autoregressive, say $u_t = \rho u_{t-1} + v_t$, where $E_{t-1}(v_t) = 0$, then the appropriate identifying restrictions can be obtained by quasi-differencing: $E_{t-1}[h_t(\vartheta) - \rho h_{t-1}(\vartheta)] = 0$. Specific assumptions about autocorrelation of the error appear ad hoc and may be purely empirically motivated (Smets and Wouters, 2007) unless they are properly microfounded. For example, Erceg and Levin (2003) show that autoregressive errors can be induced by rational learning in an environment of persistent shocks to the inflation target, when agents have less information than the central bank.

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, 2011).

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 separate overall econometric approaches. Figure 5 and Table 1 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 RE. Roberts (1995) first 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 RE forecast error into the equation's error term and instrumenting for next period's inflation and the output gap using lagged variables. Using annual U.S. data, Roberts found a significant role for the output gap.

Seminal, independent 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 fully microfounded RE 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 and Gertler further assume that a fraction of firms employ a rule of thumb when setting prices, thus introducing intrinsic persistence of the inflation rate on top of the extrinsic persistence imparted by the forcing variable. Using linear and non-linear GMM methods, Galí and Gertler (1999) find that, while the backward-looking rule-of-thumb term is significant, the forward-looking RE term dominates; they also obtain a significant and correctly signed coefficient on the labor share (unlike the output gap). In addition, the overidentifying restrictions are not rejected at conventional significance levels.

4.2   Rational expectations GMM (GIV) estimation

Many papers have followed the GMM approach of Galí and Gertler (1999) while extending the basic hybrid NKPC in various ways. Galí et al. (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 et al. (2005), 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; GMM estimation on U.S. data yields intuitively reasonable coefficients with significant forward-looking behavior. Krause et al. (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 et al. (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., that 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 GMM, 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 de-trended output (which is used by Roberts, 1995, and Galí and Gertler, 1999), and GMM estimates

of the slope of the NKPC are even more significant than when using the labor share.[25] Gagnon and Khan (2005) extend the NKPC to a more general CES production function and find that structural GMM estimates imply less price stickiness than under the usual Cobb-Douglas specification. In addition to a CES production function, McAdam and Willman (2010) add varying capacity utilization, which decreases the estimated coefficient on the inflation expectation term. Batini et al. (2005) and Rumler (2007) estimate open-economy NKPCs by GMM on data from European countries, finding a significant role for international variables. Guerrieri et al. (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. Gwin and VanHoose (2008) and Shapiro (2008) construct alternative measures of firm marginal costs from micro-level and sectoral data.

While the previously mentioned papers find a significant, and often dominant, role for the forward-looking RE 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 et al. (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 GMM estimates are quite sensitive to the choice of instrument set and estimator (see also Guay and Pelgrin, 2005), and the Galí et al. (2001) hybrid NKPC is rejected in favor of alternative, encompassing models of inflation. 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.[26] 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 GMM in iterated form (cf. section A.2.2 in the Appendix) yields a smaller coefficient 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 set-up. Galí et al. (2005) offer a partial rebuttal to the criticism.

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 RE GMM framework.

4.3   Minimum distance

Sbordone (2002) tests the pure NKPC using a two-step approach akin to 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 micro-estimates of

---

[25]See also Fukač and Pagan (2010) for a critique of "off-model" output detrending.
[26]This issue is further pursued by Mazumder (2010).

price stickiness in the U.S. Sbordone (2005) refines the estimation approach by interpreting it as minimum distance estimation and accounting for sampling uncertainty of the first-step estimated VAR (see also Kurmann, 2005). She also provides estimates of the *hybrid* NKPC, broadly confirming the conclusions in Galí and Gertler (1999). Sbordone (2006) develops a model of joint price and wage determination which is estimated by minimum distance. Tillmann (2008) uses a related two-step approach to assess the importance of the cost channel of monetary policy. Coenen et al. (2007) develop a model with a general, non-constant 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. The Calvo (1983) assumption of a constant hazard rate is found to be inconsistent with U.S. and German data.

4.4   Maximum likelihood and Bayesian analysis

Another oft-used estimation method is ML. 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 Anderson and Moore (1985) routine to solve for a RE 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 RE component is insignificant. Subsequent papers have applied the AIM-based VAR-ML method to the Galí and Gertler (1999) hybrid NKPC. Fuhrer and Olivei (2004) and Fuhrer (2006) find a small coefficient on the RE term relative to that on lagged inflation. Roberts (2005), who also considers GMM and impulse response matching, estimates the hybrid NKPC with four lags of inflation instead of one. His estimated coefficients on the forward- and backward-looking terms are of the same magnitude. Jondeau and Le Bihan (2003, 2008) seek to explain why GIV estimates of the coefficient on forward-looking expectations tend to be high, while VAR-ML estimates tend to be lower, as established empirically by Jondeau and Le Bihan (2005) on data from the U.S. and major European countries. They find that finite-sample bias is unlikely to be the only cause and demonstrate that measurement error and misspecification of the inflation lag structure can account for the disparity. Kiley (2007) considers a slightly different NKPC specification with four lags of inflation and expectations of next-period inflation taken with respect to *last*-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 arguably extraneous assumption that, for any given parameter vector, the RE solution must be unique. He shows how to write the likelihood in a way that does not require uniqueness. When applying the method to U.S. data, he finds evidence of a large share of forward-looking behavior, which contrasts with estimates obtained under the additional uniqueness assumption. Korenok et al. (2010) also eschew the AIM algorithm and instead write their model in a form that

is amenable to Kalman filtering.[27]

As emphasized by Bårdsen et al. (2004) and Nymoen et al. (2010), nonstationarity of inflation and/or the labor share may spell trouble for the standard econometric tests of the NKPC. Fanelli (2008) and Boug et al. (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 RE 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 MLE 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 pseudo-posteriors 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 et al. (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.

4.5   Survey expectations

As mentioned, Roberts (1995) uses survey measures of inflation expectations from the Michigan and Livingston surveys as an alternative to the RE GIV approach.[28] Roberts (1997) finds that the apparent sluggishness and non-rationality 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) 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 U.S. and Canada using survey forecasts and allowing for non-zero trend inflation. Gerberding (2001), Paloviita and Mayes (2005), Paloviita (2006, 2008), Henzel and Wollmershäuser (2008) and Koop and Onorante (2011) estimate NKPCs for European countries using various measures of survey expectations of inflation and various estimation procedures. These papers find that the

---

[27]Other papers that use filtering techniques include Nelson and Lee (2007), Kim and Kim (2008) and Kim and Manopimoke (2011).

[28]Roberts still instruments for the output gap as it may be correlated with the inflation error term.

coefficient on the expectation term tends to be smaller than or of the same magnitude as the coefficient on lagged inflation. 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 et al. (2008, 2009) consider both SPF, Greenbook and Michigan survey forecasts. Unlike the RE 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 and SPF forecasts in an NKPC. The GMM estimates point to non-rational 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 calculate the RE term based on a reduced-form VAR and estimate the NKPC by Bayesian and ML methods. Smith (2009) gives conditions under which it is advantageous to pool survey forecasts and rational expectations 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 full understanding of the interplay between non-rational expectation formation and price setting remains elusive.

4.6   Weak identification

The literature's awareness of the problems associated with weak identification has grown over time. Galí et al. (2001) guide the choice of their GMM instrument set by the first-stage F statistic. 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. Mavroeidis (2004, 2005) provides analytical and simulation evidence that explains why weak identification is likely to be an issue for NKPC estimation. He demonstrates the unreliability of traditional GMM point estimates, Wald confidence regions and overidentification tests, and strongly recommends using robust inference procedures. Dufour et al. (2006) compute Anderson and Rubin (1949) and Kleibergen (2002) confidence sets for the Canadian and U.S. NKPCs, both under RE and using survey forecasts as a proxy for expectations. The U.S. RE confidence region is fairly large, yet not completely uninformative, while the survey forecast one is empty; the opposite is true for Canada, but the non-empty survey forecast confidence region only contains implausible parameter values. Nason and Smith (2008a) reject the hybrid NKPC for both Canada, the U.K. and the U.S. using the Anderson and Rubin (1949) test and the Guggenberger and Smith (2008) GEL test.[29] In contrast, Martins and Gabriel (2009) find very wide robust GEL confidence sets. Kleibergen and Mavroeidis (2009) survey the literature on weak identification in the NKPC and give new evidence on its extent using a variety of GMM-based robust tests. While inflation appears to be significantly forward-looking, the confidence regions are wide. Formal parameter stability tests give mixed results about the presence of structural breaks. Motivated by the Lucas critique, Magnusson and Mavroeidis (2011) suggest using robust parameter instability

---

[29]The Anderson-Rubin test is a special case of the S test (cf. section A.2.5 in the Appendix) in a linear model with conditionally homoskedastic and serially uncorrelated errors.

tests to improve inference about the NKPC. Magnusson and Mavroeidis (2010) develop a weak identification robust version of Sbordone's (2005) minimum distance test, finding somewhat smaller confidence regions than using the GIV approach. Kleibergen and Mavroeidis (2011) show the consequences of ignoring weak identification in Bayesian analyses of the NKPC and propose ways of circumventing the problems. Dufour et al. (2010a,b) carry out robust inference on certain theoretical extensions of the NKPC with real wage rigidities or labor market frictions.

Some papers have devised methods for improving the strength of the instruments in GMM estimation of the NKPC. Fuhrer and Olivei (2004) propose employing a reduced-form VAR to compute expectations for next period's inflation and the output gap, and then using the computed expectations as instruments. Their estimate of the coefficient on the RE term is much smaller than the traditional GMM estimate. Dees et al. (2009) obtain instruments for individual-country NKPCs by estimating a multi-country cointegrating VAR. Building on Beyer et al. (2008), Kapetanios and Marcellino (2010) and Kapetanios et al. (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.

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, if not impossible.

5 Empirical Synthesis

In this section we generate estimates of the NKPC corresponding to a wide selection of modeling approaches from the literature.[30] Because we use a common data set 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 conditional on a model, the *sampling* uncertainty of the estimates tends to be large. Both these conclusions can be explained by the weakness of identification. We also demonstrate that the potential for the data to distinguish between rational and non-rational 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 non-farm business sector. See section A.4 in the Appendix for a detailed description of the data and transformations.

---

[30]Estimation results are obtained using Ox, see Doornik (2007).

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

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 UC-SV model of Stock and Watson (2007, 2010). For CPI inflation, 10-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.[31]

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 secular trends such as the recent dramatic decline in the labor share, which may arguably be attributed to demographic changes outside of the New Keynesian model.[32] 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.[33] Another stationary analog of the labor share is the cointegrating relationship between real wages and labor productivity found by Sbordone (2005, fn. 19). 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 et al. (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 RE assumption. As discussed in section 4, GIV analyses typically find point estimates of the coefficient on expectations γ_f in the 0.5–0.7 range, and the coefficient on lagged inflation γ_b, the measure of intrinsic persistence, is often significantly positive and not significantly different from $1-\gamma_f$. The coefficient on the labor share λ is generally estimated to be positive but borderline significant (using the usual strong-instrument inference). In Table 2 we replicate these findings using data of the same vintage as Galí and Gertler (1999) but with the Galí et al. (2001) instrument set.[34] Later papers have mostly obtained insignificant λ estimates, and like Rudd and

---

[31]We are grateful to Shawn Sprague for assisting us in obtaining the real-time labor share data.

[32]See Gwin and VanHoose (2008) for a discussion of the need to detrend marginal cost measures.

[33]Unfortunately, we cannot use our real-time labor share data to construct actual real-time labor share gaps. Because the 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.

[34]We obtained the 1998 vintage data from Adrian Pagan. We use 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 et al. (2001).

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 λ, 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 5, estimates of λ and γ_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), generally find forward-looking behavior to be dominant ($\gamma_f > 0.5$). The results presented in Figure 5 are a tiny subset of possible specifications. Table 3 presents various dimensions of the specification choice that have been considered in the literature.[35] These combinations of choices produce a very large number specifications that are not objectionable on a priori grounds.

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 (λ, γ_f) in the baseline NKPC (9) for various combinations of the specification choices listed in Table 3. We then plot the point estimates in (γ_f, λ)-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 λ 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.[36]

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 6 and 7 report the results for the labor share and output gap as forcing variable, respectively. Figure 6 also contains the Galí and Gertler (1999) vintage point estimate and associated Wald confidence ellipse from Table 2 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 5. Table 4 reports summary statistics for the point estimates in Figures 6 and 7.

The main messages from the figures are that (i) estimates of the coefficient of 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%) 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 λ and γ_f both appear statistically significantly positive using (one-sided) 5%-level individual t-tests is quite small, while most of the reported estimates in the literature appear to fall in that category. It is interesting that the frequency of significantly positive coefficients is almost double for the output gap specifications than the labor share ones.

---

[35]The only components of the table that have not been explored in the literature are some of the real-time data series (but see Paloviita and Mayes, 2005, and Dufour et al., 2006) and the use of survey expectations as instruments (but see Nunes, 2010). The latter is motivated by evidence that surveys typically forecast inflation better, see Ang et al. (2007).

[36]https://sites.google.com/site/sophoclesmavroeidis/research/working-papers/online-supplement-for-nkpc-review

This is not in line with the view that NKPC specifications with the output gap as forcing variable are more counterfactual than ones 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 which are robust to weak identification.

To shed some light on the issue of weak identification, the penultimate row of Table 4 reports the median value of the heteroskedasticity and autocorrelation robust first-stage F statistics of Montiel Olea and Pflueger (2011), denoted F_{HAR}. A low value of this statistic can be thought of as a warning sign for weak instruments.[37] 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).[38] Even though this is not a formal test of weak instruments, and we do not recommend the use of pre-tests in place of 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 8 displays smoothed density estimates of our first-stage F statistics for forecasting the expectations proxy, treating rational expectations separately from time-t dated SPF/GB forecasts, and using all instrument sets in Table 3. 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 and GB 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 RE counterparts, and the evidence reported in section 5.3 corroborates this conjecture.

The final row of Table 4 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% at the 5% level, so there is no systematic evidence 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.

CUE versus 2-step GMM We generate GMM estimates using both the efficient 2-step estimator (2S) and the continuous updating estimator (CUE) of Hansen et al. (1996). These are described in section A.2.1 in the Appendix. Table 5 compares summary statistics of point estimates based on 2S and CUE GMM for the various specifications listed in Table 3.[39] The 2S and CUE are very

---

[37]The well-known rule of thumb of F > 10 is a useful benchmark (Stock and Yogo, 2002), although Montiel Olea and Pflueger (2011) show that this condition is neither necessary nor sufficient for instruments to be strong in the presence of heteroskedasticity and autocorrelation.

[38]Low values of the F statistic for forecasting the forcing variable do arise in specifications that use the real-time (RT) instrument set. For labor share specifications, the median value of the F statistic is 9.5 for the RT regressions, whereas the median is 63.1 for all other instrument sets in Table 3. For output gap specifications, the corresponding medians are 69.3 and 170.6, respectively.

[39]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

similar for λ, and CUE is typically larger than 2S for γ_f. Moreover, the 2S 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 2S than for CUE (Stock et al., 2002). The final two rows give median differences for specifications using the Galí and Gertler (1999) instrument set (GG), which is considerably larger than the rest. The proximity to OLS is stronger for these specifications, but less so for the CUE. The relatively better bias properties of the CUE come at the cost of greater dispersion, which is confirmed by the 90% interquantile ranges: the ones for the CUE are more than double the corresponding ones for 2S. Bårdsen et al. (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.

VAR assumption   Our VAR-GMM estimates are based on the moment condition (15). The reduced form evolution of inflation and the forcing variable is thus restricted to be a linear function of the variables in the instrument set. Table 6 reports summary statistics comparing GIV and VAR-GMM estimates, while Figure 9 plots clouds for estimates that impose the VAR assumption and those that do not. The VAR assumption makes a big difference for the estimate of γ_f: it tends to move the estimate closer to 1. This is consistent with the results reported in Sbordone (2002, 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), which we believe to be due to the Kurmann (2007) critique. There is no noticeable difference in the estimates of λ between the VAR and GIV methods, but there is a substantial difference in γ_f. In the vast majority of cases (about 80%), imposing the VAR assumption increases γ_f, and the median difference is about 0.18 in both the labor share and output gap specifications. The median estimate of γ_f is actually 1 to two decimal points, which implies that about half of the time the VAR assumption produces estimates of $\gamma_f > 1$. As we pointed out in section 3.3, this could be a symptom of weak identification, which biases VAR estimates of γ_f towards 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 γ_f estimates to concentrate even more tightly around 1, but also increases the dispersion of the λ estimates.

Survey forecasts   There are large and systematic differences in the effect of using survey inflation forecasts relative to RE (based on GIV) across labor share and output gap specifications, sample period, inflation series (GDP deflator versus CPI) and forecast source (SPF versus GB). The results are documented in Table 7 and Figure 10. Survey forecasts typically increase the estimate of λ across most specifications and sample periods, especially when the output gap is used as forcing variable. The estimate of γ_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 10, which plots post-1984 cloud for RE 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.

Subsample variation is also quite striking. The reduction in γ_f relative to GIV is much more evident in the post 1981-sample (SPF CPI forecasts are only available from 1981q3). There is also a

fat tails (Hansen et al., 1996).

systematic difference between CPI and GDP deflator measures of inflation. CPI inflation typically yields much larger estimates of γ_f than GDP deflator inflation. SPF and GB forecasts do not yield systematically different full-sample estimates, though there are some systematic difference to the estimates of γ_f before and after 1984. Treating surveys as endogenous or exogenous does not seem to make much difference to the central tendency of the estimates, though it does make a difference to dispersion (the latter estimates are a lot less dispersed, as expected).

Instruments   The last two rows of Table 5 showed that the Galí and Gertler (1999) instrument set (GG) produces estimates for both λ and γ_f that are typically lower than average, and closer to the OLS estimates. The GG estimates are also less dispersed and more concentrated around the OLS estimates. This highlights the pitfalls of using many possibly weak instruments (Andrews and Stock, 2007). This was realized early on in the literature, and later papers typically use a smaller number of instruments. Other than GG, the choice of instrument set does not substantially change the central tendency of the estimates. The first two rows of Table 8 list the median pairwise difference in estimates across specifications that differ only in their choice of instrument set.

Number of inflation lags in the NKPC   Table 8 also explores the implications of adding lags of inflation in the NKPC. Estimates of λ seem to be unaffected, on average, but estimates of γ_f are very sensitive to the number of inflation lags included in the model. Specifically, adding lags of inflation to the NKPC tends to reduce the estimate of γ_f by about 0.25 when we add 1 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). This need not be due to misspecification of the more restrictive NKPC specifications, as was suggested by Rudd and Whelan (2005) and Mavroeidis (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 γ_f would exhibit a larger bias towards $\gamma_f = 1/2$.[40] Indeed, the first-stage F_{HAR} statistics 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.[41] 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 0 lags to 4 lags.

Inflation series   Figure 10 indicates that survey forecast specifications are more sensitive to the choice of inflation series than RE estimates are. Figure 11 compares GDP deflator and CPI estimates, pooling across all RE and survey forecast specifications and all subsamples in Table 7. Estimates of both parameters are considerably more dispersed in GDP deflator specifications, but the median difference across these inflation series is very small, cf. Table 8. This is partly a result of the general 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 10 compares CPI versus GDP

---

[40]As explained in section 3.3, the probability limit of the OLS γ_f estimator is often close to 0.5 for empirically realistic NKPC parametrizations.

[41]The corresponding numbers for output gap specifications are 25.2, 4.1 and 3.1.

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 γ_f irrespective of whether the labor share or output gap is used as the forcing variable, cf. Table 8. For λ, there is a small positive difference only in output gap specifications. The reason why these differences are not large is most probably 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 RE, PCE estimates are similar to CPI results, although in output gap specifications λ 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.

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 λ, 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.[42] 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.

Sample   Table 8 shows little systematic difference in the central tendency of estimates before and after 1984. As Figure 10 suggests, survey estimates are, however, sensitive to sample choice. This is consistent with Zhang et al. (2008). Figure 12 reports pre- and post-1984 estimates in the survey specifications with GDP deflator inflation. For RE specifications, the central tendency of estimates does not depend much on the choice of sample, but post-1984 estimates are more tightly concentrated around $\lambda = 0$.

Other specification choices   The restriction that coefficients sum to 1 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, see the last two rows of Table 8. This is consistent with Chowdhury et al. (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 sets for selected specifications based on methods that are robust to weak

---

[42]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.

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. 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.

Since S sets are not available analytically, we compute them by grid search over the parameter space, focusing only on the range used for the cloud plots (which includes over 90% of all point estimates). For each specification, we evaluate the test on over 1000 grid points. Because this procedure is computationally intensive, we consider only a subset of all the specifications listed in Table 3, consisting of about 1400 specifications, see Table 9.[43] The cloud of point estimates for the specifications in Table 9 is qualitatively similar to that for the full set of specifications in Table 3. Perhaps not surprisingly, the union of the joint 90% S sets for all specifications in Table 9 covers the entire parameter region in our plots, $(\gamma_f, \lambda) \in [-1, 2] \times [-0.3, 0.3]$.[44] 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 90% and 95% S sets across different specification choices (see section A.5 in the Appendix for details). The S sets are generally quite large on average, covering between 1/3 (90% level) and 1/2 (95% level) of the parameter space for both labor share and output gap specifications.[45] 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 0 to 4 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 RE to survey inflation expectations, and different instruments. S sets for RE are on average much larger than for surveys, as anticipated in the discussion around Figure 8 above, and it looks like most of the difference arises from using GB data. With regards to the different instruments, RT (external) instruments yield the largest S sets covering 50-80% 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 surveys as instruments produces on average smaller S sets than using lags of inflation.

It is interesting to compute how often the S sets for (λ, γ_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 λ and γ_f occur in a very small fraction of

---

[43]Computation of the S sets for VAR-GMM takes about 100 times more than using the other single-equation methods. Therefore, we only consider 16 specifications that impose the VAR assumption.

[44]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.

[45]Additionally, the S sets are, on average, between 3 to 7 times larger than the corresponding Wald ellipses.

33

the specifications considered (less than 5% at the 10% level).[46] 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 RE 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% S sets and Wald confidence ellipses (based on the CUE) for (λ, γ_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 36). Figure 13 reports the results for specifications based on GDP deflator inflation using either 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, pre-1984 and post-1984 subsamples. The confidence sets are not completely uninformative, and they are particularly tight along the λ axis over the full sample, but rather wide across the γ_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.[47]

Figure 14 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 γ_f is larger than 1 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 13. Thus, the VAR assumption appears to be informative, confirming the results in Magnusson and Mavroeidis (2010).

Figure 15 reports confidence sets based on survey specifications. Results are reported for SPF and GB GDP deflator inflation forecasts, as well as SPF CPI inflation forecasts. For SPF GDP deflator forecasts, the results are quite similar to the corresponding RE specifications, given in the bottom row of Figure 13 (post-1984 sample). However, when we use the GB forecasts, confidence sets become considerably smaller. S sets 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% Wald ellipses do not overlap, which, if identification were strong, would suggest time-variation in the coefficients of the NKPC, as suggested by Zhang et al. (2008). However, the S sets do overlap considerably over the two subsamples, so it is not clear whether the survey-based NKPC is unstable.

To assess the empirical success of the external instruments approach, we plot robust confidence sets for GDP deflator inflation using the real-time (RT) instrument set in Figure 16. These figures

---

[46]Specifically, this is the fraction of the S sets that are non-empty and lie entirely in the area $(\lambda, \gamma_f) \in (0, .3] \times (0, 2]$.

[47]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 (2011).

are comparable to the top row in Figure 13, although the sample starts in 1971 due to data availability. Figure 16 demonstrates the unfortunate fact that the most plausibly exogenous instrument set also results in very weak identification, as the 90% robust confidence sets contain all reasonable γ_f values. Post-1984 confidence sets (not reported) are even larger.

5.4 Nesting RE 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). Figure 17 reports CUE estimates and 90% S sets for the coefficients of future inflation (RE) and time-t dated one-quarter-ahead forecasts of inflation in the model

$$\pi_t = \lambda x_t + \gamma_{RE}\pi_{t+1} + \gamma_s \pi_{t+1|t}^s + \gamma_b \pi_{t-1} + u_t. \tag{25}$$

The coefficient λ here is treated as well-identified, and it is concentrated out.[48] We consider both SPF and GB 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 3) plus two lags of survey inflation forecasts. The point estimates generally indicate a dominant role for RE, consistent with the evidence in Nunes (2010), and different from the preferred estimates 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 RE term is dominant. Interestingly, all the confidence sets exclude $\gamma_{RE} = \gamma_s = 0$.[49]

6 Conclusion

Based on the foregoing comparison of more than 100 papers from the literature with our analysis of thousands of a priori reasonable new Keynesian Phillips curve 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, conditional on 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.

---

[48]This is a reasonable assumption since in all cases the first-stage F_{HAR} statistic for x_t is over 100.

[49]A number of authors have argued that non-rational expectations can account for the intrinsic persistence in the NKPC (Roberts, 1997; Brissimis and Magginas, 2008; Nunes, 2010). The results for the case with $\gamma_b = 0$ (not reported) are quite similar to the ones reported.

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 vector autoregression (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 non-rational 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, 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 data sets 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.

A 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):

$$\pi_t = \gamma_f E_t \pi_{t+1} + (1 - \gamma_f)\pi_{t-1} + \lambda mc_t,$$

$$x_t = -\frac{1}{\sigma}(i_t - E_t \pi_{t+1} - \rho) + E_t x_{t+1}, \quad mc_t = \left(\sigma + \frac{\varphi + \alpha}{1 - \alpha}\right) x_t, \qquad (26)$$

$$i_t = \rho + \phi_\pi \pi_t + \phi_x x_t + v_t, \quad v_t = \rho_v v_{t-1} + \varepsilon_t^v.$$

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 follows an AR(1) process. We call ε_t^v the monetary policy shock.

When calibrating the structural parameters of the model, we use the benchmark values in Galí (2008, p. 52). 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 $\varphi = 1$, the labor exponent in the production function is $1 - \alpha = 2/3$, the Taylor rule coefficients are $(\phi_\pi, \phi_x) = (1.5, 0.5/4)$, 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

$$S_T\left(\vartheta, \overline{\vartheta}\right) = f_T\left(\vartheta\right)' W_T\left(\overline{\vartheta}\right) f_T\left(\vartheta\right), \qquad (27)$$

where $\overline{\vartheta}$ is some preliminary estimator of ϑ, and W_T is a weighting matrix that may depend on the data and on ϑ. A GMM estimator is the minimizer of $S_T\left(\vartheta, \overline{\vartheta}\right)$ with respect to ϑ, if it exists. Given the particular choice of moments $f_T(\vartheta)$, efficient GMM estimation requires $W_T\left(\overline{\vartheta}\right)$ 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 $\overline{\vartheta}$ is obtained using some weight matrix that does not depend on ϑ. When the moment conditions are linear, $\overline{\vartheta}$ may be obtained using two-stage least squares.

Setting $\overline{\vartheta} = \vartheta$, so the efficient weight matrix estimator $W_T(\vartheta)$ is evaluated at the same parameters as the sample moments $f_T(\vartheta)$, yields the continuous updated estimator (CUE), which was proposed by Hansen et al. (1996). 2-step GMM and CUE are asymptotically equivalent under strong identification, but the latter has certain advantages under weak identification (see, e.g., Stock et al., 2002).

Strong identification For the model to be identified, the moment condition $E\left[f_T\left(\vartheta\right)\right] = 0$ must hold if and only if ϑ is equal to the true value ϑ_0. This condition is not sufficient for consistent estimation of ϑ. For this, a stronger identification condition is required. Specifically, let $f_0\left(\vartheta\right) = \lim_{T\to\infty} E\left[f_T\left(\vartheta\right)\right]$, where $f_T\left(\vartheta\right) = T^{-1}\sum_{t=1}^{T} Z_t h_t\left(\vartheta\right)$ are the sample moments. Strong identification requires that $f_0\left(\vartheta\right) = 0$ if and only if $\vartheta = \vartheta_0$. To better capture the consequences of instruments being nearly irrelevant, the weak instrument literature instead uses asymptotic approximations under which the strong identification condition does not hold.

Optimal instruments When identification is given by conditional moment restrictions of the form $E_{t-1}\left[h_t\left(\vartheta_0\right)\right] = 0$, where $h_t\left(\cdot\right)$ 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 $E\left[Z_t h_t\left(\vartheta_0\right)\right] = 0$. 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 $Z_t \in \mathcal{I}_{t-1}$, where $\mathcal{I}_{t-1}$ denotes the information set at time $t - 1$. If the residual function $h_t\left(\vartheta_0\right)$ is a martingale difference sequence (MDS), the optimal instruments are given by

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

see Chamberlain (1987).

When estimating the NKPC by VAR-GMM, the two-dimensional residual vector (16) satisfies the conditional moment restriction $E_{t-1}\left[\widetilde{h}_t\left(\vartheta_0\right)\right] = 0$. It is a MDS because it is adapted to the information set at time t. Moreover, the VAR assumption implies that $E_{t-1}\left[\partial\widetilde{h}_t\left(\vartheta_0\right)/\partial\vartheta'\right]$ 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 $\mathcal{I}_t$ since $h_t\left(\vartheta_0\right) = \tilde{u}_t \in \mathcal{I}_{t+1}$, see equation (12). Under the assumption $E_{t-1}\left(u_t\right) = 0$, $h_t\left(\vartheta_0\right)$ can be represented as a moving average of order 1, e.g., $h_t\left(\vartheta_0\right) = \upsilon_t - \varphi\upsilon_{t+1} = \varphi\left(L^{-1}\right)\upsilon_t$, say, where υ_t is an MDS with $E_{t-1}\left(\upsilon_t\right) = 0$. Following Hayashi and Sims (1983), the optimal instruments can be obtained as follows. First, forward-filter $h_t\left(\vartheta_0\right)$ to get $\upsilon_t = \varphi\left(L^{-1}\right)^{-1} h_t\left(\vartheta_0\right)$. Then compute the optimal instruments Z_t^o for $E_{t-1}\left(\upsilon_t\right) = 0$ using (28). Finally, transform these instruments to the optimal instruments $\tilde{Z}_t^o$ for $E_{t-1}\left[h_t\left(\vartheta_0\right)\right] = 0$, which are given by $\tilde{Z}_t^o = \varphi\left(L\right)^{-1} Z_t^o = \sum_{j=0}^{\infty} \varphi^j Z_{t-j}^o$.

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 $E_t\left(u_{t+j}\right) = 0$, $j > 0$, and the law of iterated expectations, we get

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

38

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

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

where Y_{t-1} is a vector of lags of π_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

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

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

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, 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 minimum distance estimator is defined as the minimizer of the objective function

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

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 $g\left(\hat{A}, \vartheta\right)$.

Define e_π and e_x to be the unit vectors with 1 in the position of π_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

$$g_1(A, \vartheta)' \equiv e_\pi' A - \beta e_\pi' A^2 - \lambda e_x' A = 0. \tag{32}$$

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

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

where $\zeta = A' e_\pi$ is the coefficients of the projection of π_{t+1} on Y_t. These are exactly the moment conditions for VAR-GMM given in (15).[50]

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

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

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

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

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

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

The distance function $g_1\left(A, \vartheta\right)$ defined in (32) satisfies $g_1\left(A, \vartheta\right) = \left(I - \beta A\right)' g_2\left(A, \vartheta\right)$. Because $\left(I - \beta A\right)$ 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.[51]

VAR-ML Suppose

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

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 a $n \times 1$ 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 implied by the structural NKPC (8).

To fix ideas, suppose $z_t = (\pi_t, x_t)'$. Given the structural parameters ϑ and the reduced-form coefficients $\zeta = A' e_\pi$ in the VAR equation for π_t, we can use the NKPC and VAR equations to solve for the reduced-form coefficients in the VAR equation for x_t (explicit calculations for the case

---

[50]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 $E[(Y_t - A Y_{t-1}) Y'_{t-1}] = 0$ for the VAR companion matrix A hold exactly.

[51]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 $g_1\left(\hat{A}, \vartheta\right) = 0$ and $g_2\left(\hat{A}, \vartheta\right) = 0$ can be solved for ϑ as a function of $\hat{A}$, in which case the VAR-MD estimator does not depend on the choice of distance function.

$l = 1$ are given below). This defines a mapping from (ϑ, ζ) to VAR matrices A_j. Under a normality assumption on the VAR errors v_t in (36), we can then write up the likelihood as a function of (ϑ, ζ) using standard formulas.

In the general case, we define an auxiliary set of parameters ψ such that the VAR coefficients can be expressed as functions of (ϑ, ψ), say $A_j(\vartheta, \psi)$, $j = 1, ..., l$. If the VAR errors are Gaussian and homoskedastic, i.e., $v_t \sim N(0, \Omega)$, where Ω is a $l \times l$ positive definite variance matrix, as is most often assumed in the literature, then after concentrating with respect to Ω, the log-likelihood function can be written as

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

where $\hat{\Omega}(\vartheta, \psi) = T^{-1} \sum \hat{v}_t(\vartheta, \psi) \hat{v}_t(\vartheta, \psi)'$ and $\hat{v}_t(\vartheta, \psi) = z_t - \sum_{j=1}^{l} A_j(\vartheta, \psi) z_{t-j}$.

There are two different implementation of VAR-ML in the literature, which are not equivalent. Fuhrer and Moore (1995) and all the papers that use their algorithm (Fuhrer, 1997; Fuhrer and Rudebusch, 2004; Fuhrer and Olivei, 2004; Fuhrer, 2006, 2012) pick ψ as the coefficients of all reduced-form equations in the VAR except the one corresponding to inflation. Then they use the AIM algorithm of Anderson and Moore (1985) to solve the resulting system of rational expectations equations consisting of the NKPC and the remaining reduced form equations to obtain $A_j(\vartheta, \psi)$. The challenge with this approach is that the solution of this system is non-unique for certain parameter combinations. To deal with that problem, Fuhrer and Moore (1995) restrict the parameter space so as to ensure existence and uniqueness of a stable VAR solution.[52]

Kurmann (2007) gave a compelling critique of this approach by demonstrating that it can lead to severe biases when the true values of the parameters lie outside the restricted parameter space. He also argued that ruling out non-uniqueness (also known as "indeterminacy") does not have any economic justification.[53] To address the problem of non-uniqueness, Kurmann (2007) proposed an alternative approach to writing the likelihood. Instead of picking ψ to be the coefficients in the reduced-form equations of all variables other than inflation, let ψ be the coefficients in the reduced-form equations of all variables other than the forcing variable x_t. It turns out that (except on a set of measure zero) there is then a *unique* map from (ϑ, ψ) to the reduced-form coefficients for x_t, so the approach can search over the entire parameter space. An illustration of Kurmann's method is given below.

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 ϑ and ψ. VAR-MD instead sets ψ equal to its OLS estimator and only minimizes this distance with respect to ϑ. Thus, the relationship between VAR-MD and VAR-ML is analogous to the relationship between 2SLS and 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. Moreover, computation of VAR-MD (like 2SLS) is easier than VAR-ML (like LIML). Another difference is that VAR-ML is invariant to nonlinear

---

[52]A stable solution is one for which the eigenvalues of the VAR companion matrix $A(\vartheta, \psi)$ lie inside the unit circle.

[53]Moreover, even if one wished to assume that the equilibrium of the underlying structural model of the economy is determinate, this does not imply that a system of equations where some of the structural equations have been replaced by their reduced form will have a determinate solution.

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 et al. (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.

Example of Kurmann's VAR-ML   As an illustration of Kurmann's method, consider the structural NKPC (8). Suppose the reduced form is a VAR(1) in π_t, x_t,

$$\pi_t = a_{\pi\pi}\pi_{t-1} + a_{\pi x}x_{t-1} + v_{\pi t}$$
$$x_t = a_{x\pi}\pi_{t-1} + a_{xx}x_{t-1} + v_{xt},$$

and let $\psi = (a_{\pi\pi}, a_{\pi x})'$ and $\vartheta = (\beta, \lambda)'$. We want to determine $a_{x\pi}, a_{xx}$ as a function of (ϑ, ψ). Substituting for $E_t(\pi_{t+1}) = a_{\pi\pi}\pi_t + a_{\pi x}x_t$ in (8) and re-arranging yields

$$(1 - \beta a_{\pi\pi})\pi_t = (\lambda + \beta a_{\pi x})x_t + u_t.$$

If $\lambda + \beta a_{\pi x} \neq 0$, then this equation uniquely pins down the reduced-form coefficients for x_t,[54]

$$a_{x\pi} = \frac{(1 - \beta a_{\pi\pi})a_{\pi\pi}}{\lambda + \beta a_{\pi x}}, \qquad a_{xx} = \frac{(1 - \beta a_{\pi\pi})a_{\pi x}}{\lambda + \beta a_{\pi x}}$$

Hence, given structural parameters ϑ and the reduced-form coefficients for inflation ψ, we can evaluate the concentrated log likelihood (37), which may then be numerically maximized to determine the MLE. In contrast, the model-implied mapping from the parameters $(\vartheta, a_{x\pi}, a_{xx})$ to the reduced-form parameters for inflation $(a_{\pi\pi}, a_{\pi x})$ is not unique, which is why Fuhrer and Moore (1995) must restrict their parameter space.

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 eq. (9) is part of Y_{t-1}):

$$\pi_t = \lambda x_t + \gamma_f \pi_{t+1}^e + \delta' Y_{t-1} + u_t. \tag{38}$$

We use π_{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. Re-arrange (38):

$$\pi_t = \lambda x_t + \gamma_f \pi_{t+1} + \delta' Y_{t-1} + \tilde{u}_t, \quad \tilde{u}_t = u_t + \gamma_f \left(\pi_{t+1}^e - \pi_{t+1}\right).$$

---

[54]Notice that these are proportional to ψ, so the VAR matrix A is of reduced rank. If $\lambda + \beta a_{\pi x} = 0$, then $(1 - \beta a_{\pi\pi})a_{\pi\pi} = (1 - \beta a_{\pi\pi})a_{\pi x} = 0$, so $a_{x\pi}$ and a_{xx} are ill-defined and the method breaks down.

Suppose that

$$E\left(u_t|Y_{t-1}, \{Y_{t-1}^r, \text{ all } r\}\right) = 0. \tag{39}$$

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

$$E\left(\tilde{u}_t|Y_{t-1}, Y_{t-1}^r\right) = 0. \tag{40}$$

If $\gamma_f = 0$, then $\tilde{u}_t = u_t$ and data revisions are exogenous by (39). 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 π_{t+1}^* denote the rational expectation of π_{t+1}, and suppose that

$$\pi_{t+1}^e = \pi_{t+1}^* + \zeta_t, \quad \text{where } E\left(\zeta_t|Y_{t-1}^{t-1}\right) = 0, \quad \text{and} \tag{41}$$

$$E\left(\pi_{t+1} - \pi_{t+1}^*|Y_{t-1}^{t-1}\right) = 0. \tag{42}$$

Condition (41) holds under rational expectations, in which case $\zeta_t = 0$, but it also holds under departures from rational expectations, in which case ζ_t is some "opinion" that is orthogonal to observable vintage-$(t-1)$ data. Condition (42) says that the information set used to compute π_{t+1}^* contains Y_{t-1}^{t-1}. Under these conditions, it can be shown that $E\left(\tilde{u}_t|Y_{t-1}^{t-1}\right) = 0$. In other words, Y_{t-1}^{t-1} is an exogenous instrument in the model (38). But this treats Y_{t-1} as endogenous, which leaves the model underidentified (there are two more endogenous variables, π_{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 (41) and (42) 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 ϑ_0, i.e., $TS_T\left(\vartheta_0, \vartheta_0\right)$. Under H_0 and some regularity conditions, this statistic is asymptotically $\chi^2\left(k\right)$ with degrees of freedom equal to the number of moment restrictions (or instruments), irrespective of whether the model is identified or not. A $(1 - a)\%$ level S set is obtained by collecting all points ϑ_0 for which $S_T\left(\vartheta_0, \vartheta_0\right)$ does not exceed the $(1 - a)$ percentile of $\chi^2\left(k\right)$. When the model also contains exogenous and predetermined variables, e.g., w_t and π_{t-j} in (9), their coefficients are concentrated out in order to improve the power of the test, see Stock and Wright (2000, Theorem 3).

A.2.6   Weak identification robust Hansen test

The minimum value of the S statistic coincides with Hansen's J test of overidentifying restrictions that is based on the continuous updated GMM objective function, see Hansen et al. (1996). Its strong-instruments asymptotic distribution under the null of correct specification is the usual $\chi^2\left(k - p\right)$, where k is the number of identifying restrictions and p is the total number of estimated

43

parameters. Under weak instruments, the asymptotic distribution of this statistic is bounded by $\chi^2\left(k-q\right)$, 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 $\chi^2\left(k-q\right)$. 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, and we provide some additional results. Table 10 lists the parameter choices for *DGPs 1–3* from the main text, as well as *DGPs 4* and *5* below. All DGPs set the constants in the reduced-form VARMA equal to 0. The innovations $\varepsilon_t = (\varepsilon_{\pi t}, \varepsilon_{xt})'$ 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 the table 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 1, 3* and *4*, the reduced-form coefficients ξ 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 1–3* are indeterminate, since none of the Blanchard and Kahn (1980) generalized eigenvalues are outside the unit circle. *DGP 3* is a variant of *DGP 1* in which the MA parameter b is non-zero. We have set

$$\pi_t = \frac{1}{\gamma_f}\pi_{t-1} - \frac{1-\gamma_f}{\gamma_f}\pi_{t-2} - \frac{\lambda}{\gamma_f}x_{t-1} + v_{\pi t},$$

where $v_{\pi t} = \varepsilon_{\pi t} + b'\varepsilon_{t-1}$. It is easy to verify that this inflation process satisfies the structural NKPC (21) with $u_t = -\gamma_f b'\varepsilon_t$. We choose $b = (-1.4, -0.3)'$ to obtain stark results. Note that *DGP 3* is much more strongly identified than *DGP 1*.

DGPs 4 and *5* set $\gamma_f = 0.3$ and are determinate. They have $\lambda = -0.03$ because we were unable to find any DGPs with $\gamma_f = 0.3$ and reasonable reduced-form parameters ξ such that both $\lambda > 0$ and a stable RE solution exists. The probability limit of the OLS estimator of γ_f is again close to 0.5. Figures 18–19 display the sampling distributions of the γ_f estimators under these two DGPs. We run 10,000 Monte Carlo repetitions per DGP. The results are similar to those in Figures 2–3. The slight bimodality of VAR-MD under *DGP 5* disappears if the strength of identification is increased further.

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 γ_f. VAR-MD uses moment conditions 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) reverse-engineering approach 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.

44

A comprehensive documentation of our approach is available in the online supplement (cf. footnote 36). This document also provides a more comprehensive set of results, including additional DGPs, the behavior of λ estimators, and rejection frequencies for various t tests, overidentification tests and the S test.

A.4   Data description

Most series mentioned in Table 3 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 36).

Wage and commodity price inflation, which we use as instruments, refer to the growth in business sector compensation per hour and commodity PPI inflation from the Bureau of Labor Statistics (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 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 λ) of 1,600 for output and 10,000 for the labor share. Our computation of *real-time* (one-sided) output gaps proceeds as 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.[55]

The real-time labor share data set is gleaned from the BLS's Productivity and Cost "Preliminary" news releases on nominal unit labor costs and the implicit price deflator. Since most of the data corresponds approximately to what was known around the middle of each quarter, the spreadsheet has the same general format as the real-time dataset available for other macroeconomic series on the website of the Philadelphia Fed. 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

Figure 20 displays point estimates and the union of the joint 90% S sets for all specifications listed in Table 9.

Table 11 reports the average size of 90% and 95% S sets as a fraction of the overall parameter space $(\gamma_f, \lambda) \in [-2, 1] \times [-0.3, 0.3]$ for various specification choices and samples. Here 'all' refers

---

[55]For the business sector log labor share, the AIC selects 15 lags on the full sample.

to the different options listed in Table 9. The 'end of sample' varies by series, for instance, for GB data the sample ends in 2005q4, while for all other series it stretches to the end of 2011.

Table 12 reports some additional statistics associated with the S sets corresponding to the specifications of Table 9. Row "% 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 non-empty and positive S sets for all specifications and for various subcategories, e.g., 'all RE' includes all GIV and VAR-GMM specifications.

References

Adam, K. and Padula, M. (2011), 'Inflation dynamics and subjective expectations in the United States', *Economic Inquiry* **49**(1), 13–25.

Adjemian, S., Bastani, H., Juillard, M., Mihoubi, F., Perendia, G., Ratto, M. and Villemot, S. (2011), 'Dynare: Reference Manual, Version 4', *Dynare Working Papers* (1).

An, S. and Schorfheide, F. (2007), 'Bayesian Analysis of DSGE Models', *Econometric Reviews* **26**(2–4), 113–172.

Anderson, G. and Moore, G. (1985), 'A linear algebraic procedure for solving linear perfect foresight models', *Economics Letters* **17**(3), 247–252.

Anderson, T. W. and Rubin, H. (1949), 'Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations', *The Annals of Mathematical Statistics* **20**(1), 46–63.

Andrews, D. W. K. (1993), 'Tests for Parameter Instability and Structural Change With Unknown Change Point', *Econometrica* **61**(4), 821–856.

Andrews, D. W. K. and Ploberger, W. (1994), 'Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative', *Econometrica* **62**(6), 1383–1414.

Andrews, D. W. and Stock, J. H. (2005), 'Inference with Weak Instruments', *National Bureau of Economic Research Technical Working Paper Series* (313).

Andrews, D. W. and Stock, J. H. (2007), 'Testing with many weak instruments', *Journal of Econometrics* **138**(1), 24–46.

Andrews, I. and Mikusheva, A. (2011), 'Maximum Likelihood Inference in Weakly Identified Models'. Working paper, December version.

Ang, A., Bekaert, G. and Wei, M. (2007), 'Do macro variables, asset markets, or surveys forecast inflation better?', *Journal of Monetary Economics* **54**(4), 1163–1212.

Angeletos, G.-M. and La'O, J. (2009), 'Incomplete information, higher-order beliefs and price inertia', *Journal of Monetary Economics* **56**(Supplement), S19–S37.

Ascari, G. (2004), 'Staggered Prices and Trend Inflation: Some Nuisances', *Review of Economic Dynamics* **7**(3), 642–667.

Atkeson, A. and Ohanian, L. E. (2001), 'Are Phillips Curves Useful for Forecasting Inflation?', *Federal Reserve Bank of Minneapolis Quarterly Review* **25**(1), 2–11.

Bårdsen, G., Jansen, E. S. and Nymoen, R. (2004), 'Econometric Evaluation of the New Keynesian Phillips Curve', *Oxford Bulletin of Economics and Statistics* **66**, 671–686.

Barnes, M. L., Gumbau-Brisa, F., Lie, D. and Olivei, G. P. (2011), 'Estimation of Forward-Looking Relationships in Closed Form: An Application to the New Keynesian Phillips Curve', *Federal Reserve Bank of Boston Working Papers* (11-3). June version.

Batini, N., Jackson, B. and Nickell, S. (2005), 'An open-economy new Keynesian Phillips curve for the U.K.', *Journal of Monetary Economics* **52**(6), 1061–1071.

Benigno, P. and López-Salido, J. D. (2006), 'Inflation Persistence and Optimal Monetary Policy in the Euro Area', *Journal of Money, Credit and Banking* **38**(3), 587–614.

Bernanke, B. S. (2007), 'Inflation Expectations and Inflation Forecasting'. Speech at the Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute.

Bernanke, B. S. (2008), 'Outstanding Issues in the Analysis of Inflation'. Speech at the Federal Reserve Bank of Boston's 53rd Annual Economic Conference.

Beyer, A., Farmer, R. E. A., Henry, J. and Marcellino, M. (2008), 'Factor analysis in a model with rational expectations', *Econometrics Journal* **11**(2), 271–286.

Blanchard, O. and Galí, J. (2007), 'Real Wage Rigidities and the New Keynesian Model', *Journal of Money, Credit and Banking* **39**, 35–65.

Blanchard, O. and Galí, J. (2010), 'Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment', *American Economic Journal: Macroeconomics* **2**(2), 1–30.

Blanchard, O. J. and Kahn, C. M. (1980), 'The Solution of Linear Difference Models under Rational Expectations', *Econometrica* **48**(5), 1305–1311.

Boug, P., Cappelen, d. and Swensen, A. R. (2010), 'The new Keynesian Phillips curve revisited', *Journal of Economic Dynamics and Control* **34**(5), 858–874.

Brissimis, S. N. and Magginas, N. S. (2008), 'Inflation Forecasts and the New Keynesian Phillips Curve', *International Journal of Central Banking* **4**(2), 1–22.

Cagliarini, A., Robinson, T. and Tran, A. (2011), 'Reconciling microeconomic and macroeconomic estimates of price stickiness', *Journal of Macroeconomics* **33**(1), 102–120.

Calvo, G. A. (1983), 'Staggered prices in a utility-maximizing framework', *Journal of Monetary Economics* **12**(3), 383–398.

Campbell, J. and Shiller, R. (1988), 'The dividend-price ratio and expectations of future dividends and discount factors', *Review of Financial Studies* **1**(3), 195–228.

Campbell, J. Y. and Shiller, R. J. (1987), 'Cointegration and Tests of Present Value Models', *Journal of Political Economy* **95**(5), 1062–1088.

Caner, M. (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.

Carvalho, C. (2006), 'Heterogeneity in price stickiness and the real effects of monetary shocks', *Frontiers in Macroeconomics* **6**(1).

Chamberlain, G. (1987), 'Asymptotic efficiency in estimation with conditional moment restrictions', *Journal of Econometrics* **34**(3), 305–334.

Chowdhury, I., Hoffmann, M. and Schabert, A. (2006), 'Inflation dynamics and the cost channel of monetary transmission', *European Economic Review* **50**(4), 995–1016.

Christiano, L. J., Eichenbaum, M. and Evans, C. L. (2005), 'Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy', *Journal of Political Economy* **113**(1), 1–45.

Cochrane, J. (2011), 'Determinacy and identification with taylor rules', *Journal of Political Economy* **119**(3), 565–615.

Coenen, G., Levin, A. T. and Christoffel, K. (2007), 'Identifying the influences of nominal and real rigidities in aggregate price-setting behavior', *Journal of Monetary Economics* **54**(8), 2439–2466.

Cogley, T. and Sbordone, A. M. (2008), 'Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve', *American Economic Review* **98**(5), 2101–26.

Dees, S., Pesaran, M. H., Smith, L. V. and Smith, R. P. (2009), 'Identification of New Keynesian Phillips Curves from a Global Perspective', *Journal of Money, Credit and Banking* **41**(7), 1481–1502.

Doornik, J. A. (2007), *Object-Oriented Matrix Programming Using Ox*, third edn, Timberlake Consultants Press, London.

Dotsey, M., King, R. G. and Wolman, A. L. (1999), 'State-Dependent Pricing And The General Equilibrium Dynamics Of Money And Output', *The Quarterly Journal of Economics* **114**(2), 655–690.

Dufour, J.-M. (2003), 'Identification, weak instruments, and statistical inference in econometrics', *Canadian Journal of Economics* **36**(4), 767–808.

Dufour, J.-M., Khalaf, L. and Kichian, M. (2006), 'Inflation dynamics and the New Keynesian Phillips Curve: An identification robust econometric analysis', *Journal of Economic Dynamics and Control* **30**, 1707–1727.

Dufour, J.-M., Khalaf, L. and Kichian, M. (2010*a*), 'Estimation uncertainty in structural inflation models with real wage rigidities', *Computational Statistics and Data Analysis* **54**(11), 2554–2561.

Dufour, J.-M., Khalaf, L. and Kichian, M. (2010*b*), 'On the precision of Calvo parameter estimates in structural NKPC models', *Journal of Economic Dynamics and Control* **34**(9), 1582–1595.

Eichenbaum, M. and Fisher, J. D. (2004), 'Evaluating the Calvo Model of Sticky Prices', *National Bureau of Economic Research Working Paper Series* (10617).

Eichenbaum, M. and Fisher, J. D. (2007), 'Estimating the frequency of price re-optimization in Calvo-style models', *Journal of Monetary Economics* **54**(7), 2032–2047.

Erceg, C. J. and Levin, A. T. (2003), 'Imperfect credibility and inflation persistence', *Journal of Monetary Economics* **50**(4), 915–944.

Estrella, A. and Fuhrer, J. C. (2002), 'Dynamic Inconsistencies: Counterfactual Implications of a Class of Rational-Expectations Models', *The American Economic Review* **92**(4), 1013–1028.

Fanelli, L. (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, J., Rubio-Ramírez, J. F., Sargent, T. J. and Watson, M. W. (2007), 'ABCs (and Ds) of Understanding VARs', *The American Economic Review* **97**(3), 1021–1026.

Fischer, S. (1977), 'Wage indexation and macroeconomic stability', *Carnegie-Rochester Conference Series on Public Policy* **5**, 107–147.

Fisher, I. (1973), 'I Discovered the Phillips Curve: "A Statistical Relation between Unemployment and Price Changes"', *The Journal of Political Economy* **81**(2), 496–502.

Friedman, M. (1968), 'The Role of Monetary Policy', *The American Economic Review* **58**(1), 1–17.

Fuhrer, J. (1997), 'The (Un)Importance of Forward-Looking Behavior in Price Specifications', *Journal of Money, Credit, and Banking* **29**(3), 338–350.

Fuhrer, J. (2006), 'Intrinsic and Inherited Inflation Persistence', *International Journal of Central Banking* **2**(3), 49–86.

Fuhrer, J. (2012), 'The Role of Expectations in Inflation Dynamics', *International Journal of Central Banking* **8**, 137–165.

Fuhrer, J. C. and Olivei, G. (2004), 'Estimating Forward Looking Euler Equations with Gmm Estimators: An Optimal Instuments Approach', *FRB of Boston Research Paper Series* (04-2).

Fuhrer, J. C. and Olivei, G. P. (2010), 'The Role of Expectations and Output in the Inflation Process: An Empirical Assessment', *Federal Reserve Bank of Boston Public Policy Briefs* (10-2).

Fuhrer, J. C. and Rudebusch, G. D. (2004), 'Estimating the Euler equation for output', *Journal of Monetary Economics* **51**(6), 1133–1153.

Fuhrer, J. and Moore, G. (1995), 'Inflation Persistence', *The Quarterly Journal of Economics* **110**(1), 127–159.

Fukač, M. and Pagan, A. R. (2007), 'Commentary on "An estimated DSGE model for the United Kingdom"', *Federal Reserve Bank of St. Louis Review* **89**(4), 233–240.

Fukač, M. and Pagan, A. R. (2010), 'Limited information estimation and evaluation of DSGE models', *Journal of Applied Econometrics* **25**(1), 55–70.

Gagnon, E. and Khan, H. (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, J. (2008), *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*, Princeton University Press.

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

Galí, J., Gertler, M. and López-Salido, J. (2001), 'European inflation dynamics', *European Economic Review* **45**(7), 1237–1270.

Galí, J., Gertler, M. and López-Salido, J. D. (2005), 'Robustness of the estimates of the hybrid New Keynesian Phillips curve', *Journal of Monetary Economics* **52**(6), 1107–1118.

Gerberding, C. (2001), 'The information content of survey data on expected price developments for monetary policy', *Deutsche Bundesbank discussion paper* (9/01).

Gordon, R. J. (1990), U.S. Inflation, Labor's Share, and the Natural Rate of Unemployment, *in* H. Konig, ed., 'Economics of Wage Determination', New York: Springer Verlag, pp. 1–34.

Gordon, R. J. (2011), 'The History of the Phillips Curve: Consensus and Bifurcation', *Economica* **78**(309), 10–50.

Guay, A. and Pelgrin, F. (2005), 'The U.S. New Keynesian Phillips Curve: An Empirical Assessment'. Working paper, September 7 version.

Guerrieri, L., Gust, C. and López-Salido, J. D. (2010), 'International Competition and Inflation: A New Keynesian Perspective', *American Economic Journal: Macroeconomics* **2**(4), 247–280.

Guggenberger, P. and Smith, R. J. (2008), 'Generalized empirical likelihood tests in time series models with potential identification failure', *Journal of Econometrics* **142**(1), 134–161.

Gumbau-Brisa, F., Lie, D. and Olivei, G. P. (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 Papers* (11-4). June version.

Gwin, C. R. and VanHoose, D. D. (2008), 'Alternative measures of marginal cost and inflation in estimations of new Keynesian inflation dynamics', *Journal of Macroeconomics* **30**(3), 928–940.

Hansen, C., Hausman, J. and Newey, W. (2008), 'Estimation with many instrumental variables', *Journal of Business & Economic Statistics* **26**, 398–422.

Hansen, L. P. (1982), 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica* **50**(4), 1029–1054.

Hansen, L. P., Heaton, J. and Yaron, A. (1996), 'Finite-Sample Properties of Some Alternative GMM Estimators', *Journal of Business & Economic Statistics* **14**(3), 262–80.

Hansen, L. P. and Singleton, K. J. (1982), 'Generalized instrumental variables estimation of nonlinear rational expectations models', *Econometrica* **50**(5), 1269–1286.

Hayashi, F. and Sims, C. A. (1983), 'Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments', *Econometrica* **51**(3), 783–798.

Henry, S. G. B. and Pagan, A. R. (2004), 'The Econometrics of the New Keynesian Policy Model: Introduction', *Oxford Bulletin of Economics and Statistics* **66**, 581–607.

Henzel, S. and Wollmershäuser, T. (2008), 'The New Keynesian Phillips curve and the role of expectations: Evidence from the CESifo World Economic Survey', *Economic Modelling* **25**(5), 811–832.

Hornstein, A. (2007), 'Evolving inflation dynamics and the New Keynesian Phillips curve', *Federal Reserve Bank of Richmond Economic Quarterly* **93**(4), 317–339.

Hume, D. (1752), Of money, *in* 'Essays', George Routledge and Sons.

Imbs, J., Jondeau, E. and Pelgrin, F. (2011), 'Sectoral Phillips curves and the aggregate Phillips curve', *Journal of Monetary Economics* **58**(4), 328–344.

Jondeau, E. and Le Bihan, H. (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, E. and Le Bihan, H. (2005), 'Testing for the New Keynesian Phillips Curve. Additional international evidence', *Economic Modelling* **22**(3), 521–550.

Jondeau, E. and Le Bihan, H. (2008), 'Examining bias in estimators of linear rational expectations models under misspecification', *Journal of Econometrics* **143**(2), 375–395.

Kapetanios, G., Khalaf, L. and Marcellino, M. (2011), 'Factor based identification-robust inference in IV regressions'. Working paper, March 22 version.

Kapetanios, G. and Marcellino, M. (2010), 'Factor-GMM estimation with large sets of possibly weak instruments', *Computational Statistics and Data Analysis* **54**(11), 2655–2675.

Kiley, M. T. (2007), 'A Quantitative Comparison of Sticky-Price and Sticky-Information Models of Price Setting', *Journal of Money, Credit and Banking* **39**(1), 101–125.

Kim, C.-J. and Kim, Y. (2008), 'Is the Backward-Looking Component Important in a New Keynesian Phillips Curve?', *Studies in Nonlinear Dynamics & Econometrics* **12**. Issue 3, Article 5.

Kim, C.-J. and Manopimoke, P. (2011), 'Trend Inflation and the New Keynesian Phillips Curve'. Working paper, July 11 version.

Kleibergen, F. (2002), 'Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression', *Econometrica* **70**(5), 1781–1803.

Kleibergen, F. (2005), 'Testing Parameters in GMM without Assuming That They Are Identified', *Econometrica* **73**(4), 1103–1123.

Kleibergen, F. and Mavroeidis, S. (2009), 'Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve', *Journal of Business & Economic Statistics* **27**(3), 293–311.

Kleibergen, F. and Mavroeidis, S. (2011), 'Identification issues in Bayesian analysis of structural macroeconomic models with an application to the Phillips curve'. Working paper, July version.

Koop, G. and Onorante, L. (2011), 'Estimating Phillips Curves in Turbulent Times using the ECBs Survey of Professional Forecasters'. Working paper, February version.

Korenok, O., Radchenko, S. and Swanson, N. R. (2010), 'International evidence on the efficacy of new-Keynesian models of inflation persistence', *Journal of Applied Econometrics* **25**(1), 31–54.

Kozicki, S. and Tinsley, P. (2002), 'Alternative Sources of the Lag Dynamics of Inflation', *Federal Reserve Bank of Kansas City Research Working Paper* (02-12).

Krause, M. U., Lopez-Salido, D. and Lubik, T. A. (2008), 'Inflation dynamics with search frictions: A structural econometric analysis', *Journal of Monetary Economics* **55**(5), 892–916.

Kuester, K., Müller, G. J. and Stölting, S. (2009), 'Is the New Keynesian Phillips curve flat?', *Economics Letters* **103**(1), 39–41.

Kurmann, A. (2005), 'Quantifying the uncertainty about the fit of a new Keynesian pricing model', *Journal of Monetary Economics* **52**(6), 1119–1134.

Kurmann, A. (2007), 'VAR-based estimation of Euler equations with an application to New Keynesian pricing', *Journal of Economic Dynamics and Control* **31**(3), 767–796.

Lindé, J. (2005), 'Estimating New-Keynesian Phillips curves: A full information maximum likelihood approach', *Journal of Monetary Economics* **52**(6), 1135–1149.

Lubik, T. A. and Schorfheide, F. (2004), 'Testing for Indeterminacy: An Application to U.S. Monetary Policy', *American Economic Review* **94**(1), 190–217.

Lucas, R. E. (1976), 'Econometric policy evaluation: A critique', *Carnegie-Rochester Conference Series on Public Policy* **1**, 19–46.

Ma, A. (2002), 'GMM estimation of the new Phillips curve', *Economics Letters* **76**(3), 411–417.

Magnusson, L. M. and Mavroeidis, S. (2010), 'Identification-Robust Minimum Distance Estimation of the New Keynesian Phillips Curve', *Journal of Money, Credit and Banking* **42**(2–3), 465–481.

Magnusson, L. M. and Mavroeidis, S. (2011), 'Identification using stability restrictions'. Working paper.

Mankiw, N. G. (2001), 'The Inexorable and Mysterious Tradeoff between Inflation and Unemployment', *The Economic Journal* **111**(471), C45–C61.

Mankiw, N. G. and Reis, R. (2011), Imperfect Information and Aggregate Supply, *in* B. M. Friedman and M. Woodford, eds, 'Handbook of Monetary Economics', Vol. 3, Elsevier, chapter 5, pp. 183–229.

Mankiw, N. G., Reis, R. and Wolfers, J. (2004), Disagreement about Inflation Expectations, *in* M. Gertler and K. Rogoff, eds, 'NBER Macroeconomics Annual 2003', Vol. 18, The MIT Press, pp. 209–270.

Martins, L. F. and Gabriel, V. J. (2009), 'New Keynesian Phillips Curves and potential identification failures: A Generalized Empirical Likelihood analysis', *Journal of Macroeconomics* **31**(4), 561–571.

Mavroeidis, S. (2004), 'Weak Identification of Forward-looking Models in Monetary Economics', *Oxford Bulletin of Economics and Statistics* **66**(supplement), 609–635.

Mavroeidis, S. (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, S. (2010), 'Monetary Policy Rules and Macroeconomic Stability: Some New Evidence', *American Economic Review* **100**(1), 491–503.

Mazumder, S. (2010), 'The new Keynesian Phillips curve and the cyclicality of marginal cost', *Journal of Macroeconomics* **32**(3), 747–765.

Mazumder, S. (2011), 'The empirical validity of the New Keynesian Phillips curve using survey forecasts of inflation', *Economic Modelling* **28**(6), 2439–2450.

McAdam, P. and Willman, A. (2010), 'Technology, Utilization and Inflation: Re-assessing the New Keynesian Fundamental'. Working paper, October 15 version.

McCallum, B. T. (1976), 'Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates', *Econometrica* **44**(1), 43–52.

Montiel Olea, J. L. and Pflueger, C. (2011), 'A Heteroskedasticity and Autocorrelation Robust Pre-Test for Weak Instruments'. Working paper, February version.

Moreira, M. J. (2003), 'A Conditional Likelihood Ratio Test for Structural Models', *Econometrica* **71**(4), 1027–1048.

Nakamura, E. and Steinsson, J. (2013), 'Price Rigidity: Microeconomic Evidence and Macroeconomic Implications', *National Bureau of Economic Research Working Paper Series* (18705).

Nason, J. M. and Smith, G. W. (2008*a*), 'Identifying the new Keynesian Phillips curve', *Journal of Applied Econometrics* **23**(5), 525–551.

Nason, J. M. and Smith, G. W. (2008*b*), 'The New Keynesian Phillips Curve: Lessons From Single-Equation Econometric Estimation', *Federal Reserve Bank of Richmond Economic Quarterly* **94**(4), 361–395.

Neiss, K. S. and Nelson, E. (2005), 'Inflation Dynamics, Marginal Cost, and the Output Gap: Evidence from Three Countries', *Journal of Money, Credit and Banking* **37**(6), 1019–1045.

Nelson, C. R. and Lee, J. (2007), 'Expectation horizon and the Phillips Curve: the solution to an empirical puzzle', *Journal of Applied Econometrics* **22**(1), 161–178.

Newey, W. K. and McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. F. Engle and D. L. McFadden, eds, 'Handbook of Econometrics', Vol. 4, Elsevier, pp. 2111–2245.

Newey, W. K. and West, K. D. (1987), 'A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica* **55**(3), 703–708.

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

Nunes, R. (2010), 'Inflation Dynamics: The Role of Expectations', *Journal of Money, Credit and Banking* **42**(6), 1161–1172.

Nymoen, R., Swensen, A. R. and Tveter, E. (2010), 'The New Keynesian Phillips Curve: A meta-analysis'. Working paper, October 7 version.

Ólafsson, T. T. (2006), 'The New Keynesian Phillips Curve: In Search of Improvements and Adaptation to the Open Economy', *Central Bank of Iceland Working Papers* (31).

Paloviita, M. (2006), 'Inflation Dynamics in the Euro Area and the Role of Expectations', *Empirical Economics* **31**, 847–860.

Paloviita, M. (2008), 'Comparing alternative Phillips curve specifications: European results with survey-based expectations', *Applied Economics* **40**(17), 2259–2270.

Paloviita, M. and Mayes, D. (2005), 'The use of real-time information in Phillips-curve relationships for the euro area', *The North American Journal of Economics and Finance* **16**(3), 415–434.

Phelps, E. S. (1967), 'Phillips Curves, Expectations of Inflation and Optimal Unemployment over Time', *Economica* **34**(135), 254–281.

Phillips, A. W. (1958), 'The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957', *Economica* **25**, 283–299.

Preston, B. (2005), 'Learning about Monetary Policy Rules when Long-Horizon Expectations Matter', *International Journal of Central Banking* **1**(2), 81–126.

Ravenna, F. and Walsh, C. E. (2006), 'Optimal monetary policy with the cost channel', *Journal of Monetary Economics* **53**(2), 199–216.

Ravenna, F. and Walsh, C. E. (2008), 'Vacancies, unemployment, and the Phillips curve', *European Economic Review* **52**(8), 1494–1521.

Roberts, J. M. (1995), 'New Keynesian Economics and the Phillips Curve', *Journal of Money, Credit and Banking* **27**(4), 975–984.

Roberts, J. M. (1997), 'Is inflation sticky?', *Journal of Monetary Economics* **39**(2), 173–196.

Roberts, J. M. (2005), 'How Well Does the New Keynesian Sticky-Price Model Fit the Data?', *Contributions to Macroeconomics* **5**. Issue 1, Article 10.

Rotemberg, J. J. (1982), 'Sticky Prices in the United States', *Journal of Political Economy* **90**(6), 1187–1211.

Rudd, J. and Whelan, K. (2005), 'New tests of the new-Keynesian Phillips curve', *Journal of Monetary Economics* **52**(6), 1167–1181.

Rudd, J. and Whelan, K. (2006), 'Can Rational Expectations Sticky-Price Models Explain Inflation Dynamics?', *The American Economic Review* **96**(1), 303–320.

Rudd, J. and Whelan, K. (2007), 'Modeling Inflation Dynamics: A Critical Review of Recent Research', *Journal of Money, Credit and Banking* **39**, 155–170.

Rudebusch, G. D. (2002), 'Assessing Nominal Income Rules for Monetary Policy with Model and Data Uncertainty', *The Economic Journal* **112**(479), 402–432.

Rumler, F. (2007), 'Estimates of the Open Economy New Keynesian Phillips Curve for Euro Area Countries', *Open Economies Review* **18**(4), 427–451.

Sahuc, J.-G. (2006), 'Partial indexation, trend inflation, and the hybrid Phillips curve', *Economics Letters* **90**(1), 42–50.

Samuelson, P. A. and Solow, R. M. (1960), 'Analytical Aspects of Anti-Inflation Policy', *The American Economic Review* **50**(2), 177–194.

Sargent, T. and Wallace, N. (1975), '"Rational" Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule', *Journal of Political Economy* **83**(2), 241–254.

Sbordone, A. M. (2002), 'Prices and unit labor costs: a new test of price stickiness', *Journal of Monetary Economics* **49**(2), 265–292.

Sbordone, A. M. (2005), 'Do expected future marginal costs drive inflation dynamics?', *Journal of Monetary Economics* **52**(6), 1183–1197.

Sbordone, A. M. (2006), 'U.S. Wage and Price Dynamics: A Limited-Information Approach', *International Journal of Central Banking* **2**(3), 155–191.

Schorfheide, F. (2011), 'Estimation and Evaluation of DSGE Models: Progress and Challenges', *National Bureau of Economic Research Working Paper Series* (16781).

Shapiro, A. H. (2008), 'Estimating the New Keynesian Phillips Curve: A Vertical Production Chain Approach', *Journal of Money, Credit and Banking* **40**(4), 627–666.

Sheedy, K. D. (2010), 'Intrinsic inflation persistence', *Journal of Monetary Economics* **57**(8), 1049–1061.

Smets, F. and Wouters, R. (2007), 'Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach', *American Economic Review* **97**(3), 586–606.

Smith, G. W. (2009), 'Pooling forecasts in linear rational expectations models', *Journal of Economic Dynamics and Control* **33**(11), 1858–1866.

Sowell, F. (1996), 'Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework', *Econometrica* **64**(5), 1085–1107.

Stock, J. H. and Watson, M. W. (2007), 'Why Has U.S. Inflation Become Harder to Forecast?', *Journal of Money, Credit and Banking* **39**(1), 3–33.

Stock, J. H. and Watson, M. W. (2010), 'Modeling Inflation After the Crisis', *National Bureau of Economic Research Working Papers* (16488). Prepared for the Federal Reserve Bank of Kansas City Symposium at Jackson Hole.

Stock, J. H. and Wright, J. (2000), 'GMM with Weak Identification', *Econometrica* **68**(5), 1055–1096.

Stock, J. H., Wright, J. H. and Yogo, M. (2002), 'A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments', *Journal of Business & Economic Statistics* **20**(4), 518–529.

Stock, J. H. and Yogo, M. (2002), 'Testing for Weak Instruments in Linear IV Regression', *National Bureau of Economic Research Working Papers* (0284).

Taylor, J. B. (1979), 'Staggered Wage Setting in a Macro Model', *American Economic Review* **69**(2), 108–113.

Taylor, J. B. (1980), 'Aggregate Dynamics and Staggered Contracts', *Journal of Political Economy* **88**(1), 1–23.

Thomas, L. B. (1999), 'Survey Measures of Expected U.S. Inflation', *Journal of Economic Perspectives* **13**(4), 125–144.

Tillmann, P. (2008), 'Do interest rates drive inflation dynamics? An analysis of the cost channel of monetary transmission', *Journal of Economic Dynamics and Control* **32**(9), 2723–2744.

Tsoukis, C., Kapetanios, G. and Pearlman, J. (2011), 'Elusive Persistence: Wage and Price Rigidities, the New Keynesian Phillips Curve and Inflation Dynamics', *Journal of Economic Surveys* **25**(4), 737–768.

White, H. (1980), 'A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity', *Econometrica* **48**(4), 817–838.

Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press.

Yao, F. (2011), 'Monetary Policy, Trend Inflation and Inflation Persistence'. SFB 649 Discussion Paper 2011-008.

Yun, T. (1996), 'Nominal price rigidity, money supply endogeneity, and business cycles', *Journal of Monetary Economics* **37**(2), 345–370.

Zhang, C. and Clovis, J. (2010), 'The New Keynesian Phillips Curve of rational expectations: A serial correlation extension', *Journal of Applied Economics* **13**(1), 159–179.

Zhang, C., Osborn, D. R. and Kim, D. H. (2008), 'The New Keynesian Phillips Curve: From Sticky Inflation to Sticky Prices', *Journal of Money, Credit and Banking* **40**(4), 667–699.

Zhang, C., Osborn, D. R. and Kim, D. H. (2009), 'Observed Inflation Forecasts and the New Keynesian Phillips Curve', *Oxford Bulletin of Economics and Statistics* **71**(3), 375–398.

OVERVIEW OF ESTIMATION APPROACHES IN THE LITERATURE

Papers	Estimation approach	Expectation vs. lags	Slope	Rejection of model?
Galí and Gertler (1999), Galí et al. (2001, 2005)	RE GMM.	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)	GMM, VAR-ML, IRF matching; RE and survey forecasts.	Sluggish survey forecasts impart necessary persistence. For RE, need more than 1 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 GMM (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 GMM, 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 stochastic volatility.	Backward-looking term in-significant once trend infla-tion is accounted for.	(Not directly estimated.)	No, based on visual fit and magnitude of forecast errors.

Table 1: 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.

Data vintage	const	λ	γ_f	γ_b	Hansen test
1998	0.041	0.026	0.615	0.340	5.263
	(0.030)	(0.013)	(0.057)	(0.058)	[0.628]
2012	-0.049	0.018	0.719	0.240	9.816
	(0.040)	(0.012)	(0.099)	(0.095)	[0.199]

Table 2: Comparison of 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. The instrument set includes four lags of inflation and two lags of the labor share, wage inflation, and quadratically-detrended output. The estimation method is CUE GMM. Newey and West (1987) weight matrix with automatic lag truncation (4 lags). Standard errors in parentheses and p-values in square brackets.

Specification settings	Options
Inflation (π_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, infl. and forcing variable evolve as VAR
Survey forecasts ($\pi^s_{t\mid\tau}$)	SPF CPI, SPF GDP def., GB GDP def.
Expectations	π_{t+1} (endogenous), $\pi^s_{t+1\mid t}$ (endog.), $\pi^s_{t+1\mid t}$ (exogenous) $\pi^s_{t+1\mid t-1}$ (endog.), $\pi^s_{t+1\mid t-1}$ (exog.)
Instruments	GG: 4 lags of π_t, ls, ygap, 10y–90d yield spread, wage infl., commodity price infl. GGLS: 4 lags of π_t and 2 lags of ls, ygap, wage infl. small: 4 lags of π_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, 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 coef's sum to 1) With $\gamma(1) = \gamma_f$, use lags of $\Delta\pi_t$ instead of π_t as instr's
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

Table 3: Different specifications of the NKPC (9).

Parameter	labor share		output gap	
	λ	γ_f	λ	γ_f
median	0.004	0.753	0.004	0.760
5th percentile	-0.068	-0.648	-0.070	-0.771
95th percentile	0.135	1.814	0.133	1.831
fraction both positive	0.525		0.505	
. . . and signif. (5% 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% Hansen test	0.033		0.032	

Table 4: Summary statistics of estimated results across specifications listed in Table 3, excluding real-time and survey instrument sets. Hansen test is evaluated at the CUE using larger critical values that are robust to weak identification; results for this statistic do not include VAR-GMM specifications.

Parameter	labor share				output gap			
	λ		γ_f		λ		γ_f	
Method	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% 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

Table 5: Comparison of 2-step and CUE GMM estimates for the specifications in Table 3 (excluding VAR specifications, for which we only computed 2-step GMM).

Parameter	labor share				output gap			
	λ		γ_f		λ		γ_f	
Method	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% 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	

Table 6: Effect of imposing the VAR identification assumption.

	labor share		output gap	
Parameter	λ	γ_f	λ	γ_f
SPF vs RE: post-81	0.002	-0.200	0.042	0.092
SPF vs RE: CPI post-81	-0.002	0.030	0.049	0.286
SPF vs RE: GDP def. post-81	0.004	-0.326	0.038	-0.043
SPF vs RE: GDP def. post-68	0.002	-0.200	0.042	0.092
GB vs RE: post-68	0.039	-0.093	0.047	-0.000
endogenous vs exogenous	-0.000	-0.000	-0.001	-0.045
GB vs SPF	0.003	-0.016	-0.007	-0.063
GB vs SPF: pre-1984	-0.010	0.069	-0.017	0.007
GB vs SPF: post-1984	0.009	-0.076	0.003	-0.103

Table 7: 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 vs RE yields median differences of coefficient estimates of survey specifications from the corresponding GIV (RE) specifications.

	labor share		output gap	
Parameter	λ	γ_f	λ	γ_f
RT vs small instr.	-0.003	0.008	-0.003	-0.012
survey vs small instr.	0.001	0.001	-0.001	-0.018
1 vs 0 lags	-0.001	-0.256	-0.000	-0.256
4 vs 1 lags	-0.004	-0.223	0.009	-0.277
post- vs pre-84 sample	-0.008	-0.024	-0.002	0.000
GDP def. vs CPI inflation	0.001	0.011	0.004	0.018
gap vs level of inflation	-0.001	-0.054	0.013	-0.072
oil vs no controls	0.000	0.001	-0.001	-0.001
Treas. rate vs no controls	-0.000	0.005	0.001	0.000

Table 8: Median pairwise differences in estimates across the specifications listed specifications in Table 3 that differ by one characteristic, keeping all other specification aspects constant. For example, row "RT vs small instr." compares the specifications that use "small" and "RT" instruments, and similarly for "survey vs small instr."; row "1 vs 0 lags" gives the median difference of estimates in the hybrid NKPC with 1 lag from the corresponding estimates in the pure NKPC, and similarly for "4 vs 1 lags". The remaining rows follow the same pattern.

Specification settings	Options
Inflation (π_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 ($\pi^s_{t\mid\tau}$)	SPF CPI, SPF GDP def., GB GDP def.
Expectations	π_{t+1} (endogenous), $\pi^s_{t+1\mid t}$ (endogenous)
Instruments	GGLS, small, exact, RT, survey
Inflation lags	0, 1, 4
Parameter restrictions	No restrictions, $\gamma(1) = \gamma_f$ (inflation coef's sum to 1)
Sample	Full available, pre–1984, 1984–end of sample

Table 9: Different specifications of the NKPC for which we compute robust confidence sets.

	NKPC		Red. form: forcing var.				Red. form: inflation				MA	
DGP	γ_f	λ	$\xi_{\pi 1}$	ξ_{x1}	$\xi_{\pi 2}$	ξ_{x2}	$\zeta_{\pi 1}$	ζ_{x1}	$\zeta_{\pi 2}$	ζ_{x2}	b'	Conc.
1	0.7	0.03	0.20	0.80	−0.10	0.10	1.00	−0.07	0.01	−0.01	—	1.8
2	0.7	0.03	0.10	0.70	0.10	−0.10	0.79	0.25	0.05	−0.05	—	108.4
3	0.7	0.03	0.20	0.80	−0.10	0.10	1.43	−0.05	−0.43	0.00	$(-1.4, -0.3)$	88.7
4	0.3	−0.03	0.20	0.80	−0.10	0.10	0.98	−0.06	0.01	−0.01	—	0.7
5	0.3	−0.03	0.70	1.05	0.08	−0.08	0.91	−0.07	−0.01	0.01	—	30.1

Table 10: List of our DGPs. Columns 2 and 3 list the true values of (γ_f, λ). The following 8 columns list the VAR(2) reduced form coefficients in the equation for the forcing variable (ξ) and inflation (ζ). The last two columns list the MA(1) parameter b and the minimum eigenvalue of the population concentration matrix ("Conc."). Numbers in columns 8–11 have been rounded off to 2 decimal points, while numbers in the last column have been rounded off to 1 decimal point.

		labor share		output gap	
	Confidence level	90%	95%	90%	95%
spec.	Sample				
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
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	1984q1–end of sample	0.33	0.48	0.43	0.60
small	1984q1–end of sample	0.46	0.66	0.43	0.59
exact	1984q1–end of sample	0.29	0.42	0.32	0.52
RT	1984q1–end of sample	0.64	0.78	0.54	0.73
survey	1984q1–end of sample	0.32	0.47	0.35	0.52

Table 11: Size of 90% and 95% S sets corresponding to the different specifications of the NKPC in Table 9, as a fraction of the parameter space shown in Figure 20 (using 1025 grid points).

	labor share		output gap	
Level	90%	95%	90%	95%
No. of specs	698	698	698	698
fraction empty S set	0.038	0.013	0.016	0.004
fraction non-empty 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	0.062	0.021	0.042	0.000
small	0.035	0.014	0.035	0.021
exact	0.000	0.000	0.079	0.043
RT	0.007	0.007	0.000	0.000
survey	0.069	0.008	0.092	0.038

Table 12: Anatomy of 90% and 95% S sets for specifications listed in Table 9. S sets are computed on the parameter space shown in Figure 20 using 1025 grid points. "Fraction empty" lists fraction of cases with empty S set. "Fraction non-empty and positive" lists fraction of cases for which S set is non-empty and lies entirely in positive orthant. Rows "all RE" to "survey" give the latter fraction for different subcases.
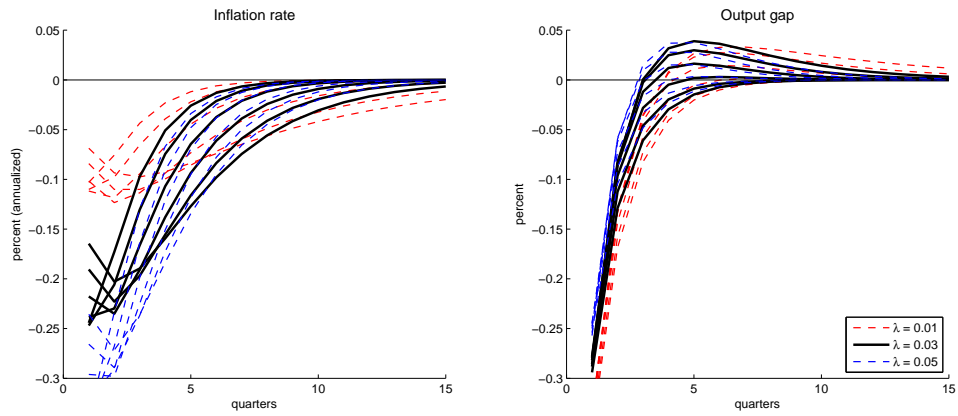
Figure 1: Impulse responses of inflation and the output gap to a 25 basis point monetary policy shock in a standard three-equation New Keynesian model with $\gamma_f = 0.3, 0.4, \ldots, 0.8$ and $\lambda = 0.01, 0.03, 0.05$. The other parameters are calibrated to the benchmark values listed in Appendix A.1. More sluggish responses correspond to lower values of γ_f. The figure was generated using Dynare (Adjemian et al., 2011).
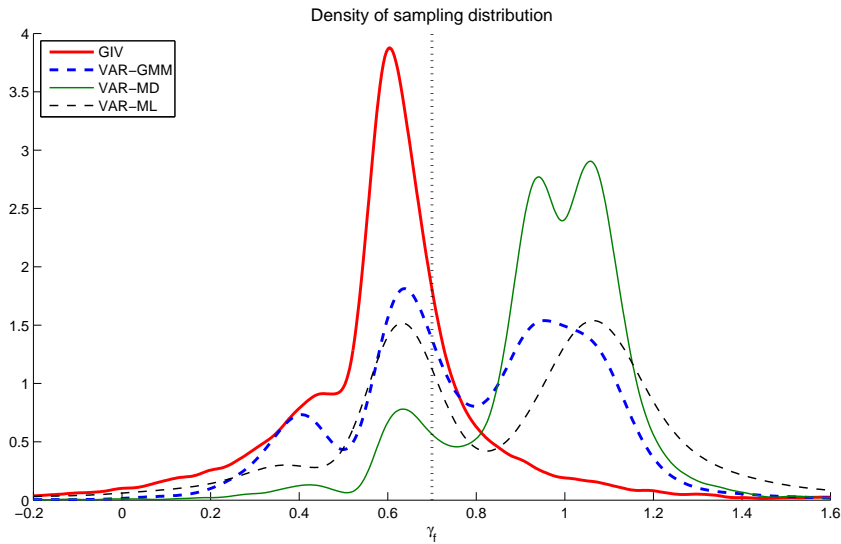
Figure 2: Kernel-smoothed density estimates of the sampling distribution of γ_f estimators, *DGP 1*. The dotted vertical line marks the true parameter value.

66
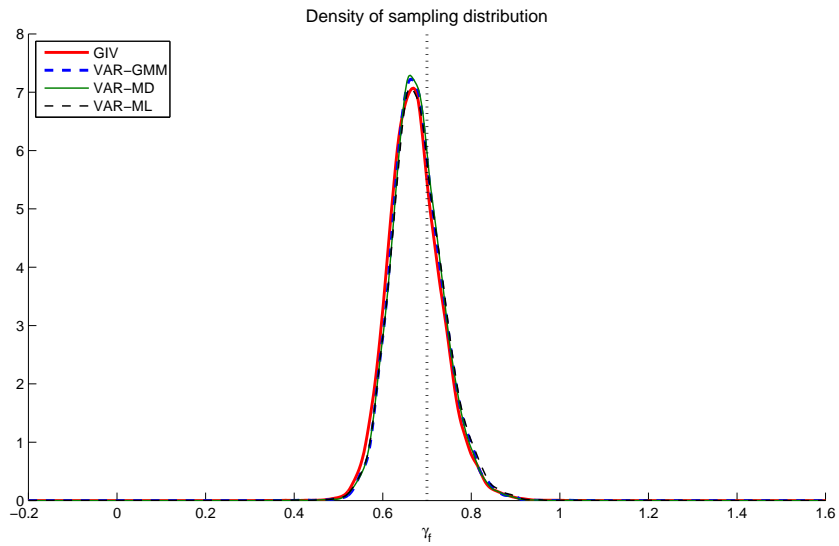
Figure 3: Sampling distribution of γ_f estimators, *DGP 2*.
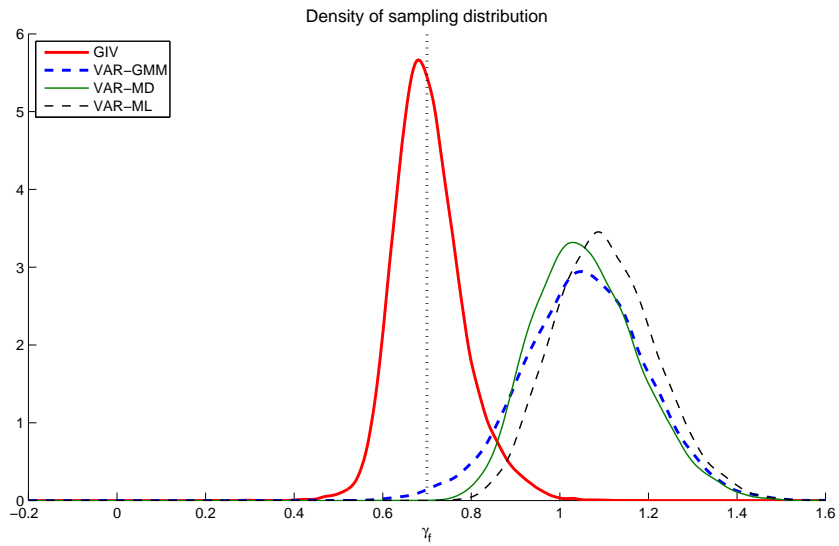
Figure 4: Sampling distribution of γ_f estimators, *DGP 3*.
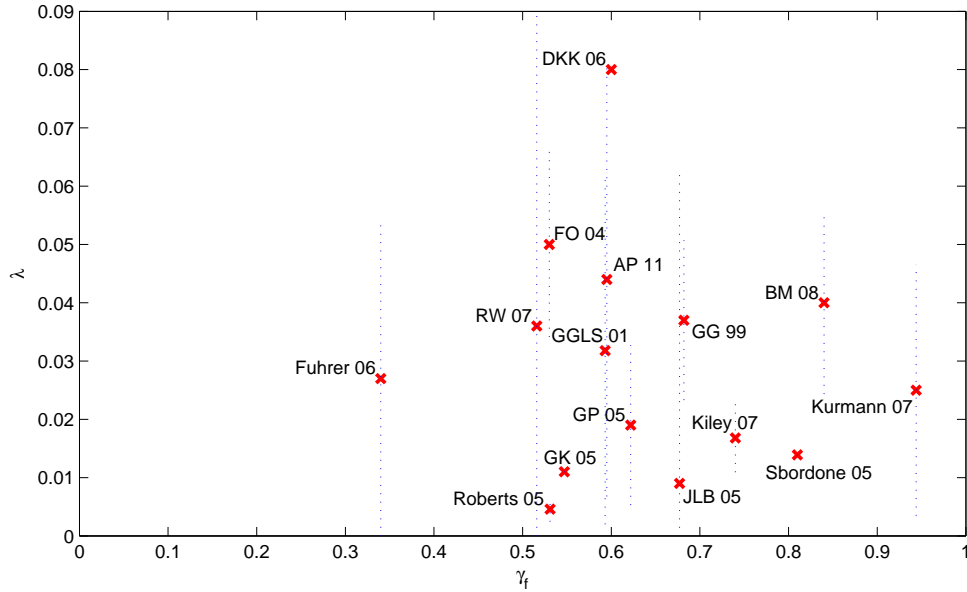
Figure 5: Point estimates of λ (vertical axis) and γ_f (horizontal axis) reported in the literature. Only estimates that use U.S. data and the labor share as forcing variable are plotted. For some papers the semi-structural point estimates have been imputed from point estimates of deeper parameters. The dotted blue lines indicate 95% confidence intervals for λ where available. We include papers with readily available estimates and more than 25 Google Scholar citations as of mid-September 2012: Galí and Gertler (1999), Galí et al. (2001), Fuhrer and Olivei (2004), Gagnon and Khan (2005), Guay and Pelgrin (2005), 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).
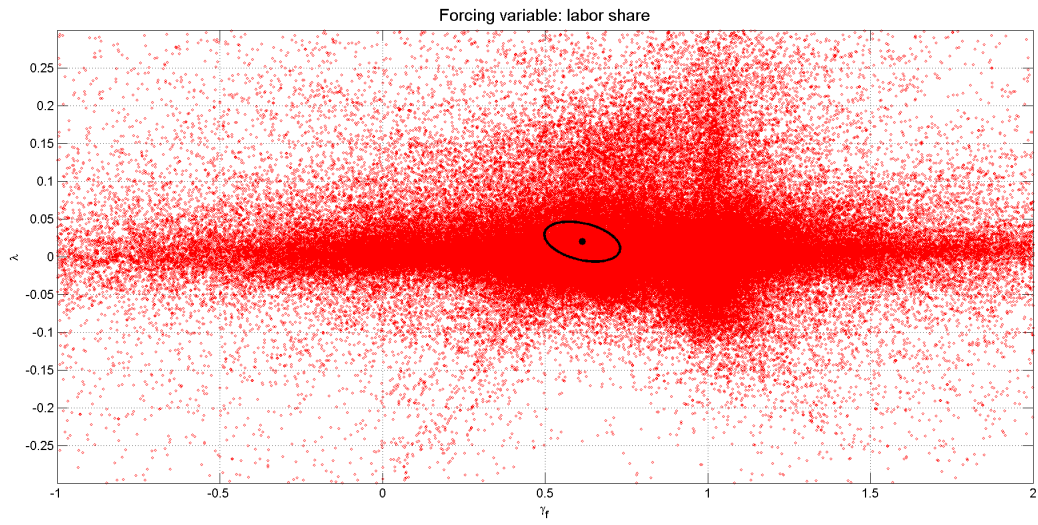
Figure 6: Point estimates of λ, γ_f from the various specifications listed in Table 3 that use the labor share as forcing variable, excluding real-time and survey instrument sets. The black dot and ellipse represent the point estimate and 90% joint Wald confidence set from the 1998 vintage results in Table 2.
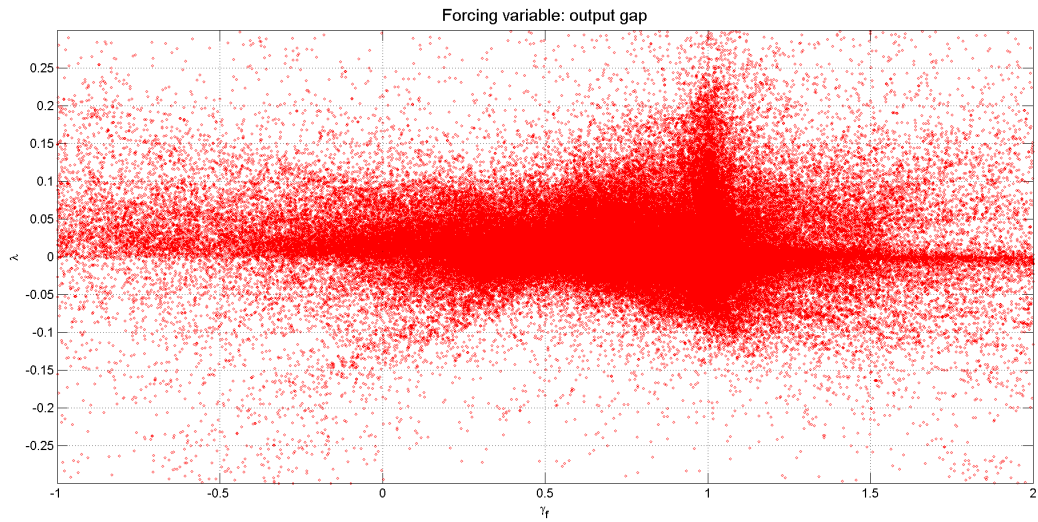
Figure 7: Point estimates of λ, γ_f from the various specifications listed in Table 3 that use the output gap as forcing variable, excluding real-time and survey instrument sets.
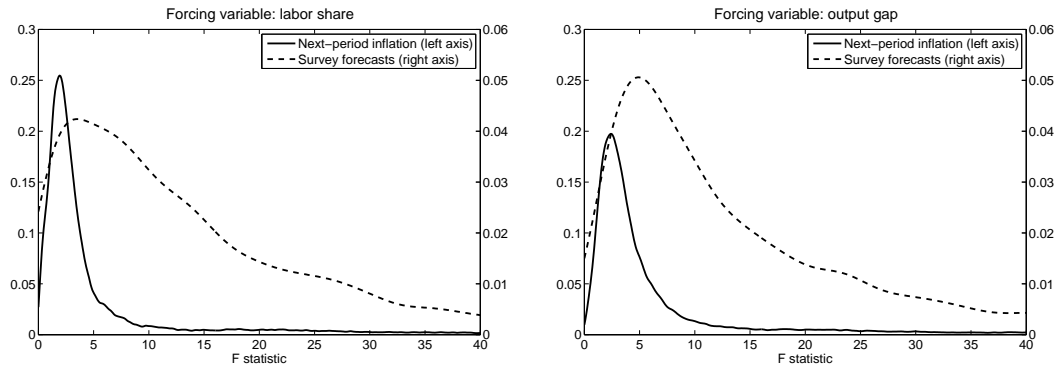
Figure 8: Smoothed density estimates of robust first-stage F statistics for forecasting the expectation proxy, using realized next-period inflation (solid line, left axis) or time-t dated endogenous SPF/GB survey forecasts of inflation (dotted line, right axis). The left and right panels show results for all labor share and output gap specifications, respectively, that are listed in Table 3.
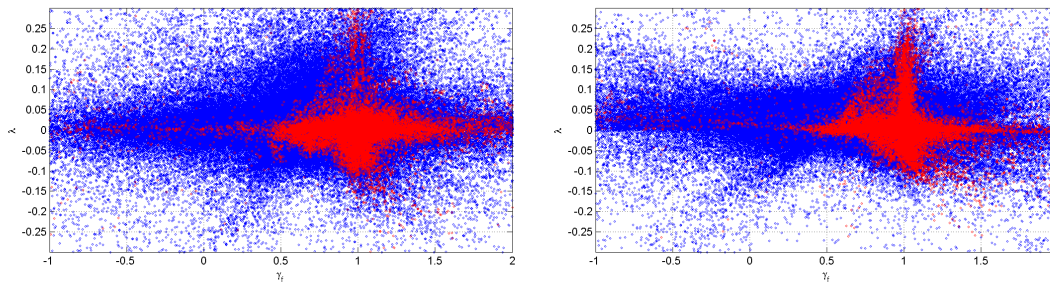
Figure 9: Point estimates of the coefficient on the forcing variable and inflation expectations for the specification combinations listed in Table 3. The red points correspond to estimates that impose the VAR assumption, while the blue points do not impose the assumption. The left and right panel plot specifications with the labor share and output gap as forcing variable, resp.
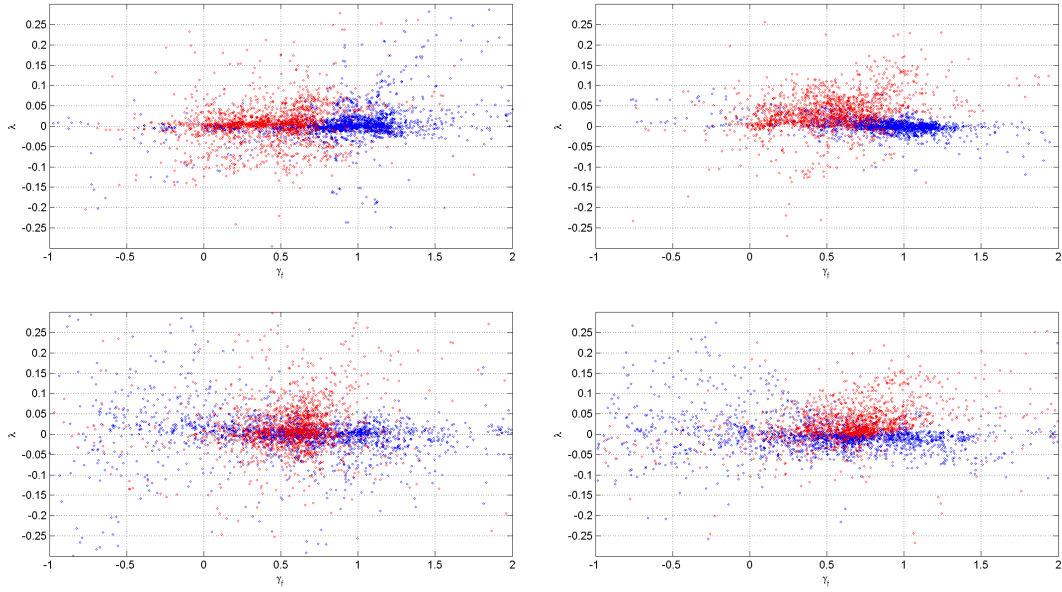
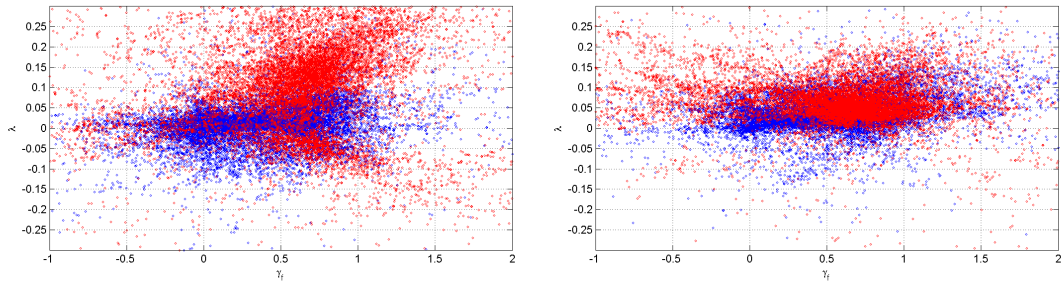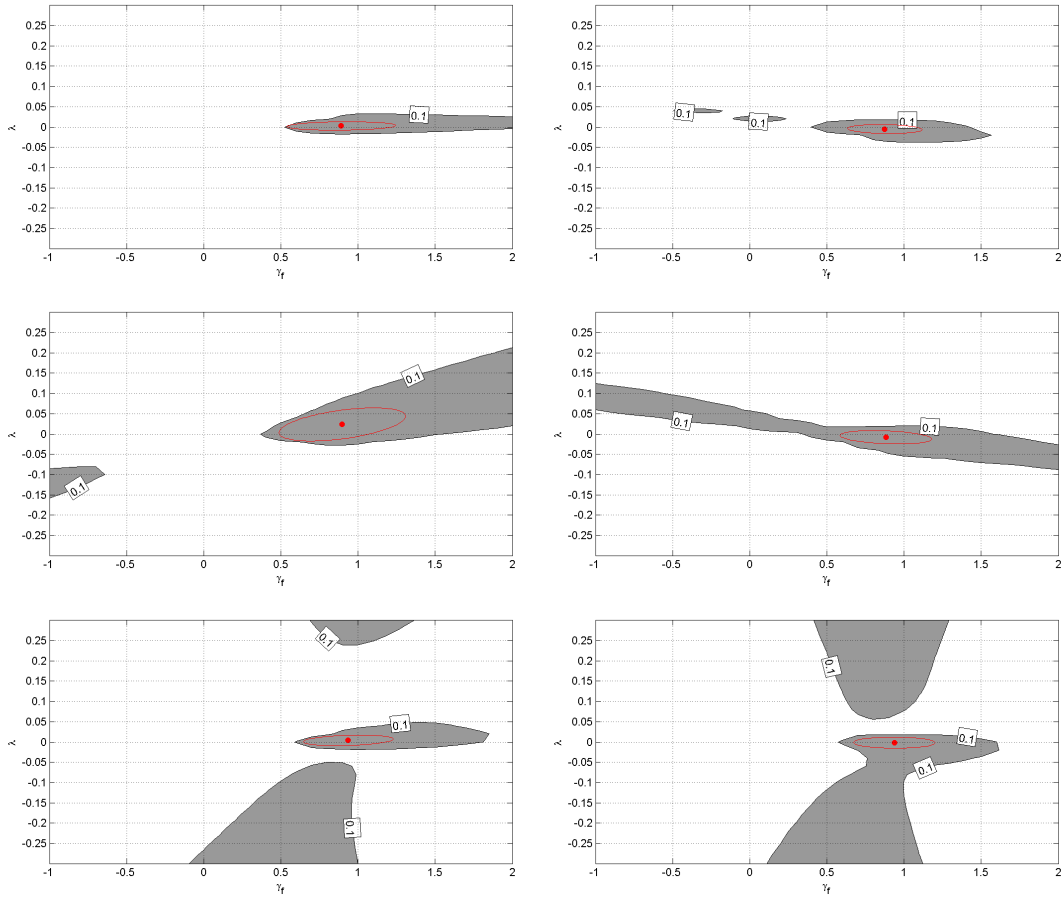**Figure 10:** Point estimates of the coefficient on the forcing variable and inflation expectations using time-$(t-1)$ dated exogenous SPF forecasts (red) versus RE (blue), for the various specification settings listed in Table 3. The top row of figures use GDP deflator inflation, while the bottom row uses CPI inflation. In each row, the left and right panels correspond to labor share and output gap specifications, resp. Sample: 1984q1–2011q4.

71
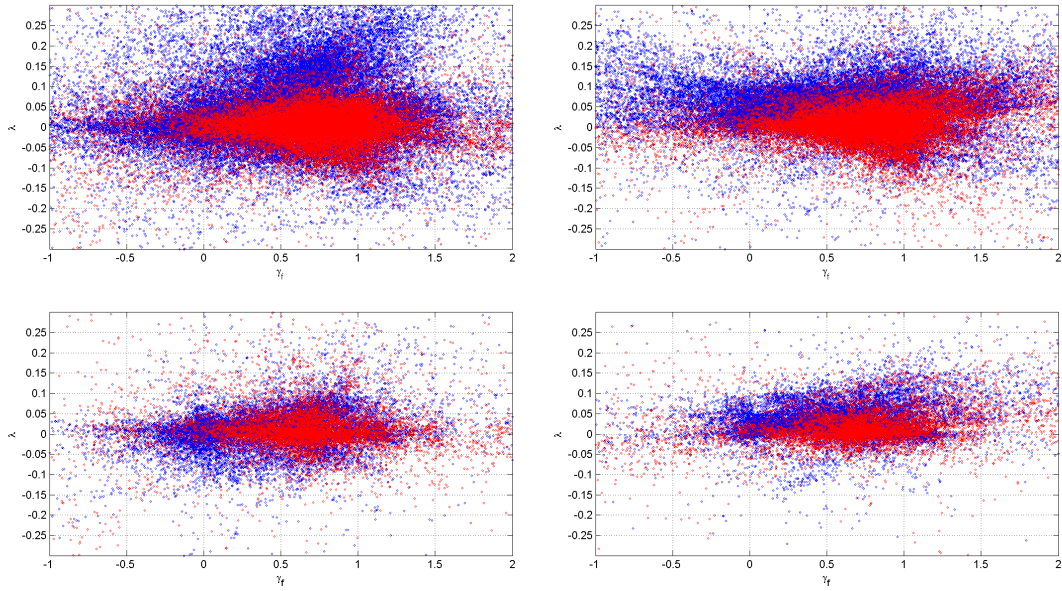
Figure 11: Point estimates of the coefficient on the forcing variable and inflation expectations using CPI (red) versus GDP deflator inflation (blue), for the various specification settings listed in Table 3. The top row gives results for all available samples, while the bottom row selects only the post-1984 sample. The left and right panels plot specifications with the labor share and output gap as forcing variable, resp.

Figure 12: Point estimates of the coefficient on the forcing variable and inflation expectations in survey forecast specifications (SPF and GB) on pre-1984 (red) and post-1984 (blue) 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, resp.
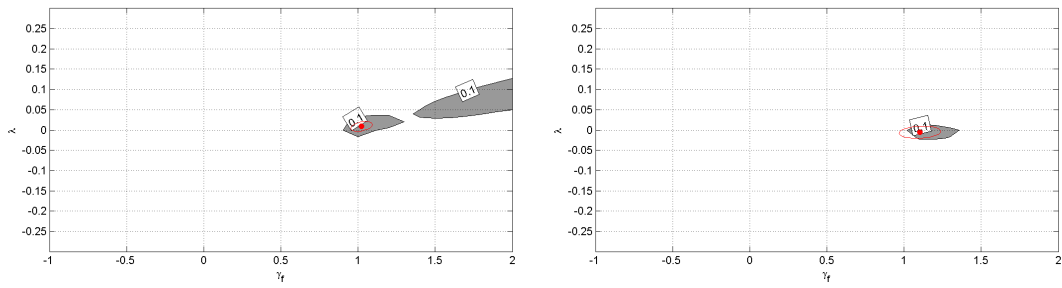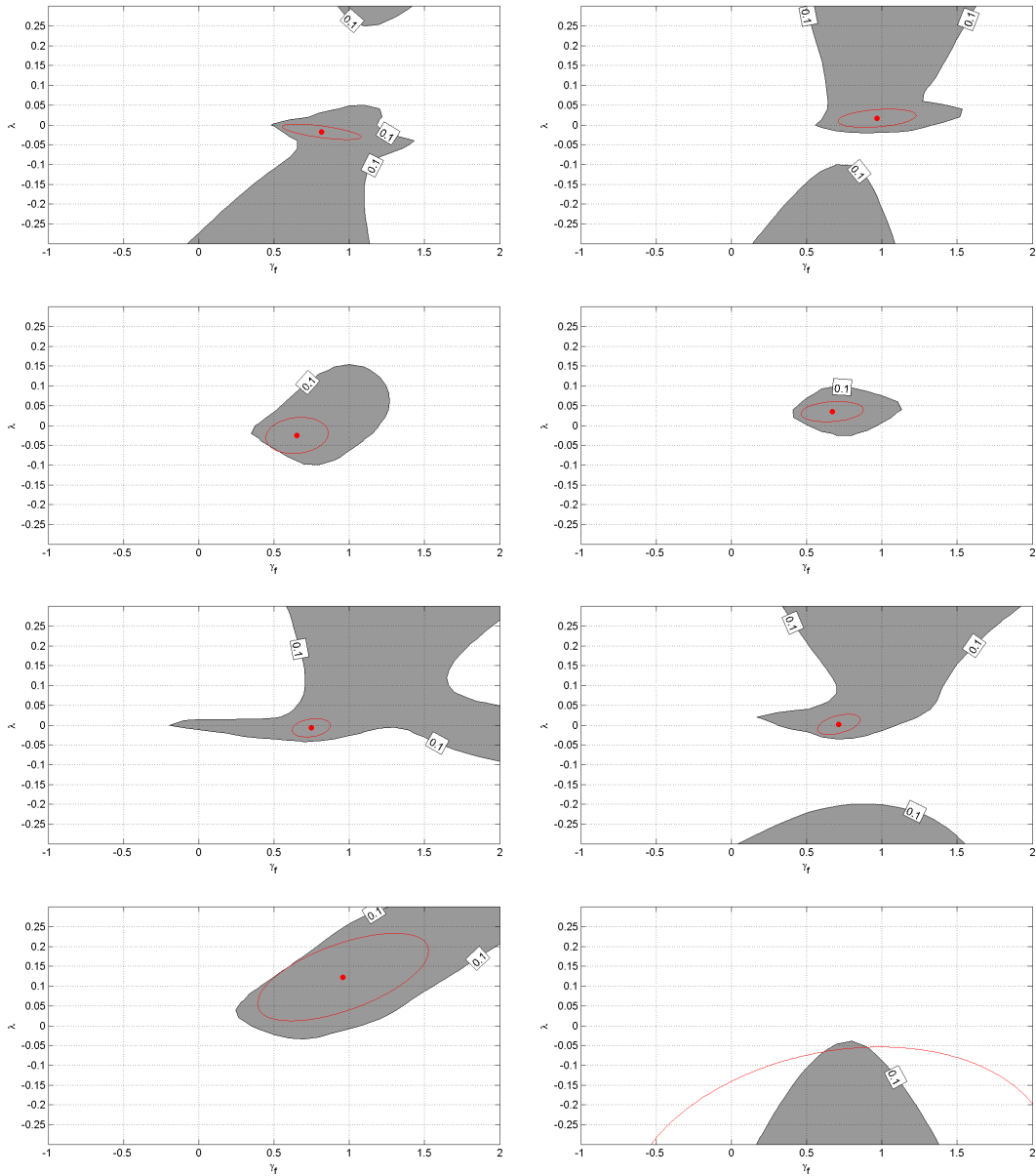
Figure 13: 90% S sets (grey), Wald ellipses and CUE GMM point estimate (red bullet) of the coefficients of the labor share and future inflation in the hybrid NKPC specification with one lag of inflation, where inflation coefficients sum to 1. Inflation: GDP deflator. Forcing variable: NFB labor share (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.

Figure 14: 90% S sets, Wald ellipses 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 1. Reduced form is VAR(3) in $(\Delta\pi_t, x_t)$. Inflation: GDP deflator. Forcing variable (x_t): 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. Weight matrix: White (1980) heteroskedasticity-consistent.
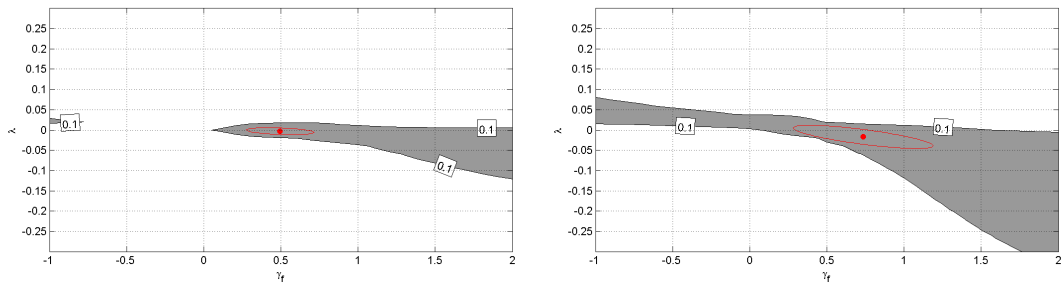
Figure 15: 90% S sets (dark grey), Wald ellipses and CUE GMM point estimate (red bullet) of the coefficients of the labor share and future inflation in the hybrid NKPC specification with one lag of inflation, where inflation coefficients sum to 1. Forcing variable: NFB labor share (left panels), CBO 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 $\pi_{t+1|t}^s$. Weight matrix: Newey-West with automatic lag truncation.

Figure 16: 90% S sets (dark grey), Wald ellipses and CUE GMM point estimate (red 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 1. Inflation: GDP deflator. Forcing variable: NFB labor share (left panels), CBO output gap (right panels). Instruments: RT. Sample: 1971q1-2011q2. Weight matrix: Newey-West with automatic lag truncation.
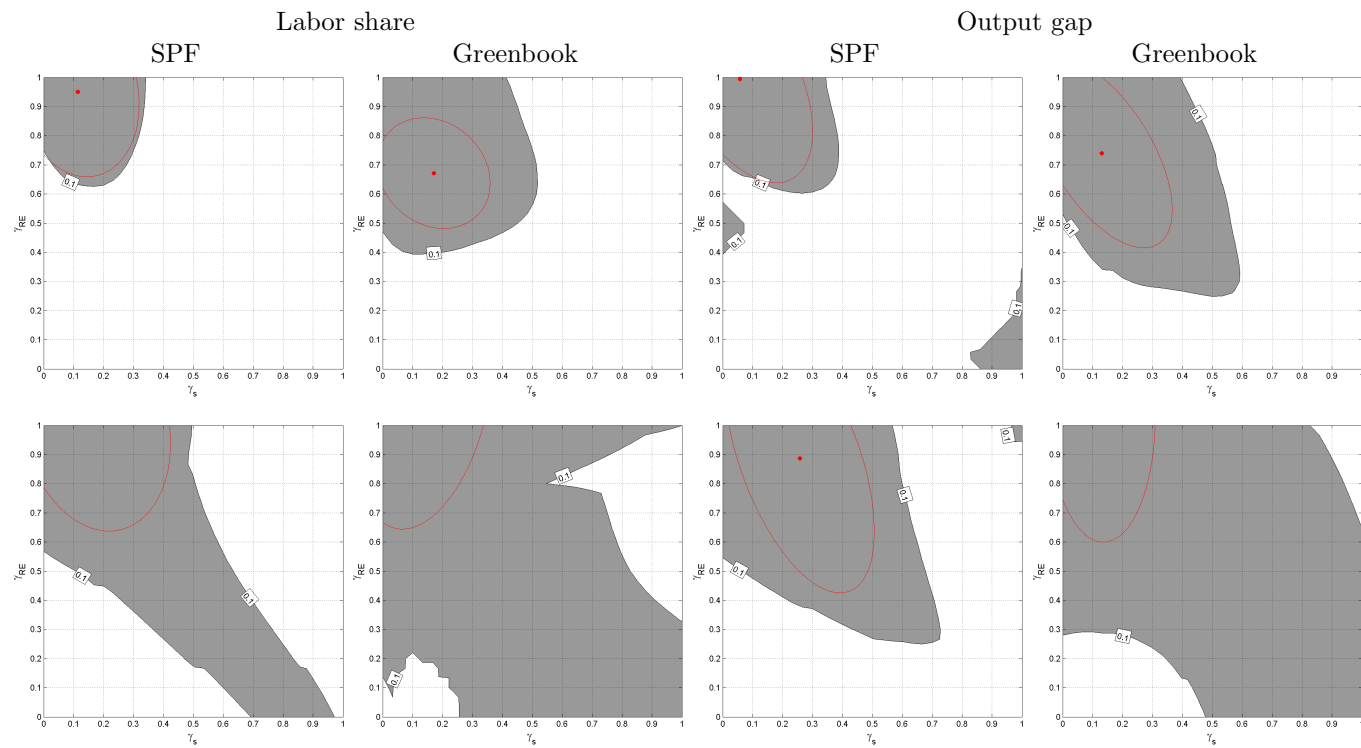
76

Figure 17: 90% S sets, Wald ellipses and CUE GMM point estimate (bullet) of the coefficients of future inflation (γ_{RE}, vertical axis) and 1-quarter forecasts (γ_s, horizontal axis) in the nesting NKPC specification (25). The instrument set is GGLS plus two lags of survey forecasts. Inflation: GDP deflator. Forcing variable: NFB labor share (left two columns), CBO output gap (right two columns). Sample: full available (top row), 1984–end of sample (bottom row). Newey-West Weight matrix with automatic lag truncation.
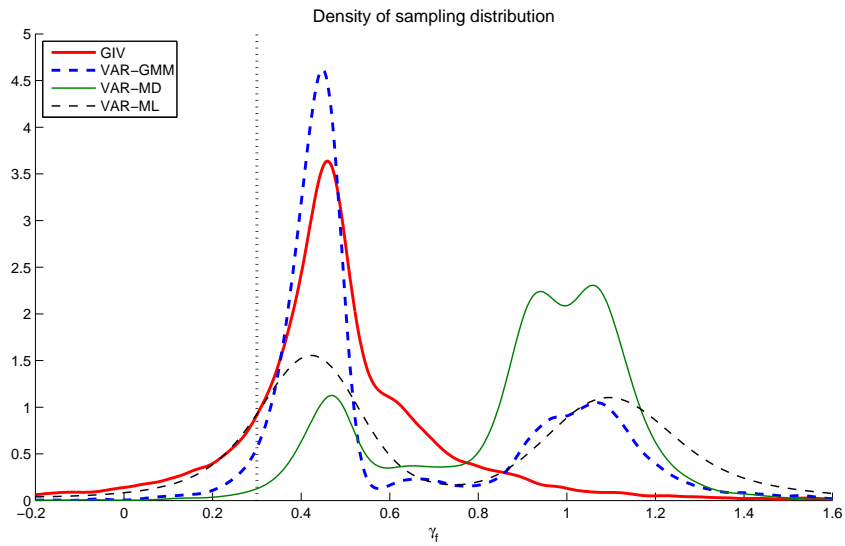
Figure 18: Sampling distribution of γ_f estimators, *DGP 4*.
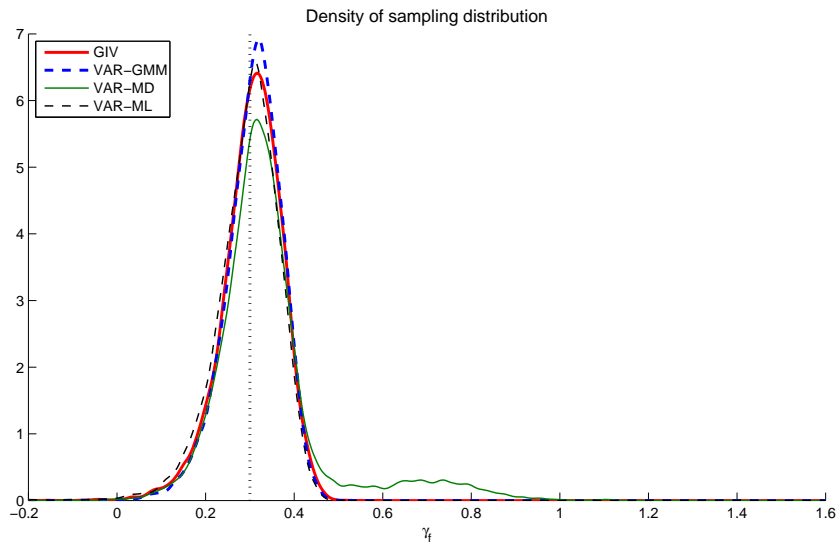

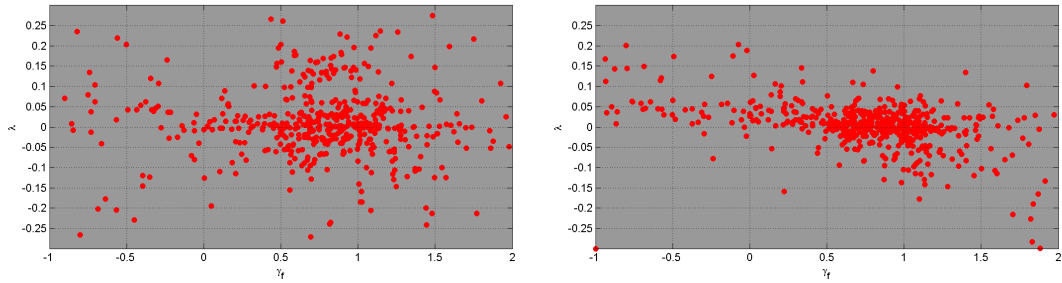**Figure 19:** Sampling distribution of γ_f estimators, *DGP 5*.

Figure 20: Point estimates (red dots) and joint 90% S set (grey area) of the coefficient on the forcing variable and inflation expectations for the various specification settings listed in Table 9. The left and right panels plot specifications with the labor share and output gap as forcing variable, resp.