Testing the New Keynesian Phillips Curve without assuming identification *

Sophocles Mavroeidis

Brown University

sophocles_mavroeidis@brown.edu

First version: 15 April 2006

This version March 12, 2007

Abstract

We re-examine the evidence on the new Phillips curve model of Gali and Gertler (Journal of Monetary Economics 1999) using inference procedures that are robust to weak identification. In contrast to earlier studies, we find that US postwar data are consistent both with the view that inflation dynamics are forward-looking, and with the opposite view that they are predominantly backward-looking. Moreover, the labor share does not appear to be a relevant determinant of inflation. We show that this is an important factor contributing to the weak identification of the Phillips curve.

Keywords: Weak instruments, Rational Expectations, GMM, robust inference

JEL classification: C22, E31

*I would like to thank Frank Kleibergen and the participants in seminars at Bonn and Boston Universities, the EC^2 conference, the CRETE conference and the Macroeconometrics conference in Halle for useful comments and discussion.

1 Introduction

The new Keynesian Phillips curve (NKPC) is a forward-looking model of inflation dynamics, according to which short-run dynamics in inflation are driven by the expected discounted stream of real marginal costs. Researchers often use a specification that includes both forward-looking and backward-looking dynamics (Buiter and Jewitt (1989), Fuhrer and Moore (1995), Galí and Gertler (1999)):

$$\pi_t = \lambda s_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \varepsilon_t \qquad (1)$$

where π_t denotes inflation, s_t is some proxy for marginal costs, E_t denotes mathematical expectation conditional on information up to time t, and ε_t is an unobserved innovation process, namely $E_{t-1}\varepsilon_t = 0$.

In a seminal paper, Galí and Gertler (1999) proposed a version of this model in which the forcing variable s_t is the labor share and the structural parameters $\lambda, \gamma_f, \gamma_b$ are functions of some deeper parameters: the fraction of backward-looking price-setters, the average duration an individual price is fixed (the degree of price stickiness) and a discount factor. Using postwar data on the U.S., Galí and Gertler (1999) reported that real marginal costs are statistically significant and inflation dynamics are predominantly forward-looking. They found γ_b to be statistically significant but quantitatively small relative to γ_f. In terms of their deep parameters, they reported that 60-80% of firms exhibited forward-looking behavior, and the average duration over which prices remain fixed was 6 to 7 quarters.

Mavroeidis (2005) argued that the above results are unreliable because the model appears to be weakly identified. He showed that the possibility of weak identification cannot be ruled out a priori, and that usual pre-tests of identification, such as the ones proposed by Cragg and Donald (1997) and Stock and Yogo (2003) are inappropriate in this context and can be misleading.

In this paper, we address the issue of identification, by re-evaluating the conclusions of Galí and Gertler (1999) using estimation and inference methods that are partially or fully robust to failure of identification. These procedures include the tests proposed by Stock and Wright (2000) and

Kleibergen (2005), which are applicable to Euler equation models estimated using the Generalized Method of Moments (GMM). Identification-robust tests are not yet available for full-information likelihood-based inference on this model. Therefore, we take a limited-information approach and use the continuously updated GMM estimator (CUE) proposed by Hansen (1996), instead of the most commonly used 2-step GMM estimator, as advocated by Stock, Wright, and Yogo (2002).

We use the same data as Galí and Gertler (1999). Our main findings are as follows. In accordance with Galí and Gertler (1999), we find some evidence that forward-looking dynamics in inflation are statistically significant at the 10% level. Unlike them, we find that postwar US inflation history is consistent both with a purely forward-looking Phillips curve as well as with a model in which the majority of firms are backward-looking. Moreover, we do not find strong evidence that real marginal costs drive inflation.

Galí and Gertler (1999) report that the sum of the backward and forward-looking coefficients is not significantly different from one. This restriction corresponds to the assumption that the discount factor in their model is known (and equal to one). Using identification-robust tests, we corroborate the above finding, but we also find that imposing that restriction does not improve the identifiability of the remaining parameters of the model.

The above results confirm the criticism of Mavroeidis (2005) regarding the poor identifiability of the parameters of the NKPC. They also help explain the large differences in empirical estimates reported by other researchers using alternative methods (Fuhrer (1997), Jondeau and Le Bihan (2003), Linde (2005)).

Ma (2002), Dufour, Khalaf, and Kichian (2006) and Nason and Smith (2005) also apply identification-robust methods to the NKPC, and report identification problems. The present study differs from those papers in several respects. First, it provides simulation evidence on the finite sample properties of the identification-robust statistics when applied to the NKPC. Second, in addition to the Anderson-Rubin and conditional score test used in the other papers, the present study makes use of the conditional likelihood ratio test, which is sometimes more powerful than the other identification-robust tests, see Andrews and Stock (2005). Third, the paper establishes some additional results on the identification of the NKPC which help understand the source of weak identification. Specifically, identification failure will arise when the labor share is not relevant as a forcing variable for inflation. Fourth, the paper examines a common restriction on the parameters of the NKPC (that the discount factor is known) and finds that it is insufficient to resolve the identification problem. Finally, the present study differs from the aforementioned papers also in terms of specific assumptions made on the model. Dufour, Khalaf, and Kichian (2006), impose the restriction that there is no error term in Eq. (1), i.e., that the NKPC is an "exact" rational expectations model in the language of Hansen and Sargent (1991). This restriction enables them to avoid using a heteroskedasticity and autocorrelation consistent (HAC) weighting matrix. One could think of this as a strict version of the NKPC, which is not the one typically encountered in applied work (see, e.g., Lubik and Schorfheide (2004) or Smets and Wouters (2003)). Moreover, it is relatively easy to test (and reject) the exact form of the model.[1] So, it is important to re-examine the conclusions on the NKPC without imposing that assumption. Nason and Smith (2005) do not impose exactness, and therefore use a HAC weighting matrix, but they impose the restriction that the labor share is exogenous (so the model has a single endogenous regressor). This restriction is also uncommon (Galí and Gertler (1999) do not impose it), and somewhat unrealistic since the labor share is a proxy for the true marginal costs, which are unobserved.

The structure of the paper is as follows. We begin by describing the model and the identification-robust methods in section 2. Before engaging in the empirical analysis, we re-examine the identification of model (1) in section 3, where we uncover another important case in which identification of the model fails. This happens when the candidate driving process s_t is unrelated to inflation, namely when $\lambda = 0$ in Eq. (1). This is of particular relevance because several studies (including that of Galí and Gertler (1999)) report estimates of λ that are very close to zero. In section 4, we present the empirical results, and in the subsequent section, we analyze the implications of the restriction $\gamma_f + \gamma_b = 1$ in Eq. (1).

[1] Exactness implies that the error $\pi_t - \lambda s_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1}$ should be serially uncorrelated. In fact, it is very significantly negatively autocorrelated at lag 1, which would occur only if $\epsilon_t \neq 0$ in Eq. (1).

Derivations are provided in Appendix A at the end. In Appendix A, we report the results of an extensive simulation study calibrated to the specifics of the NKPC for US data. The purpose of this study is to investigate the finite sample properties of the identification-robust tests and compare them to the Wald test. We find that the asymptotically robust tests have approximately correct size in all the cases that we considered, while the Wald tests can be severely over or under-sized: nominal 5%-level t tests reject the true null hypothesis more than 70% of the time in some cases, and do not reject at all in other cases. Moreover, the identification-robust tests do not waste power relative to the Wald test when the parameters are well-identified.

2 Methodology

The generic NKPC, Eq. (1), can be derived in a general equilibrium framework by log-linearizing the Euler equations of monopolistically competitive firms facing constraints in the adjustment of their prices, see Woodford (2003). The model of Galí and Gertler (1999) is a particular version of the NKPC that is derived from a model with Calvo (1983) frictions, in which a fraction of firms are backward-looking, in the sense that they do not adjust their prices in response to anticipated future deviations of marginal costs from their steady state value. Because of its prominence in the literature, we analyze this model in detail here.

The Galí and Gertler (1999) model has three structural parameters: ω is the fraction of backward-looking firms; θ is the probability that a firm will be unable to change its price in a given period, so that $1/(1-\theta)$ is the average time over which a price is fixed; and β is the discount factor. These parameters relate to λ, γ_f and γ_b in Eq. (1) as follows

$$\lambda = (1-\omega)(1-\theta)(1-\beta\theta)\phi^{-1} \tag{2}$$

$$\gamma_f = \beta\theta\phi^{-1} \tag{3}$$

$$\gamma_b = \omega\phi^{-1} \tag{4}$$

$$\phi = \theta + \omega\left[1-\theta(1-\beta)\right]. \tag{5}$$

Galí and Gertler (1999) refer to the parameters $(\lambda, \gamma_f, \gamma_b)$ in Eq. (1) as the reduced-form parameters, and (ω, θ, β) as the structural parameters. We will adopt this distinction hereafter.

Let ϑ denote the vector of parameters, including a constant.[2] Define the moment function

$$f_t(\vartheta) = Z_t(\pi_t - c - \lambda s_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1}) \tag{6}$$

where Z_t is a vector of k variables known at time $t-1$. The moment function can be equivalently expressed in terms of the structural parameters (ω, θ, β) using the expressions (2) through (4), as in Galí and Gertler (1999, p. 213, Equations 27 and 28).[3] The assumption of rational expectations implies that $Ef_t(\vartheta)$ vanishes at the true value of the parameters for any vector Z_t in the $t-1$ dated information set. In this application, Z_t will consist of the first four lags of π_t, s_t and four other variables used in the Galí and Gertler (1999) study.

Let $f_T(\vartheta) = \sum_{t=1}^{T} f_t(\vartheta)$ and $V_{ff}(\vartheta) = \lim_{T\to\infty} var\left[T^{-1/2} f_T(\vartheta)\right]$. The objective function for the CUE of Hansen, Heaton, and Yaron (1996) is

$$S(\vartheta) = T^{-1} f_T(\vartheta)' V_{ff}^{-1}(\vartheta) f_T(\vartheta) \tag{7}$$

and the CUE $\widehat{\vartheta}$ is the minimizer of $S(\vartheta)$ w.r.t. ϑ. The CUE is an alternative to the more traditionally used iterative GMM estimators, originally proposed by Hansen (1982). We use the CUE because it has been recently shown to have better finite sample properties under weak or many instruments, see Stock, Wright, and Yogo (2002) and Newey and Smith (2004).

Let $\mathcal{F}_t$ denote the nondecreasing information set available at time t, which is adapted to the sequence $\{\pi_t, \pi_{t-1}, \ldots; s_t, s_{t-1}, \ldots; \varepsilon_t, \varepsilon_{t-1}, \ldots\}$. Due to the presence of π_{t+1} in Eq. (6), the stochastic process $f_t(\vartheta)$ is not measurable w.r.t. $\mathcal{F}_t$ but rather to $\mathcal{F}_{t+1}$. This means that $E_{t-1} f_t(\vartheta) f_{t-1}(\vartheta) \neq 0$ in general, where $E_t(\cdot) \equiv E(\cdot|\mathcal{F}_t)$, so the moment functions $f_t(\vartheta)$ can exhibit first-order autocorrelation without contradicting the model.[4]

[2] For simplicity, we omit the constant in the ensuing discussion.

[3] Galí and Gertler (1999) discuss alternative normalizations of the moment conditions, because iterative GMM methods are not invariant to parameter transformations. However, the CUE is invariant..

[4] An exception occurs when $\gamma_f = 0$, in which case the model implies that $f_t(\vartheta)$ should be serially independent. This is relevant only when testing the null hypothesis $\gamma_f = 0$.

To operationalize (7) we need a heteroskedasticity and autocorrelation consistent (HAC) estimator of $V_{ff}(\vartheta)$. Popular choices of HAC estimators are those proposed by Newey and West (1987) and Andrews (1991), though it is well-known that they often result in large size distortions in finite samples, see den Haan and Levin (1997). A more efficient and potentially more reliable choice is the parametric MA-l estimator of West (1997), which exploits the first-order moving average pattern of dependence in the moment function $f_t(\vartheta)$.[5]

Identification of ϑ requires that $Ef_t(\vartheta) = 0$ if and only if $\vartheta = \vartheta_0$. This is a necessary condition for any estimator $\widehat{\vartheta}$ to be consistent and asymptotically normal. Our objective here is to do inference without imposing that assumption. In the remainder of this section, we briefly describe the inference procedures used in this paper, and explain why they are robust to potential failure of identification.

2.1 Identification-robust inference

Stock and Wright (2000) showed that if a central limit theorem applies to the moment function $f_T(\vartheta)$, then the asymptotic distribution of the objective function $S(\vartheta)$ evaluated at the true value of ϑ, is χ^2 with degrees of freedom equal to the number of instruments, k. This result requires no identification assumptions, and is a generalization of the Anderson and Rubin (1949) test. Hypotheses on ϑ can be tested at the α level of significance by comparing $S(\vartheta)$ to the $(1-\alpha)$ quantile of the $\chi^2(k)$ distribution. We will refer to this test as the AR test.

One weakness of the AR test is that it may have low power relative to the usual Wald, LR or LM tests when identification is strong. Therefore, use of the AR test incurs a cost that reflects the usual trade-off between efficiency and robustness. Another weakness of the AR test is in its interpretation: a rejection may reflect either the violation of the overidentifying restrictions or evidence against the particular null hypothesis on ϑ.

Recently, Kleibergen (2002) and Moreira (2003) developed testing procedures that overcome

[5] Monte Carlo evidence reported in West and Wilcox (1996) suggests the MA-l outperforms nonparametric alternatives in Euler equations models. See also Mavroeidis (2005) for simulations based on the NKPC.

the above two weaknesses of the AR test in the context of the linear IV regression model. In this study, we will apply the GMM versions of the conditional score (KLM) test and the conditional likelihood ratio (CLR) test developed by Kleibergen (2005). Here we just give a brief explanation of how these methods work. Stock, Wright, and Yogo (2002) and Andrews and Stock (2005) provide excellent reviews of those methods.

Consider first the KLM statistic. Let B_T be a sequence of stochastic matrices of dimensions $k \times m$ (where m is the number of parameters, $m \leq k$) that are asymptotically independent of the sample moment conditions $f_T(\vartheta)$ and converge in probability to a (possibly stochastic) matrix B of full rank m. Then, the m-vector $T^{-1/2} B_T' f_T(\vartheta)$ is asymptotically normal with variance $p\lim_{T\to\infty} (B_T' V_{ff}(\vartheta) B_T)$, and consequently, the quadratic form $T^{-1} f_T(\vartheta)' B_T (B_T' V_{ff}(\vartheta) B_T)^{-1} B_T' f_T(\vartheta)$ is $\chi^2(m)$. Kleibergen (2005) showed that a natural choice of B_T is the first derivative of the objective function (7), $D_T(\vartheta)' V_{ff}(\vartheta)^{-1} f_T(\vartheta)$, where $D_T(\vartheta)$ depends on $\partial f_T(\vartheta)/\partial\vartheta$ and $\partial V_{ff}(\vartheta)/\partial\vartheta$, which results in a score test statistic, see Kleibergen (2005, Eq. (16)). This statistic requires smoothness of the objective function, which applies to the NKPC, since $f_T(\vartheta)$ is linear and $V_{ff}(\vartheta)$ is quadratic in ϑ. The resulting quadratic form is called the KLM statistic and the test based on comparing $KLM(\vartheta)$ to critical values of the $\chi^2(m)$ distribution is the KLM test.

A weakness of the KLM test is that it may suffer from lack of power against alternatives that are close to points of inflection of the CUE objective function (7). This occurs because the derivative of $S(\vartheta)$ (and hence the KLM statistic) is zero at all those points. To overcome this weakness, Kleibergen suggests using the JKLM statistic defined as $JKLM(\vartheta) = AR(\vartheta) - KLM(\vartheta)$ which is asymptotically $\chi^2(k-m)$ and independent of the KLM statistic. The JKLM statistic is interpreted as testing the overidentifying restrictions when the true value of the parameters is ϑ and can be used to provide an upper bound on the GMM objective function. Thus, an approximately 10%-level test can be constructed by first pre-testing whether the JKLM test rejects at the 1%-level, and then, provided it does not reject, using a 9%-level KLM test, see Kleibergen (2005, section 5).

The final identification robust procedure that we shall employ is the GMM version of Moreira's (2003) CLR test, proposed in Kleibergen (2005, section 5.1). This statistic is also a function of the KLM and JKLM statistics, but this function is random and depends on a statistic that tests the rank condition for identification (i.e., the rank of the Jacobian of the moment conditions). We implement the test using the rank test statistic of Kleibergen and Paap (2006). The asymptotic distribution of the CLR statistic, conditional on the rank test statistic, is independent of any nuisance parameters. This conditional distribution is not analytically available but p-values can be derived by simulation to any desired degree of accuracy (we use 10^5 replications). In the linear IV regression model with homoscedastic and serially uncorrelated innovations, Andrews, Moreira, and Stock (2006) show that the CLR test enjoys certain optimality properties relative to the AR and KLM tests. However, these optimality results have not been established in more general GMM settings, and do not, therefore, apply in the study of the NKPC. Here, we report some evidence regarding the size and power of those tests in the context of the NKPC derived by means of Monte Carlo simulation. Despite no apparent power advantages, we use all three methods in our empirical analysis of the NKPC.

The KLM and CLR tests do not require any identification assumption for ϑ. However, they do require some additional regularity conditions relative to the AR test. These conditions refer to the limiting behavior of the Jacobian of the moment conditions $\partial f_T(\vartheta)/\partial\vartheta'$, which in our model is (proportional to) the covariance between the regressors and the instruments. Moreover, we need a consistent estimator of the asymptotic covariance between $f_T(\vartheta)$ and $\partial f_T(\vartheta)/\partial\vartheta$, $V_{f\vartheta}(\vartheta)$ as well as the variance of the latter. Note that, even though $f_t(\vartheta)$ is MA(1), the pattern of serial dependence in $\partial f_t(\vartheta)/\partial\vartheta$ is unrestricted by our model. This means that West's (1997) MA-l estimator is inappropriate for the KLM and CLR tests in this study, and we need to use an unrestricted HAC estimator, such as Newey and West (1987), see Kleibergen (2005, Section 4).[6]

[6] We computed all tests and confidence sets also using West's MA-l instead of the Newey and West (1987) weighting matrix. The p-values were always higher than the ones associated with Newey and West (1987), and the confidence sets were wider. The results are omitted for brevity, but are available on request.

2.2 Testing composite hypotheses

We are mainly interested in testing hypotheses on subsets of the parameters in (1) leaving the remaining parameters unrestricted. In particular, we are interested in testing the following hypotheses: $\lambda = 0$, $\gamma_f = 0$, $\gamma_f = \gamma_b$, and $\gamma_f + \gamma_b = 1$.

To perform α-level tests on such hypotheses, we can follow two alternative approaches. One approach is to construct joint AR, KLM and CLR tests on all parameters, and then choose the values of the unrestricted parameters that minimize the test statistics. For instance, when testing $\lambda = \lambda_0$, compute $\tau = \min_{\gamma_f, \gamma_b} AR(\lambda_0, \gamma_f, \gamma_b)$. The asymptotic distribution of this statistic is bounded above by a $\chi^2(k)$, so if we compare the statistic τ to the $1 - \alpha$ quantile of $\chi^2(k)$, the resulting test will have a size that is at most α under the null. This is the projection method, discussed in detail in Dufour (2003).

The disadvantage of the projection method is that it wastes power when it is known that certain parameters are well-identified under the null. This is true of γ_b when λ and γ_f are fixed, since γ_b can be recovered from a regression of $\pi_t - \gamma_f \pi_{t+1} - \lambda s_t$ on π_{t-1}. In that case, it is preferable to partial out the identified parameters by concentrating the objective function with respect to them, i.e., by deriving the restricted CUE $\widehat{\vartheta}_0$, say, and evaluating all the test statistics at $\widehat{\vartheta}_0$. The resulting tests are sometimes referred to as subset tests. Stock and Wright (2000) and Kleibergen (2005) derived the distribution of the subset tests under the assumption that the unrestricted parameters are well-identified. This assumption is not always plausible (see below for the null hypothesis $\lambda = 0$). In the linear IV regression model, Kleibergen (2007) showed that the distribution of the subset AR, KLM and CLR statistics is bounded from above by the asymptotic distribution that arises when the unrestricted parameters are well-identified, otherwise, these subset tests become conservative. This result has not yet been extended to GMM, but simulation evidence reported in Appendix B shows it may apply to the present model. In the interest of maximizing power, and subject to this caveat, we chose to use subset tests instead of projection methods in this study. Note that the latter will always produce wider confidence sets that the former. Therefore, the main conclusions of this paper that the key structural parameters of the NKPC are not well-identified, would remain unchanged had we used projection methods instead.

We shall also report two-dimensional confidence sets for various parameters of the model. A $(1-\alpha)$-level confidence set is a random set that contains the true value of the parameter with probability at least $1-\alpha$. $(1-\alpha)$-level confidence sets can be derived by inverting α-level tests, that is, finding all the points ϑ in the parameter space such that a given test τ does not reject the hypothesis that ϑ is the true value at the α-level of significance. Wald-based confidence sets are elliptical by construction, but they do not have correct coverage when identification is weak. In contrast, confidence sets based on inverting the AR, KLM and CLR tests have correct coverage asymptotically. Except in the special case of the Wald test, confidence sets derived by inverting tests can be asymmetric and disjoint (e.g., when the objective function has multiple local minima).

3 Some new results on the identification of the new Phillips curve

3.1 What happens when $\lambda = 0$

Mavroeidis (2005) showed that the identification of the parameters $(\lambda, \gamma_f, \gamma_b)$ depends on the dynamics of the forcing variable. Here, we focus on another source of weak identification: weak association between the forcing variable s_t and inflation.

Suppose we are interested in testing the null hypothesis $\lambda = 0$. The interpretation of that hypothesis in the Galí and Gertler (1999) model is that real marginal costs s_t are unimportant as determinants of inflation. In other applications, one may wish to test whether some other measure, e.g. GDP growth or output gap, is the relevant forcing variable for inflation. We will show that this hypothesis cannot be tested by a usual t-test, because the t statistic does not have an asymptotically normal distribution under the null.

Under the null hypothesis, the model (1) becomes

$$\pi_t = \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \varepsilon_t, \tag{8}$$

so inflation is driven solely by a sequence of innovations ε_t that are unobserved by the econometrician. The dynamics of inflation, under the null, depend on whether the rational expectations model (8) admits a determinate or an indeterminate solution. By a solution, we mean a stochastic process $\{\pi_t\}$ that satisfies equation (8) and is not explosive, i.e., satisfies the following limiting condition $\lim_{t\to\infty} \varsigma^{-t} E_s \pi_t = 0$ for all $|\varsigma| > 1$. Blanchard and Kahn (1980) showed that existence and uniqueness of such a solution depends on the roots of the characteristic polynomial

$$1 - \gamma_f z^{-1} - \gamma_b z = 0. \tag{9}$$

Existence requires that at most one root of this polynomial is inside the unit circle.

A determinate solution arises when the characteristic polynomial (9) has exactly one root inside the unit circle. The determinate solution can be represented as a first-order autoregression [AR(1)]

$$\pi_t = \delta \pi_{t-1} + \frac{1}{1 - \delta \gamma_f} \varepsilon_t \tag{10}$$

where the autocorrelation coefficient δ is the inverse of the largest root of (9), see Appendix A.

If none of the roots is inside the unit circle, the solution is indeterminate, and can be characterized by

$$\pi_t = \frac{1}{\gamma_f} \pi_{t-1} - \frac{\gamma_b}{\gamma_f} \pi_{t-2} - \frac{1}{\gamma_f} \varepsilon_{t-1} + \eta_t \tag{11}$$

where the one-step ahead forecast error in inflation $\eta_t = \pi_t - E_{t-1}\pi_t$ is some indeterminate martingale difference sequence. Equation (11) can be represented as an Autoregressive Moving Average process, denoted ARMA(2,1), see Pesaran (1987). It can also be represented by the equations

$$\pi_t = \delta_1 \pi_{t-1} + \frac{1}{1 - \delta_1 \gamma_f} \varepsilon_t + \xi_t \tag{12}$$

$$\xi_t = \delta_2 \xi_{t-1} + \eta_t - \frac{1}{1 - \delta_1 \gamma_f} \varepsilon_t \tag{13}$$

where ξ_t is an autoregressive process that is unobserved by the econometrician, and δ_1, δ_2 are the inverses of the roots of (9), see Appendix A. The martingale difference sequence $\eta_t - \varepsilon_t / (1 - \delta_1 \gamma_f)$ is often called a sunspot shock, and the indeterminate solution is said to exhibit sunspot dynamics whenever the sunspot shock (or equivalently ξ_t) is not identically equal to zero for all t.

When inflation exhibits only first order autoregressive dynamics, the parameters γ_f and γ_b in the model (8) are not separately identifiable, because all the combinations of γ_f and γ_b that yield the same autocorrelation coefficient are observationally equivalent. In this case, we say that γ_f and γ_b are partially identified. This occurs either when the solution is determinate (10) or when it is indeterminate, but inflation does not exhibit sunspot dynamics, i.e., follows Eq. (12) with $\xi_t \equiv 0$.

Although alternative models may differ in the way in which they specify price rigidities and partial adjustment, a common feature of many such models (Buiter and Jewitt (1989), Galí and Gertler (1999)) is that the parameters γ_f and γ_b satisfy the restrictions $\gamma_f, \gamma_b \geq 0$, $\gamma_f + \gamma_b \leq 1$ and $\lambda \geq 0$. We show in Appendix A that unless $\gamma_f + \gamma_b = 1$ and $\gamma_f > \gamma_b$, these restrictions imply that the solution of (8) is determinate and, therefore, γ_f, γ_b are only partially identified when $\lambda = 0$. In the special case $\gamma_f + \gamma_b = 1$ and $\gamma_f > \gamma_b$, γ_f and γ_b are identified when inflation exhibits sunspot dynamics.

When interest centers on the composite null hypothesis $\lambda = 0$, the remaining parameters of the model, γ_f and γ_b, can be viewed as 'nuisance parameters'. The usual t-statistic for this hypothesis is the ratio of the parameter estimator $\hat{\lambda}$ to the estimator of its standard-error $\hat{\sigma}_{\hat{\lambda}}$. If all the parameters of the model were identified, the asymptotic distribution of $\hat{\lambda}/\hat{\sigma}_{\hat{\lambda}}$ would be standard normal (under standard regularity conditions), and one could interpret the t-statistics in the usual way. However, when either λ or the nuisance parameters γ_f, γ_b are not identified under the null, then the t-statistic does not have an asymptotically normal distribution, and tests based on standard normal critical values could be very misleading. We summarize the above discussion in the following proposition.

Proposition 1 *When $\lambda = 0$, the parameters γ_f and γ_b of the NKPC (1) are partially identified, unless $\gamma_f + \gamma_b = 1$, $\gamma_f > \gamma_b$ and inflation exhibits sunspot dynamics. Hence, the t-statistic for the hypothesis $H_0 : \lambda = 0$ does not have an asymptotically normal distribution under the null, and is therefore not interpretable in the usual way.*

The inferential problem that arises when a nuisance parameter is not identified under the null is well-known in econometrics. This problem was studied by Andrews and Ploberger (1994) and

Hansen (1996) in the context of nonlinear regression models, but their proposed methods are not applicable to the NKPC. However, we can still perform valid, albeit conservative, tests using the subset AR KLM and CLR tests described in the previous section.

3.2 Identification of the structural version of the NKPC

Given their interpretation, the deep parameters of the Galí and Gertler (1999) model (ω, θ, β) must lie in the unit cube. The compactness of the parameter space of (ω, θ, β) is computationally attractive, since confidence sets can be feasibly derived by grid search over the entire parameter space, with a reasonable degree of precision.

There are two identification issues that need to be pointed out. First, as we show in the Appendix, the mapping from (ω, θ, β) to $(\lambda, \gamma_f, \gamma_b)$ given by equations (2), (3) and (4) is not generally invertible. This means that the structural parameters (ω, θ, β) are not globally identified, and so estimation and inference on (ω, θ, β) could be problematic, even when $(\lambda, \gamma_f, \gamma_b)$ are well-identified.

Second, there are regions in the admissible parameter space of (ω, θ, β) in which those parameters become *locally* unidentified. We will refer to these as partial identification regions. These correspond to three limiting cases. At $\theta = 0$ (i.e., in the absence of frictions), the parameter β is not identified, since it only appears in the model through the product $\theta\beta$. At $\theta = 1$, i.e., when prices are fixed forever, $\lambda = 0$ and hence ω and β are not separately identified for the same reasons as those given in proposition 1. At $\omega = 1$, i.e., when all firms are backward-looking, neither θ nor β are identified, since inflation follows a random walk whose distribution is independent of θ and β. We summarize the above discussion in the following proposition.

Proposition 2 *The parameters (ω, θ, β) in the structural NKPC model of Galí and Gertler (1999) are not globally identified. Moreover, they are locally unidentified in the following three limiting cases: (i) when $\theta = 0$; (ii) when $\theta = 1$ and (iii) when $\omega = 1$.*

Even if the aforementioned limiting cases are considered implausible, they do provide useful insights into possible sources of identification problems. In particular, if the true values of (ω, θ, β) lie 'close' to those partial identification regions, the model will be weakly identified. How far they need to be in order for the model to be well-identified depends on other aspects of the data generating process, such as the dynamics of the forcing variable. One could characterize the regions in the parameter space where identification might be weak using some measure of the strength of identification, e.g., the concentration parameter which is a measure of the correlation between instruments and endogenous regressors, see Stock, Wright, and Yogo (2002). To do this, one would need to specify the law of motion of the forcing variables s_t, as was done in Mavroeidis (2005).

The objective of the present study is not to assess the identifiability of the NKPC, but rather to do inference on its parameters without assuming identification. As already explained in the previous section, this can be done without taking a stance on whether the model is well-identified or not, using tests that are robust to weak identification. One may wish to interpret our results as indirect evidence on the identifiability of the NKPC. This could be done by comparing the conclusions drawn from robust and non-robust tests. If the model is well-identified, any observed differences would be due to sampling variation, and thus, they should be small.

4 New estimates of the new Phillips curve

Table 1 reports estimates of the parameters in the two aforementioned specifications of the NKPC, using two sets of instruments. The small set of instruments includes only four lags of inflation and the labor share, while the extended set includes all the instruments used by Galí and Gertler (1999).

We also report the Hansen-Sargan test of over-identifying restrictions, as well as a test of excess serial correlation in the model's residuals $u_t = \pi_t - c - \lambda s_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1}$. Since u_t is MA(1) under the null of correct specification, we test against higher order autocorrelation using the test of Cumby and Huizinga (1992).[7] There is no evidence that the over-identifying restrictions are

[7] Details of the implementation are given in Mavroeidis (2002).

Table 1: Estimates of the new hybrid Phillips curve

	reduced-form param.			deep param.			diagnostics	
	λ	γ_f	γ_b	ω	θ	β	OR	SC
Instruments								
small set	0.132	0.850	0.144	0.112	0.663	0.992	5.605	10.539
	(0.073)	(0.212)	(0.205)	(0.180)	(0.083)	(0.035)	[0.347]	[0.032]
large set	0.126	0.690	0.281	0.246	0.635	0.948	27.899	5.147
	(0.050)	(0.035)	(0.034)	(0.043)	(0.051)	(0.028)	[0.143]	[0.273]

The model is $E[Z_t(\pi_t - c - \lambda s_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1})] = 0$. Instruments include 4 lags of π_t, s_t (small set), plus commodity price and wage inflation, output gap and long-short yield spread (large set). CUE-GMM with Newey and West (1987) Weight matrix, bandwidth: 4. Sample: 1960 (1) - 1997 (4). Diagnostics: OR is Hansen-Sargan test of overidentifying restrictions, $\chi^2(k-4)$; SC is Cumby and Huizinga (1992) test of residual autocorrelation from lags 2 to 5, $\chi^2(4)$.

violated at the conventional 5%. There is some evidence of excess serial correlation in the residuals when using the small instrument set, but this is not robust to extensions of the instrument set. Unreported plots of the residual correlogram do not show substantial autocorrelation beyond lag 1.

Turning to the parameter estimates, we note that they are broadly in line with the results reported by Galí and Gertler (1999).[8] As found by Galí and Gertler (1999), the results suggest that forward dynamics dominate backward dynamics. In terms of specification 2, the fraction of backward-looking agents is close to 0, albeit statistically significant according to the t-test.

4.1 Tests of various hypotheses

Table 2 reports tests of various hypotheses on the reduced form parameters of the NKPC, see Eq. (1). For each hypothesis we report the p-values associated with the Wald, AR, CLR, KLM and JKLM statistics. We report results both using own lags of inflation and the share, and using

[8] They use a 2-step GMM procedure, and they consider alternative normalizations of the moment conditions. We do not need to do that here, because the CUE is invariant to re-normalization.

Table 2: Hypothesis tests in the reduced form specification of the new hybrid Phillips curve

Hypothesis	Instruments	Test p-values				
		Wald	AR	CLR	K	J
λ=0	small set	0.07	0.19	0.55	0.08	0.34
	large set	0.01	0.05	0.30	0.05	0.09
γ_f=0	small set	0.00	0.10	0.37	0.61	0.07
	large set	0.00	0.10	0.34	0.76	0.08
γ_b=0	small set	0.48	0.41	0.95	0.37	0.38
	large set	0.00	0.06	0.12	0.16	0.07
$\gamma_f + \gamma_b = 1$	small set	0.82	0.46	1.00	0.80	0.34
	large set	0.08	0.09	0.51	0.10	0.12

The model is $E[Z_t(\pi_t - c - \lambda s_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1})] = 0$. Instruments include 4 lags of π_t, s_t (small set), plus commodity price and wage inflation, output gap and long-short yield spread (large set). CUE-GMM with Newey-West Weight matrix, bandwidth: 4. Sample: 1960 (1) - 1997 (4).

additional variables as instruments.

We consider first the null hypothesis $\lambda = 0$. By proposition 1, we note that the Wald test is inappropriate, and we expect the identification-robust tests to be potentially conservative. Using the small instrument set the AR test does not reject at the 19% level or higher, but the KLM test rejects at the 10% level. The evidence that the labor share is the relevant forcing variable for inflation becomes stronger when we use more instruments, but λ is still barely significantly different from 0 at the 5% level.

Next, we turn to the hypothesis $\gamma_f = 0$, which received considerable attention in the literature. The Wald test here is simply the square of the usual t-test on γ_f and its p-value is 0 to three decimals, suggesting overwhelming evidence against the null. However, none of the identification-robust tests reject at the 10% level. Note that under this null hypothesis $\gamma_f = 0$, the moment conditions must not exhibit any serial correlation, so a valid AR test can be performed without using a HAC weighting matrix. When we use Eicker-White instead, the p-value drops to below 5% but still above 1%. This version of the AR provides some evidence against the view that inflation dynamics are purely backward-looking. But note that this is far weaker than the conclusions drawn from the Wald test.

The evidence on $\gamma_b = 0$ is mixed. Using the small instrument set, none of the tests rejects, but with more instruments there is some evidence against the null, albeit very weak. Unlike Galí and Gertler (1999), our results suggest that the data is consistent with a purely forward-looking Phillips curve.

Finally, the hypothesis $\gamma_f + \gamma_b = 1$ is not rejected by any of the tests. Indeed, we also notice that the Wald test is very similar to the KLM test in this case. This is sometimes seen as an indication that a parameter is well-identified. We will see more evidence of that in two-dimensional confidence sets below.

4.2 Confidence sets

We compute two-dimensional identification-robust 10%-level confidence sets by inverting the AR KLM and CLR tests. (KLM confidence sets are based on a combination of a 1% JKLM pretest and a 9% KLM test.)

Reduced-form specification

To examine the relative importance of forward versus backward-looking adjustment, we consider a two-dimensional confidence set for γ_f, γ_b. We consider only the parameter region $0 \leq \gamma_f, \gamma_b \leq 1$ and $\gamma_f + \gamma_b \leq 1$. Thus, the reported confidence sets contain the values of γ_f and γ_b that are consistent both with the theoretical model and with the data at the given level of significance.

Figure 1 shows AR-based 90%-level sets, where λ has been partialled out. The set on the left-hand side uses the small instrument set (four lags of inflation and the share) while the set on the right uses all of Gali and Gertler's (1999) instruments. Comparison of these two graphs suggests that the additional instruments are informative. In both cases, it appears that there is little information in the data about γ_f. Even though the data are more informative about γ_b, they are consistent both with the hypothesis $\gamma_f > \gamma_b$ and with $\gamma_f < \gamma_b$ at the 10% level of significance.

Large AR-based confidence sets are not necessarily an indication of weak identification, due to the potential lack of power of the AR test when the number of over-identifying restrictions is large.

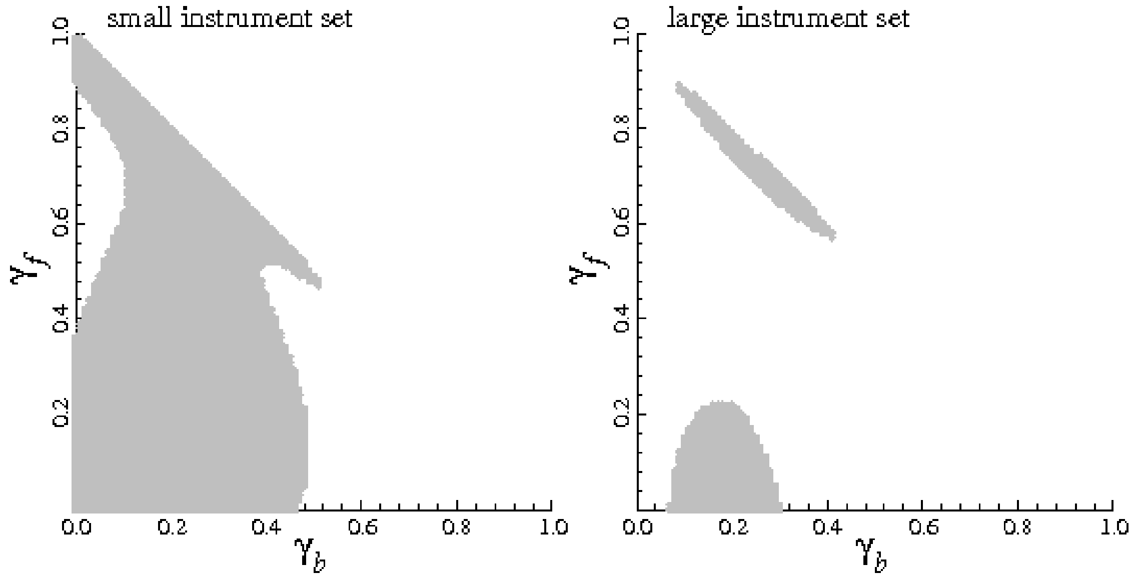

Figure 1: Joint 90%-level confidence sets for the parameters γ_f, γ_b of the NKPC $\pi_t = \lambda s_t + \gamma_f E_t(\pi_{t+1}) + \gamma_b \pi_{t-1} + \epsilon_t$. Instruments include 4 lags of π_t, s_t (small set), plus commodity price and wage inflation, output gap and long-short yield spread (large set). λ is partialled out.

On the other hand, AR-based sets can be empty if the over-identifying restrictions are violated even when the parameters are unidentified (Stock and Wright (2000)). In the present model, a tight AR set could reflect near violation of the orthogonality conditions, which could be interpreted as evidence against the assumption of rational expectations.

To shed further light on those issues, we now turn to the confidence sets derived by inverting the KLM test and CLR tests. Figure 2 plots these two 90%-level confidence sets on (γ_f, γ_b) based on the large instrument set.[9] We see that both the KLM and the CLR confidence sets are very similar, and much larger than the corresponding AR confidence set (see right-hand panel of Figure 1). This suggests that the large AR confidence sets reported earlier are not due to lack of power of the AR test. According to these pictures the forward-looking coefficient is completely unidentified. Moreover, we cannot rule out either that forward-looking dynamics dominate ($\gamma_f > \gamma_b$), or the opposite ($\gamma_f < \gamma_b$).

The last confidence set reported in Figure 2 is derived by inverting the JKLM test, which

[9]The KLM and CLR confidence sets corresponding to the smaller instrument set are much wider, and are omitted for brevity.

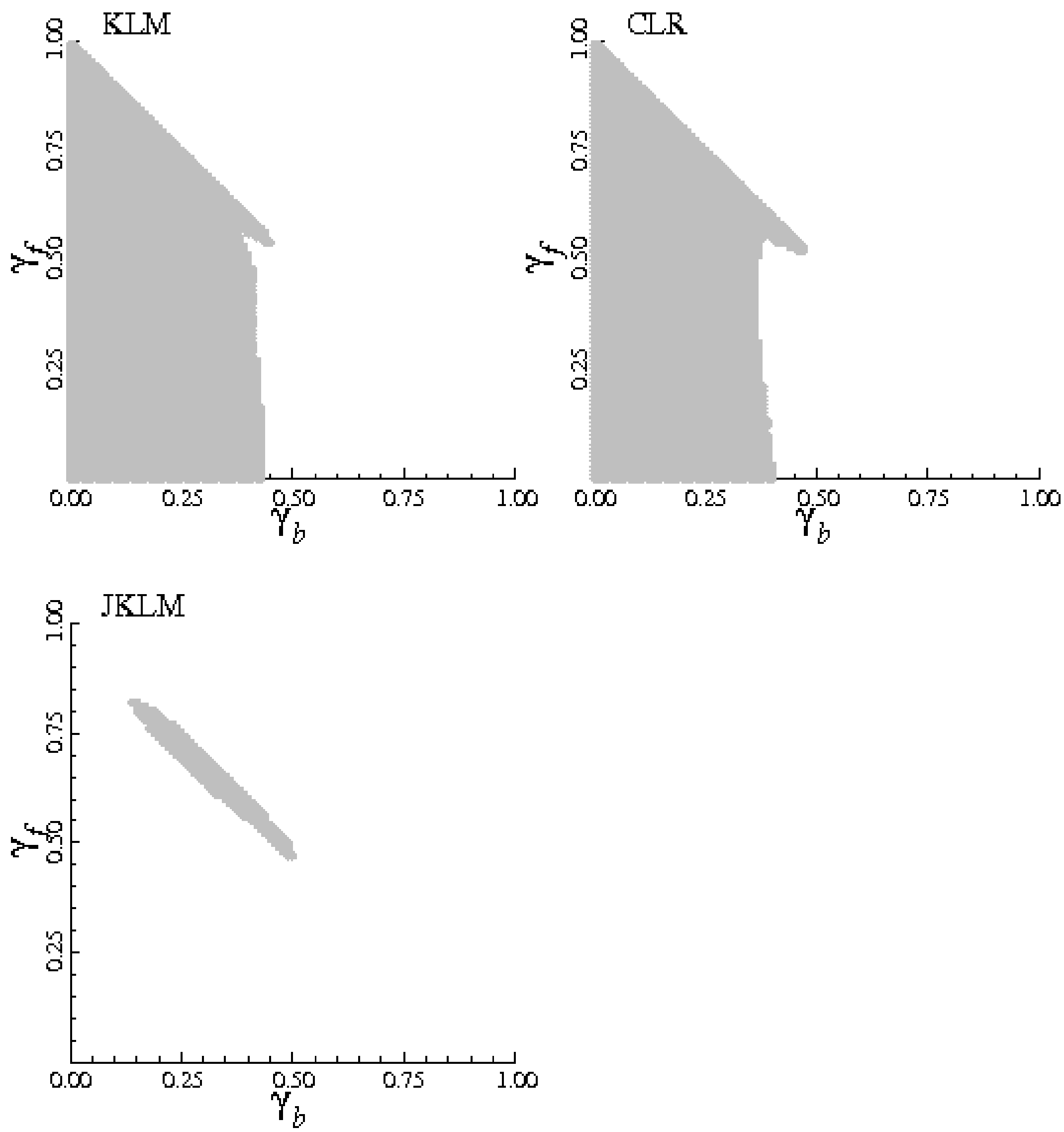

Figure 2: Joint 90%-level confidence sets for the parameters γ_f, γ_b of the NKPC $\pi_t = \lambda s_t + \gamma_f E_t(\pi_{t+1}) + \gamma_b \pi_{t-1} + \epsilon_t$ derived by inverting the KLM CLR and JKLM tests. Instruments include 4 lags of π_t, s_t, commodity price and wage inflation, output gap and long-short yield spread. λ is partialled out.

has power against violation of the overidentifying restrictions. The fact that this set is smaller than all the others suggests that it is near violation of the over-identifying restrictions that is the most important source of information in these data.[10] In this smallest 90%-level confidence set, γ_f lies between 0.47 and 0.82, and γ_b lies between 0.14 and 0.5. This implies that backward-looking dynamics are statistically significant at the 10% level, and we cannot rule out the possibility that $\gamma_f < \gamma_b$. It is important to keep in mind that this apparent identification comes primarily from the over-identifying restrictions.

In sum, the above results show that the data are consistent both with a pure forward-looking Phillips curve as well as with a hybrid Phillips curve that puts most weight on backward-looking dynamics. This finding reconciles the conflicting results reported by different researchers on the relative importance of forward versus backward-looking adjustment.

Structural specification

Next, we consider the structural specification of Galí and Gertler (1999). Figure 3 presents for alternative joint 90%-level confidence sets for ω and θ, partialling out β. Also drawn in the figures are confidence ellipses based on the Wald test, which are not robust to weak instruments.

The striking difference between the Wald-based confidence ellipses and the identification-robust confidence sets is suggestive of weak identification. The confidence sets most comparable, in terms of power and interpretation, to the Wald confidence ellipse are the KLM and CLR ones. In sharp contrast to the Wald ellipse, these sets include the entire parameter space of θ, suggesting that this parameter is completely unidentified. While ω is better identified, we still cannot rule out the possibility that backward-looking price setting is dominant ($\omega > 1/2$).

The AR-based confidence set is tighter but noticeably disjoint, and still wider than the Wald ellipse. Based on the AR confidence set, one could conclude with 90% confidence that the fraction of backward-looking agents is less than a half ($\omega < 0.45$), and prices remain fixed between 1.2 and 1.6 quarters ($\theta \in [0.17, 0.36]$) or 1.8 and 5.5 quarters ($\theta \in [0.44, 0.85]$).

[10] If the parameters γ_f and γ_b were well-identified, one would expect to see the opposite pattern, namely, the KLM and CLR sets should be tighter, since, at least in theory, these tests are more powerful than the AR and JKLM tests.

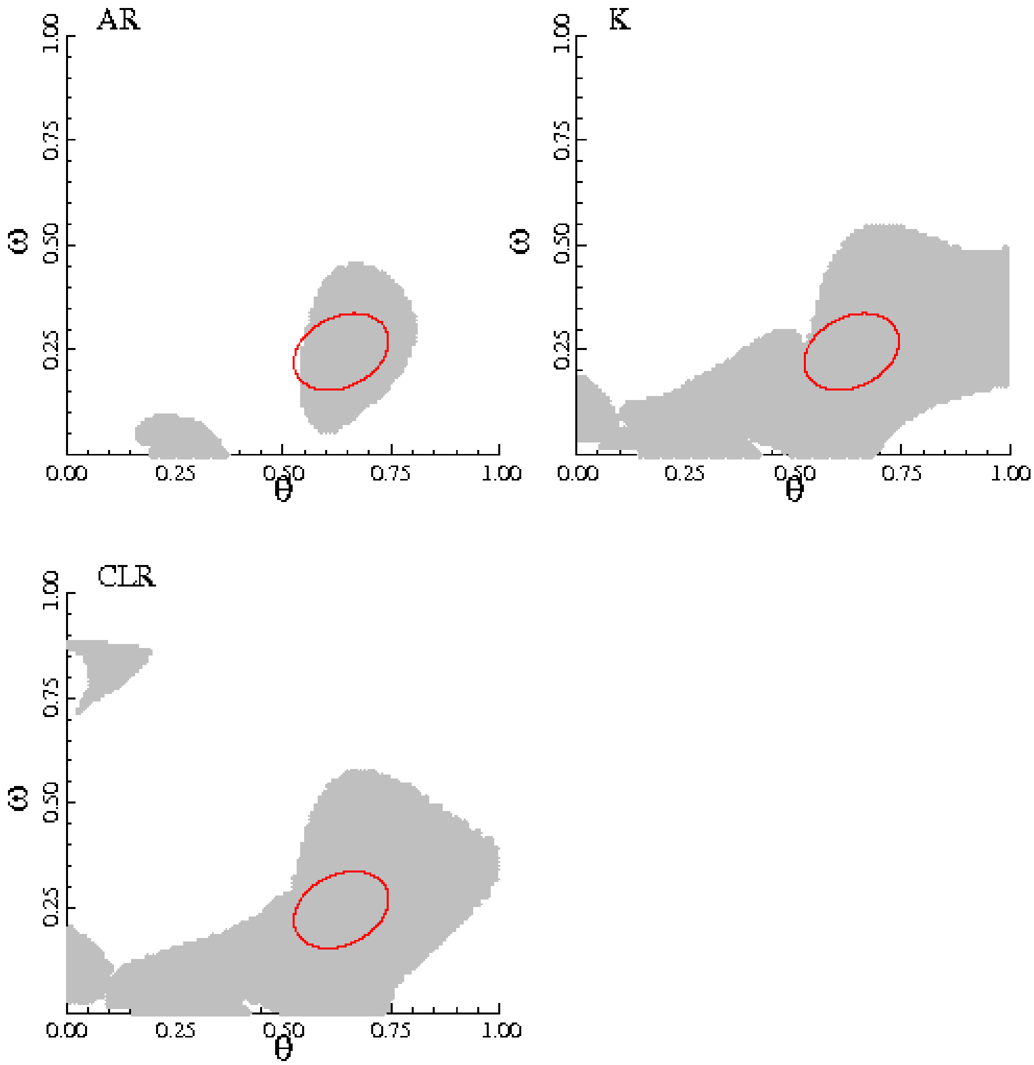

Figure 3: Joint 90%-level confidence sets for fraction of backward-looking firms ω and the probability prices remain fixed θ of the structural specification of the NKPC. Instruments include 4 lags of π_t, s_t, commodity price and wage inflation, output gap and long-short yield spread. The discount factor β has been partialled out.

5 Implications of the restriction $\beta = 1$

Early versions of the hybrid NKPC (Buiter and Jewitt (1989)) imposed the restriction that the backward and forward-looking coefficients sum to 1. As we saw earlier, this restriction is indeed not rejected by the data using identification-robust tests. Since it is reducing the number of estimable parameters, this restriction is essentially freeing up one instrument (lagged inflation), and can potentially improve the identifiability of the remaining structural parameters. It is therefore worth examining whether this is in fact the case.

The restriction $\gamma_f + \gamma_b = 1$ corresponds to $\beta = 1$ or $\theta = 0$ or $\omega = 1$ in the structural parametrization of Galí and Gertler (1999). In line with Galí and Gertler (1999), we will consider the restriction $\beta = 1$, both because it is the most appealing one given the interpretation of the parameters, and because both $\theta = 0$ and $\omega = 1$ would lead to the identification problems described in proposition 2. When $\gamma_f + \gamma_b = 1$, the model (1) becomes

$$(1 - \gamma_f)\,\Delta\pi_t = \lambda s_t + \gamma_f E_t \Delta\pi_{t+1} + \varepsilon_t.$$

Substituting for γ_f and λ using (3) and (2) when $\beta = 1$, the restricted model can also be written as

$$\omega\Delta\pi_t = (1 - \omega)(1 - \theta)^2 s_t + \theta E_t \Delta\pi_{t+1} + (\omega + \theta)\varepsilon_t. \tag{14}$$

For consistency, we also need to use lagged *changes* in inflation rather than levels of inflation as instruments.[11] Table 3 reports estimates of the parameters in the two specifications of the NKPC. Upon comparison with the unrestricted estimates reported in Table 1, we note that the qualitative results on λ, γ_f or ω, θ remain roughly unchanged when the restriction $\gamma_f + \gamma_b = 1$ is imposed.

Next we investigate the implication of the restriction $\beta = 1$ for the identification of the structural parameters of the model, ω and θ. Figure 4 presents 90%-level confidence sets for the structural parameters ω and θ. These should be compared to the confidence sets in the unrestricted specification reported in Figure 3. When we compare the AR-based sets, we find they are somewhat

[11] This is also done in the recent study by Rudd and Whelan (2006).

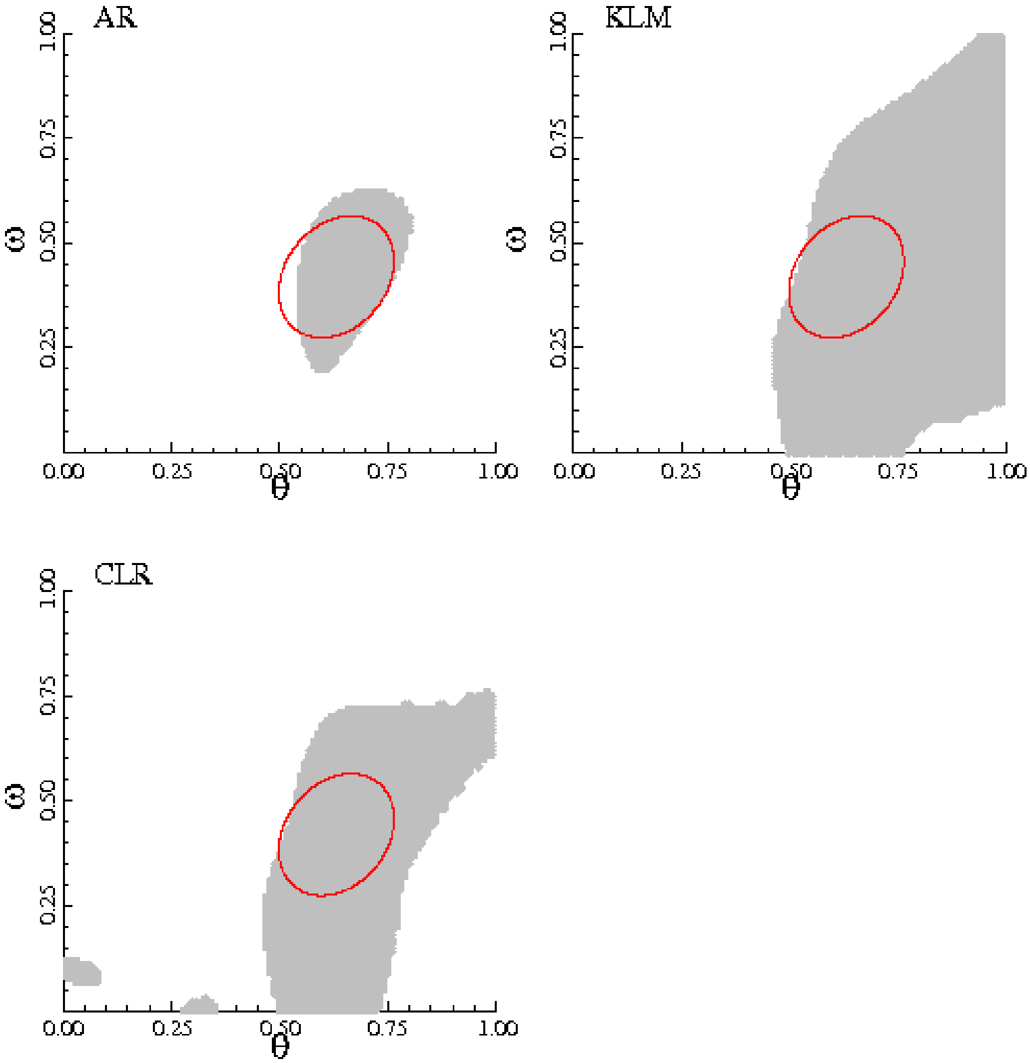

Figure 4: Joint 90%-level confidence sets for fraction of backward-looking firms ω and the probability prices remain fixed θ, subject to the restriction that the discount factor β is equal to 1. Instruments include 3 lags of $\Delta\pi_t$ and 4 lags of s_t, commodity price and wage inflation, output gap and long-short yield spread.

Table 3: Estimates of the new hybrid Phillips curve

	reduced-form param.		deep param.		diagnostics	
	λ	γ_f	ω	θ	OR	SC
Instruments						
small set	0.129	0.862	0.107	0.666	5.689	10.498
	(0.082)	(0.205)	(0.178)	(0.087)	[0.338]	[0.033]
large set	0.075	0.601	0.420	0.632	29.225	7.556
	(0.041)	(0.040)	(0.068)	(0.062)	[0.109]	[0.109]

The model is $E[Z_t(\pi_t - c - \lambda s_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1})] = 0$. Instruments include 3 lags of $\Delta\pi_t$ and 4 lags of s_t (small set), plus commodity price and wage inflation, output gap and long-short yield spread (large set). CUE-GMM with Newey and West (1987) Weight matrix, bandwidth: 4. Sample: 1960 (1) - 1997 (4). Diagnostics: OR is Hansen-Sargan test of overidentifying restrictions, $\chi^2(k-4)$; SC is Cumby and Huizinga (1992) test of residual autocorrelation from lags 2 to 5, $\chi^2(4)$.

tighter in the restricted specification at the 90% level.[12] However, the KLM and CLR-based sets are much wider than the AR ones, as was the case for the unrestricted model, see Figure 3. The main difference in the two specifications is that in the restricted model, it is ω rather than θ that is effectively unidentified. In the restricted specification, prices remain fixed for at least 2 quarters ($\theta > 1/2$).

In sum, there is no evidence that imposing the restriction $\beta = 1$ helps identify the other two structural parameters ω and θ. The intuition behind this is simple. In order to improve identifiability, restrictions must be placed on those parameters that are weakly identified. Here, β appears to be well-identified, so fixing it has little effect in improving the identifiability of the remaining parameters.

[12]The fact that the restricted confidence sets are tighter than the unrestricted ones is merely due to the fact that the over-identifying restrictions become sharper in the restricted specification. Compare the Hansen-Sargan test of over-identifying restrictions between the two models: the p-value drops from 0.14 for the unrestricted model (see Table 1) to 0.11 for the restricted one (see Table 3).

6 Conclusions

In this paper, we applied identification-robust inference procedures to test the parameters of the new Keynesian Phillips curve. Our results show that these parameters are weakly identified, and therefore help explain the conflicting estimates reported recently in the literature. The postwar US inflation history is consistent both with the view that inflation dynamics are purely forward-looking as well as with the view they are predominantly backward-looking.

The Phillips curve can also be estimated by full-information methods, such as maximum likelihood (FIML). These methods require the specification of a model for the observable forcing variables. They are typically more efficient than GMM, but not robust to mis-specification of the dynamics of the forcing variables. Recently, Fuhrer and Olivei (2004) argued that FIML may help alleviate the identification problem by making more efficient use of the information available in the data. However, the standard likelihood-based inference procedures, such as Wald, likelihood ratio and score tests, are not robust to failure of the identification assumption. This is the main reason why we took the limited information approach in this paper. The study of the identifiability of the Phillips curve by full-information methods is an important topic for future research.

A Appendix

A.1 Derivation of proposition 1

If $\gamma_f > 0$, the inverses of the roots of the characteristic polynomial (9) are given by (in ascending order)

$$\delta_1 = \frac{1 - \sqrt{1 - 4\gamma_f \gamma_b}}{2\gamma_f}, \qquad \delta_2 = \frac{1 + \sqrt{1 - 4\gamma_f \gamma_b}}{2\gamma_f}. \tag{15}$$

If $\gamma_f + \gamma_b < 1$ or $\gamma_f + \gamma_b = 1$ and $\gamma_f \leq \gamma_b$, then $|\delta_2| > 1$, which is explosive, and hence, the model has the unique non-explosive/determinate solution given by Eq. (10), where $\delta = \delta_1$ if $\gamma_f \neq 0$ and γ_b otherwise. If $\gamma_f > \gamma_b$, and $\gamma_f + \gamma_b = 1$, the inverses of the roots are $\delta_1 = (1 - \gamma_f)/\gamma_f$ and $\delta_2 = 1$, neither of which is explosive, so the solution is indeterminate, and is given by Eq. (11). To see that this is equivalent to (12), observe that $1 - \gamma_f^{-1} L - \gamma_b / \gamma_f L^2 = (1 - \delta_1 L)(1 - \delta_2 L)$, by definition. So, premultiplying Eq. (12) by $1-\delta_2 L$ and substituting for ξ_t from Eq. (13) we obtain

$$\pi_t = \frac{1}{\gamma_f}\pi_{t-1} - \frac{\gamma_b}{\gamma_f}\pi_{t-2} + \eta_t - \frac{\delta_2}{1-\delta_1\gamma_f}\varepsilon_{t-1}$$

It suffices to show that $\delta_2/\left(1-\delta_1\gamma_f\right) = 1/\gamma_f$, or, equivalently, that $1-\delta_1\gamma_f-\delta_2\gamma_f = 0$. This follows immediately from the definitions (15).

A.2 Derivation of proposition 2

Lack of global identification arises because there are multiple values of ω, θ, β that solve the equations (2) to (4) in terms of $\lambda, \gamma_f, \gamma_b$. It suffices to consider the case $\lambda \neq 0$, since otherwise, the model is partially identified by proposition 1.

First, assume that $\gamma_b \neq 0$. Dividing (3) by (4) yields:

$$\theta\beta = \omega\frac{\gamma_f}{\gamma_b}. \tag{16}$$

Dividing (2) by (3) and substituting for $\theta\beta$ using (16) yields

$$\lambda\omega\frac{\gamma_f}{\gamma_b} = (1-\omega)(1-\theta)\left(1-\omega\frac{\gamma_f}{\gamma_b}\right)\gamma_f.$$

The solution for θ depends on whether $\gamma_b - \gamma_b\omega - \gamma_f\omega + \gamma_f\omega^2 = (\omega-1)(\gamma_f\omega-\gamma_b) \neq 0$. Sufficient for this is that $\lambda \neq 0$, because $\omega = 1$ or $\omega = \gamma_b/\gamma_f$ imply $\lambda = 0$ from Eq. (2). Thus:

$$\theta = \frac{\gamma_b - \lambda\omega - \gamma_b\omega - \gamma_f\omega + \gamma_f\omega^2}{(\omega-1)(\gamma_f\omega-\gamma_b)} \tag{17}$$

Substituting (5) and (16) into (4) we obtain

$$\theta\gamma_b(1-\omega) + (\gamma_b+\gamma_f\omega-1)\omega = 0$$

and substituting (17), we have

$$\gamma_b(1-\omega)\frac{\gamma_b - \lambda\omega - \gamma_b\omega - \gamma_f\omega + \gamma_f\omega^2}{(\omega-1)(\gamma_f\omega-\gamma_b)} + (\gamma_b+\gamma_f\omega-1)\omega = 0$$

or

$$\gamma_b^2 - (1+\lambda+\gamma_f)\gamma_b\omega + (1+\gamma_b)\gamma_f\omega^2 - \gamma_f^2\omega^3 = 0.$$

This is a cubic in ω, and can generally have 3 real roots. This establishes that the structural parameters are not globally identified even when the reduced form parameters are.

If $\gamma_b = 0$, then $\omega = 0$ by Eq. (4), $\beta = \gamma_f$ by Eq. (3), and θ solves

$$\lambda\theta = (1 - \theta)(1 - \gamma_f \theta)$$

which has up to two real solutions if $(\lambda + \gamma_f + 1)^2 - 4\gamma_f \geq 0$.

Partial identification is discussed in the main text.

B Simulations

We perform a number of simulation experiments to examine the finite sample size and power of the identification-robust statistics AR, KLM, JKLM and CLR, as well as the non-robust Wald test. The data generating process (DGP) is specified as follows. We consider a range of values of the structural parameters in the following two specifications. Specification 1 is the reduced-form NKPC in Eq. (1), with parameters $(\lambda, \gamma_f, \gamma_b)$ satisfying the restrictions $\gamma_f, \gamma_b \geq 0,\ \gamma_f + \gamma_b \leq 1$ and $\lambda \geq 0$. Specification 2 is the Galí and Gertler (1999) version of the model with parameters $(\omega, \theta, \beta) \in [0,1] \times [0,1] \times [0,1]$. For simplicity, we impose the restriction $\gamma_f + \gamma_b = 1$ or $\beta = 1$. The results for the unrestricted models are similar and are omitted.

To simulate data we need a description of the dynamics of the forcing variable s_t. We assume s_t is a stable linear process that can be represented by an autoregressive distributed lag model:

$$\rho(L) s_t = \varphi(L) \pi_t + v_t \tag{18}$$

where $\rho(L)$ and $\varphi(L)$ are fourth order lag polynomials with $\rho_0 = 1$ and $\varphi_0 = 0$. The disturbances (ε_t, v_t) are assumed to be Gaussian innovations, with variance parameters $\sigma_\varepsilon^2, \sigma_{v\varepsilon}$ and σ_v^2. Mavroeidis (2005) showed that identification is invariant to the scale of v_t, so, without loss of generality, we normalize σ_v^2 to 1. The parameters $\sigma_\varepsilon^2, \sigma_{v\varepsilon}, \rho_i, \varphi_i$, $i = 1, ..., 4$ are calibrated to the US data over the period 1960-1997 as follows. First, we estimate the model (18) by OLS and derive the residuals $\hat{v}_t$. Then we combine (18) with (1) to find the solution for the law of motion for π_t, which we solve for the innovations $\hat{\varepsilon}_t$. Finally, we back out estimates of σ_ε^2 and $\sigma_{v\varepsilon}$ from the covariance matrix of the $\hat{\varepsilon}_t$ and $\hat{v}_t$.

We conduct three different Monte Carlo simulation experiments. In the first experiment, we study the rejection probabilities under the null hypothesis (NRP) of the Wald, Anderson-Rubin, KLM and JKLM and CLR tests at a nominal 5% level of significance for different values of the structural parameters $(\lambda, \gamma_f, 1-\gamma_f)$ in the reduced-form specification (1). In the second experiment, we compute the power curves for tests of the null hypothesis of no forward-looking dynamics ($\gamma_f = 0$). In the last experiment, we study the coverage probability of 95%-level confidence sets on the parameters (ω, θ) of the structural-form specification of the model, for a wide range of true values of those parameters.

The sample size in all the experiments is set to 150, to match the sample size available for the empirical analysis. The estimation method is the continuously updated GMM estimator of Hansen, Heaton, and Yaron (1996) and the weight matrix is the inverse of the Newey and West (1987) HAC estimator. Finally, we consider two instrument sets: a small set that includes only the first two lags of π_t and s_t, and a larger instrument set that includes the first four lags of those variables.

B.1 Experiment 1: size comparisons of alternative tests

Table 4 reports the null rejection probabilities (NRP) of Wald, AR, KLM, JKLM and CLR tests with nominal level 5% for hypotheses on each (λ, γ_f) at different true values. Several features of these results are noteworthy. First, consider the NRPs of the Wald test. They vary from 0 to over 70%, showing that the test is severely size-distorted, and therefore cannot be used for valid inference. The size distortion is more pronounced when the number of instruments is larger. This contrasts sharply with the NRPs for the identification-robust statistics. Although the AR and JKLM tests exhibit some mild overrejection when the number of instrument is large, their size appears to be close to 5% in most cases. The KLM test is the one whose NRP is closest to its nominal size.

Notable under-rejections occur for the just-identified model (k=4), for tests of null hypotheses

Table 4: Null rejection frequencies of various tests with nominal level 5% in the reduced-form hybrid Phillips curve

		$\lambda = 0.1$				$\lambda = 1$			
	Test on:	λ		γ_f		λ		γ_f	
	Instruments:	2 lags	4 lags	2 lags	4 lags	2 lags	4 lags	2 lags	4 lags
	Test								
$\gamma_f = 0$	W	0.149	0.340	0.371	0.736	0.353	0.656	0.355	0.673
	AR	0.033	0.061	0.085	0.127	0.072	0.121	0.074	0.124
	KLM	0.035	0.065	0.076	0.109	0.075	0.102	0.077	0.101
	JKLM	0.034	0.054	0.064	0.107	0.061	0.102	0.062	0.105
	CLR	0.011	0.037	0.043	0.087	0.036	0.079	0.037	0.082
$\gamma_f = .6$	W	0.008	0.025	0.086	0.117	0.040	0.068	0.094	0.122
	AR	0.040	0.077	0.070	0.094	0.047	0.098	0.070	0.112
	KLM	0.013	0.013	0.061	0.044	0.030	0.043	0.067	0.063
	JKLM	0.059	0.092	0.068	0.095	0.058	0.107	0.061	0.107
	CLR	0.006	0.012	0.025	0.026	0.008	0.018	0.025	0.027
$\gamma_f = 1$	W	0.004	0.011	0.111	0.164	0.027	0.045	0.111	0.173
	AR	0.039	0.072	0.076	0.106	0.050	0.086	0.076	0.110
	KLM	0.004	0.005	0.069	0.078	0.020	0.026	0.071	0.083
	JKLM	0.067	0.091	0.070	0.094	0.067	0.094	0.070	0.097
	CLR	0.001	0.002	0.024	0.035	0.006	0.012	0.027	0.040

W stands for Wald test; AR: Anderson-Rubin test, (Stock and Wright (2000)); KLM, JKLM and CLR tests from Kleibergen (2005). The model is the reduced-form NKPC with the restriction $\gamma_b + \gamma_f = 1$ imposed.

on λ, when λ is close to zero. This can be understood by proposition 1, since, as λ gets close to zero, we expect that the remaining parameters γ_f and γ_b, that are partialled out when testing hypotheses on λ, are poorly identified. These results are consistent with the conservativeness of the subset tests when identification is weak in the linear IV regression model, established by Kleibergen (2007).

All in all, this experiment demonstrates the superior size properties of the AR, KLM, JKLM and CLR tests relative to the Wald test in the context of the NKPC.

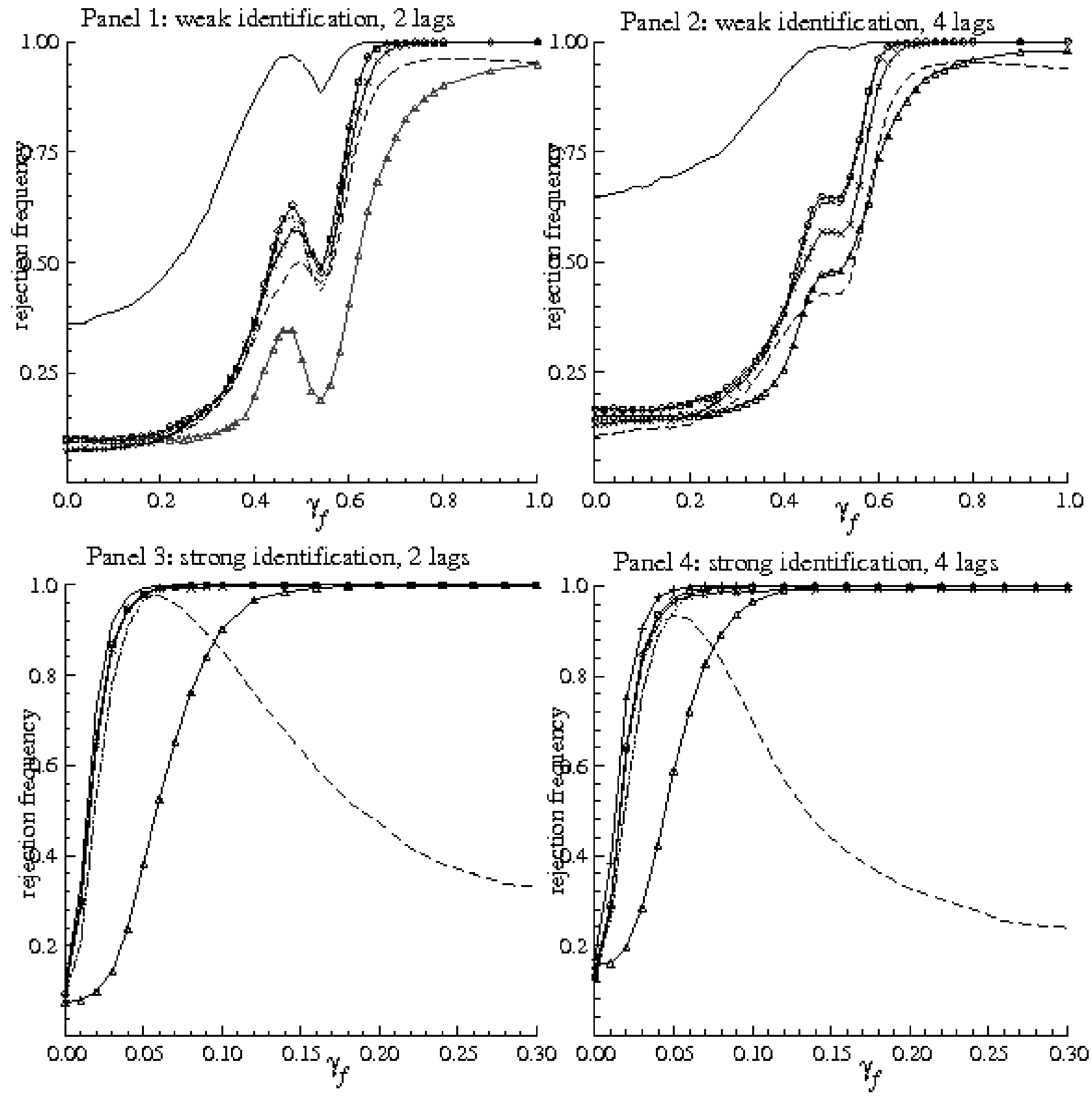

Figure 5: Power curves of tests of the null hypothesis $\gamma_f = 0$ at the 5% level: Wald (solid line), AR (dotted), KLM (dashed), JKLM (triangles), CLR (circles), KLM-JKLM $\alpha_K = 0.4, \alpha_J = 0.01$.

B.2 Experiment 2: power comparisons

In the first two panels of Figure 5, we plot the power curves corresponding to the DGP of experiment 1, with all nuisance parameters matched to the data. This DGP results in weak identification and causes the Wald test to be highly size-distorted, especially when we use more instruments. The remaining tests have little power for alternatives close to 0, but power picks up when γ_f exceeds 0.6. They are also much less sensitive to increasing the number of instruments.

Panels 3 and 4 of Figure 5 correspond to a DGP that results in strong identification. Following Mavroeidis (2005), this is achieved by setting the variance of the structural shock to be 16 times smaller than in the observed data, and changing the second autoregressive coefficient in s_t from -0.05 to -0.8. Since γ_f is well-identified, power picks up almost immediately, so we only look at the range 0 to 0.3. It is clear that the CLR and Wald tests have approximately the same power. We also see that the KLM test exhibits non-monotonic power, which is a well-known phenomenon (Kleibergen (2005)). The power of the KLM test can be made almost identical to the CLR when we combine it with a JKLM pretest. All in all, these results show that the identification robust tests do not waste power when identification is strong.

It is notable that even the identification-robust statistics appear over-sized in all panels of Figure 5. When we repeated those experiments without using a HAC weighting matrix, the size distortion disappeared completely in all cases. Therefore, we think that this size distortion is attributable to the HAC estimator, which is, in fact, a well-known problem, see den Haan and Levin (1997). It is clear from our results that the distortion arising from the HAC estimator is unrelated to the degree of identification.

B.3 Experiment 3: Coverage probabilities of confidence sets

The coverage probability of a 95%-level confidence set is the actual probability that the set will contain the true value of the parameter. In this experiment, we consider two-dimensional confidence sets for the parameters in the structural specification, ω and θ when β is fixed at 1. These confidence sets are constructed by inverting each test statistic, i.e., they contain all the values of the parameters that cannot be rejected by the corresponding test at the 5% level of significance.

Table 5 reports the coverage probabilities of the confidence sets for various true values of the parameters ω, θ. The contrast in the behavior of the Wald-based set relative to the identification-robust alternatives is even sharper than in experiment 1. The coverage probabilities for the Wald-based set are always considerably smaller than their nominal 95%. The smallest coverage rate 35% occurs when the true values are $\omega = 0.9$ and $\theta = 0.1$. These results show that the Wald set is typically much tighter than it is supposed to be, giving a very misleading sense of accuracy in the estimation of the parameters.

In contrast, the identification-robust statistics have coverage rates that are much closer to their nominal size in all cases, and typically not very different from 95%. Notable exceptions occur in the cases when ω is large and θ is small, when the coverage rate can drop to as low as 86%. This apparently large size distortion is not surprising, once we observe that at those values of the parameters, inflation becomes very persistent (the highest autoregressive root is about 0.99). We do not expect the asymptotic chi square critical values to be reliable in such cases when we have a moderately-sized sample.

In sum, this experiment shows that confidence sets for the structural specification of the Galí and Gertler (1999) model that are based on the Wald test can be very misleading, while identification-robust confidence sets are generally reliable, except in cases when the data are very persistent. But even in those cases, identification-robust sets are still more reliable than Wald tests, since they suffer from much smaller size distortion.

References

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statistics 20*, 46–63.

Andrews, D. W. and J. H. Stock (2005, August). Inference with weak instruments. NBER Technical Working Papers 0313, National Bureau of Economic Research, Inc.

Table 5: Coverage probabilities of 95%-level confidence sets on the structural specification of the new Keynesian Phillips curve

ω		θ :	0.1		0.25		0.75		0.9	
			2 lags	4 lags	2 lags	4 lags	2 lags	4 lags	2 lags	4 lags
0.1	W		0.912	0.861	0.897	0.833	0.827	0.746	0.767	0.677
	AR		0.936	0.884	0.936	0.883	0.947	0.912	0.947	0.912
	K		0.948	0.938	0.946	0.929	0.974	0.959	0.973	0.958
	J		0.933	0.882	0.934	0.889	0.928	0.903	0.929	0.903
	CLR		0.946	0.920	0.943	0.912	0.952	0.914	0.949	0.912
0.25	W		0.856	0.759	0.911	0.873	0.780	0.782	0.723	0.706
	AR		0.926	0.870	0.941	0.891	0.949	0.915	0.947	0.915
	K		0.939	0.924	0.951	0.952	0.975	0.964	0.974	0.964
	J		0.930	0.872	0.934	0.882	0.925	0.904	0.925	0.904
	CLR		0.929	0.897	0.949	0.932	0.953	0.919	0.949	0.915
0.5	W		0.657	0.488	0.700	0.553	0.710	0.747	0.657	0.705
	AR		0.916	0.863	0.933	0.893	0.950	0.921	0.950	0.919
	K		0.923	0.896	0.950	0.940	0.974	0.975	0.975	0.972
	J		0.933	0.884	0.929	0.887	0.927	0.905	0.923	0.905
	CLR		0.914	0.862	0.936	0.899	0.956	0.928	0.951	0.919
0.75	W		0.622	0.391	0.670	0.460	0.631	0.500	0.558	0.530
	AR		0.916	0.858	0.928	0.878	0.954	0.926	0.952	0.924
	K		0.922	0.889	0.941	0.917	0.974	0.985	0.973	0.981
	J		0.933	0.884	0.930	0.887	0.933	0.903	0.926	0.901
	CLR		0.914	0.855	0.928	0.880	0.956	0.932	0.953	0.924
0.9	W		0.635	0.345	0.640	0.382	0.609	0.442	0.563	0.403
	AR		0.918	0.860	0.928	0.874	0.954	0.925	0.954	0.928
	K		0.923	0.890	0.941	0.909	0.977	0.983	0.975	0.986
	J		0.933	0.885	0.929	0.887	0.928	0.902	0.931	0.904
	CLR		0.916	0.857	0.927	0.874	0.957	0.929	0.955	0.929

10000 Monte Carlo replications used. MC standard error is 0.0022.
$\beta = 1$ and fixed.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica 74*(3), 715–752.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*(6), 1383–1414.

Blanchard, O. J. and C. M. Kahn (1980). The solution of linear difference models under rational expectations. *Econometrica 48*, 1305–11.

Buiter, W. and I. Jewitt (1989). Staggered wage setting with real wage relativities: Variations on a theme of Taylor. In W. Buiter (Ed.), *Macroeconomic Theory and Stabilization Policy*, pp. 183–199. Ann Arbor: University of Michigan Press.

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

Cragg, J. G. and S. G. Donald (1997). Inferring the rank of a matrix. *J. Econometrics 76*(1-2), 223–250.

Cumby, R. E. and J. Huizinga (1992). Testing the autocorrelation structure of disturbances in Ordinary Least Squares and Instrumental Variables regressions. *Econometrica 60*(1), 185–195.

den Haan, W. J. and A. T. Levin (1997). A practitioner's guide to robust covariance matrix estimation. In *Robust inference*, pp. 299–342. Amsterdam: North-Holland.

Dufour, J.-M. (2003). Identification, weak instruments and statistical inference in econometrics. Cahiers de recherche 2003-12, Universite de Montreal, Departement de sciences economiques. available at http://ideas.repec.org/p/mtl/montde/2003-12.html.

Dufour, J.-M., L. Khalaf, and M. Kichian (2006). Inflation dynamics and the new keynesian phillips curve: An identification robust econometric analysis. *Journal of Economic Dynamics* *and Control 30*(9-10), 1707–1727.

Fuhrer, J. C. (1997). The (un)importance of forward-looking behavior in price specifications. *Journal of Money, Credit and Banking 29*(3), 338–50.

Fuhrer, J. C. and G. R. Moore (1995). Inflation persistence. *Quarterly Journal of Economics 110*, 127–159.

Fuhrer, J. C. and G. P. Olivei (2004). Estimating forward-looking euler equations with gmm estimators: an optimal instruments approach. Working paper, 04-2, Federal Reserve Bank of Boston, USA.

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

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica 64*(2), 413–430.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics 14*, 262–280.

Hansen, L. P. and T. J. Sargent (1991). Exact linear rational expectations models: specification and estimation. In L. P. Hansen and T. J. Sargent (Eds.), *Rational Expectations Econometrics*. Boulder: Westview Press.

Jondeau, E. and H. Le Bihan (2003). ML vs. GMM estimates of hybrid macroeconomic models (with an application to the "New Phillips Curve"). Working paper, Banque de France.

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. (2007). Subset statistics in the linear IV regression model. Working paper, Brown University, USA.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics 127*(1), 97–126.

Linde, J. (2005, September). Estimating new-keynesian phillips curves: A full information maximum likelihood approach. *Journal of Monetary Economics 52*(6), 1135–1149. available at http://ideas.repec.org/a/eee/moneco/v52y2005i6p1135-1149.html.

Lubik, T. A. and F. Schorfheide (2004). Testing for indeterminacy: An application to U.S. monetary policy. *American Economic Review 94*(1), 190–216.

Ma, A. (2002). GMM estimation of the New Keynesian Phillips Curve. *Economics Letters 76*, 411–417.

Mavroeidis, S. (2002). *Econometric Issues in Forward-Looking Monetary Models*. DPhil thesis, Oxford University, Oxford.

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), 421–449.

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

Nason, J. M. and G. W. Smith (2005). Identifying the new keynesian phillips curve. Working Paper 2005-01, Federal Reserve Bank of Atlanta.

Newey, W. K. and R. J. Smith (2004). Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica 72*(1), 219–255.

Newey, W. K. and K. D. West (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Pesaran, M. H. (1987). *The limits to Rational Expectations*. Oxford: Blackwell Publishers.

Rudd, J. and K. Whelan (2006). Can rational expectations sticky-price models explain inflation dynamics? *American Economic Review 96*(1), 303–320.

Smets, F. and R. Wouters (2003). An estimated Dynamic Stochastic General Equilibrium model for the Euro area. Working paper, European Central Bank.

Stock, J., J. Wright, and M. Yogo (2002). GMM, weak instruments, and weak identification. *Journal of Business and Economic Statistics 20*, 518–530.

Stock, J. and M. Yogo (2003). Testing for weak instruments in linear IV regression. NBER technical working paper, 284, NBER, USA.

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

West, K. D. (1997). Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator. *Journal of Econometrics 76*, 171–191.

West, K. D. and D. W. Wilcox (1996). A comparison of alternative instrumental variables estimators of a dynamic linear model. *J. Bus. Econom. Statist. 14*(3), 281–293.

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