

Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)  
<http://hdl.handle.net/11245/2.99018>

---

File ID        uvapub:99018  
Filename      478fulltext.pdf  
Version        unknown

---

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type            pre-print - working paper  
Title            Identification issues in forward-looking models estimated by GMM, with an  
                  application to the Phillips curve  
Author(s)      S. Mavroeidis  
Faculty         FEB: Amsterdam School of Economics Research Institute (ASE-RI)  
Year             2004

FULL BIBLIOGRAPHIC DETAILS:

<http://hdl.handle.net/11245/1.346864>

---

*Copyright*

*It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content licence (like Creative Commons).*

---

Discussion Paper: 2004/05

# Identification Issues in Forward-Looking Models Estimated by GMM, with an Application to the Phillips Curve

Sophocles Mavroeidis

[www.fee.uva.nl/ke/UvA-Econometrics](http://www.fee.uva.nl/ke/UvA-Econometrics)

Department of Quantitative Economics  
Faculty of Economics and Econometrics  
Universiteit van Amsterdam  
Roetersstraat 11  
1018 WB AMSTERDAM  
The Netherlands

UvA  UNIVERSITEIT VAN AMSTERDAM



# Identification Issues in Forward-Looking Models Estimated by GMM, with an Application to the Phillips Curve \*

Sophocles Mavroeidis  
UvA-Econometrics

Department of Quantitative Economics  
University of Amsterdam

August 2004

## Abstract

Limited-information methods are commonly used to estimate forward-looking models with rational expectations, such as the “New Keynesian Phillips Curve” of Galí and Gertler (1999). In this paper, we address issues of identification that have been overlooked due to the incompleteness of the single-equation formulation. We show that problems of weak instruments may arise, depending on the properties of the ‘exogenous’ variables, and that they are empirically relevant. We also uncover a link between identification and dynamic mis-specification, and examine the (lack of) power of Hansen’s (1982)  $J$  test to detect invalid over-identifying restrictions. With regards to the New Phillips curve, we find that problems of identification cannot be ruled out, and they deserve further attention.

*JEL classification:* C22, E31

*Keywords:* Weak instruments, Rational Expectations, GMM, New Phillips Curve.

## 1 Introduction

Forward-looking Rational Expectations (RE) models are common in the macroeconomic literature. These models are typically of the form:

$$y_t = \beta E(y_{t+1}|\mathcal{F}_t) + \gamma y_{t-1} + d_t \quad (1)$$

where  $y_t$  is a decision variable,  $d_t$  is a ‘driving’ or ‘forcing’ variable, usually thought of as ‘exogenous’, and  $E(y_{t+1}|\mathcal{F}_t)$  is the expectation of  $y_{t+1}$  conditional on the information set  $\mathcal{F}_t$ . The popularity of

---

\*An earlier version of this paper was entitled “Identification and mis-specification issues in forward-looking models”. I am grateful to David Hendry and Adrian Pagan for numerous comments and insightful discussions. I also benefited greatly from comments by Jean Boivin, Peter Boswijk, Maurice Bun, Guillaume Chevillon, Kees Jan van Garderen, Massimo Guiliodori, Eilev Jansen, Søren Johansen, Frank Kleibergen, George Konaris, Sebastiano Manzan, John Muellbauer, Hashem Pesaran, Hugo Rodriguez, Jim Stock, Jan Tuinstra, Ken West and two anonymous referees.

such models derives from the fact that they make the notion of forward-lookingness in economic decisions explicit and address the so-called Lucas (1976) critique.

The estimation of those models has been the subject of considerable research (see Hansen and Sargent 1991 or Binder and Pesaran 1995 for reviews). The various methods proposed in the literature can be divided into full and limited information methods, such as Full Information Maximum Likelihood (FIML) and Generalized Method of Moments (GMM), respectively. The former require the specification of a completing model for the forcing variables and the derivation of the solution to the model, the restricted reduced form. Their implementation is computationally more demanding and their results are not robust to mis-specification of the completing model. In contrast, limited information methods obviate the need to model the forcing variables, and are generally robust to alternative distributions for them, at the cost of being less efficient than full information methods.

However, there is a substantive condition that must be satisfied for any estimation method to provide consistent estimates and reliable inference on the parameters of interest  $\beta$  and  $\gamma$ . Namely, these parameters must be *identified* on the available information. The question of identifiability of such RE models has been originally studied by Pesaran (1981). A thorough analysis of the order and rank conditions for the identification of RE models is given in Pesaran (1987). In that book, Pesaran warns against the “indiscriminate application of the Instrumental Variable (IV) method to rational expectation models”, stressing that the necessity of the identification condition “is often overlooked in the literature”. Pesaran also urges that the conditions for identification of the RE model under consideration must be checked prior to estimation by limited information methods.

Unfortunately, this problem appears to have been overlooked in the recent monetary economics literature, where it has become common practice to estimate forward-looking RE models by GMM. One example is the use of single-equation GMM for the estimation of forward looking monetary policy rules, popularized by Clarida, Galí, and Gertler (1998). Another important example is the influential paper of Galí and Gertler (1999) (henceforth GG), which uses the same econometric methodology in estimating the New Keynesian Phillips curve, a forward-looking model for inflation dynamics (see also Batini, Jackson, and Nickell 2000 and Galí, Gertler, and López-Salido 2001).

The main objective of the present study is to raise some methodological points regarding the identification analysis of forward-looking rational expectation models. We show that good use of limited-information methods, such as GMM, requires a structural identification analysis, which is typical of a full information approach. This analysis requires modelling the forcing variables explicitly, solving the model, and using the restrictions implied by rational expectations to assess the identifiability of the parameters.

Furthermore, we examine the problem of identification of the forward-looking model (1), in the light of the recent weak instruments literature (see Stock, Wright, and Yogo 2002 for a review). We distinguish between the problem of under-identification (the failure of the usual rank condition for identification), and weak identification (when the rank condition is satisfied, but GMM remains unreliable). Problems of weak identification have been well-documented across the spectrum of applied econometrics (Stock, Wright, and Yogo 2002). This paper contributes to this literature by addressing these issues in the context of the New Keynesian Phillips curve.

The specification of the Phillips curve proposed by GG consists of a term in expected future inflation, a term in lagged inflation and a term in the labor share. Using conventional GMM esti-

mation GG report that forward dynamics dominate backward-looking behavior, with a coefficient on future inflation equal to 0.6 relative to less than 0.4 for the coefficient on lagged inflation. In this paper we show that the identification of those parameters depends on the dynamics and variability of the labor share. When we calibrate these to US data we find that the model of the Phillips curve proposed by GG is weakly identified. In this case, simulations indicate that one would tend to estimate a substantially positive coefficient on expected inflation *irrespective* of what the population value of that coefficient is. More specifically, GMM estimates would tend to find dominant forward looking behavior even when the Phillips curve is purely backward looking. This means that existing estimates of the GG model are not reliable evidence that the Phillips curve has a significant forward looking component.

In fact, it seems that the GG model is mis-specified because it does not adequately capture the dynamics of inflation. If the model is mis-specified, then we find that its parameters will be spuriously identified in the following sense: the instruments will correlate strongly with the endogenous regressors (future inflation and the labor share) but they will not be exogenous. This has implications for standard pretests of identification that are based on the *unrestricted* covariance between the endogenous regressors and the instruments (e.g., Galí, Gertler, and López-Salido 2001). Such tests do not impose the restrictions on the dynamics that are implied by the model, and they can be very misleading because they have power to detect identification even when the model is mis-specified.

In principle, mis-specification can be tested using Hansen’s (1982)  $J$  test of overidentifying restrictions. We conduct a simulation experiment, calibrated to match the observed inflation dynamics in the data, in order to study the power of the  $J$  test in this context. We find that the common practice of using a very large number of instruments, and unnecessarily general corrections for serial correlation (based on a Newey and West 1987 GMM weighting matrix), virtually annihilates the power of the  $J$  test to detect this mis-specification in finite samples of order less than 1000. The power of the test can be increased substantially by using fewer instruments and a different weighting matrix proposed by West (1997), which is more appropriate for inference on forward-looking models.

The structure of the paper is as follows. Section 2 introduces the hybrid Phillips curve model of GG and defines the relevant concepts of identification. Section 3 analyzes the identifiability of the model, assuming it is correctly specified. Section 4 relaxes the assumption of correct specification and studies the implications of dynamic mis-specification. Finally, section 5 concludes. Algebraic derivations are given in the appendix at the end. An unpublished additional appendix is available on request.

## 2 Preliminaries

### The model

‘Phillips curve’ is the name economists use to refer to an equation that describes the evolution of prices (or of the inflation rate) in a macroeconomic system. The New Keynesian Phillips curve is a pure forward-looking model of inflation dynamics, which typically takes the form:

$$\pi_t = \lambda x_t + \beta \mathbf{E}(\pi_{t+1} | \mathcal{F}_t) \tag{2}$$

where  $x_t$  is a forcing variable, usually a measure of the output gap or marginal costs. The set  $\mathcal{F}_t$  contains, in principle, all of the information that is available to the agents at time  $t$ , which is usually more than a handful of macroeconomic variables that the econometrician may have at their disposal. Following the convention in the literature (Binder and Pesaran 1995), we assume that  $\mathcal{F}_t$  contains at least current and past values of the endogenous variable  $\pi_t$ , and the forcing variable  $x_t$ , namely  $\mathcal{F}_t = (\pi_t, \pi_{t-1}, \dots; x_t, x_{t-1}, \dots; \dots)$ .

Model (2) can be seen as a limiting case of a more general model that accommodates both forward- and backward-looking price-setting behavior. This prompted a number of researchers to put forward a hybrid version of new and old Phillips curves (Fuhrer and Moore 1995 and Buiter and Jewitt 1989):

$$\pi_t = \lambda x_t + \gamma \mathbb{E}(\pi_{t+1} | \mathcal{F}_t) + (1 - \gamma) \pi_{t-1} \quad (3)$$

GG proposed a new hybrid version, which is motivated by the idea of combining both forward- and backward-looking price-setting behavior. This leads to:

$$\pi_t = \lambda s_t + \gamma_f \mathbb{E}(\pi_{t+1} | \mathcal{F}_t) + \gamma_b \pi_{t-1} + \epsilon_t \quad (4)$$

where  $\epsilon_t$  is an exogenous inflation shock, which is an innovation w.r.t.  $\mathcal{F}_{t-1}$  with variance  $\sigma_\epsilon^2$ , and  $s_t$  is a measure of real unit labour costs in deviations from their steady state, which is used as a proxy for marginal costs following Sbordone (2002).<sup>1</sup>

### GMM estimation

Model (4) cannot be estimated directly due to the fact that  $\mathbb{E}(\pi_{t+1} | \mathcal{F}_t)$  is a latent variable. Therefore, we replace it by  $\pi_{t+1}$  in order to derive the GMM estimating equation:

$$\pi_t = \lambda s_t + \gamma_f \pi_{t+1} + \gamma_b \pi_{t-1} + e_t \quad (5)$$

where  $e_t = \epsilon_t - \gamma_f \eta_{t+1}$  is the ‘GMM residual’, and  $\eta_{t+1} \equiv \pi_{t+1} - \mathbb{E}(\pi_{t+1} | \mathcal{F}_t)$  is the forecast error in predicting future inflation, and hence a mean innovation process with respect to  $\mathcal{F}_t$ . Note that  $e_t$  exhibits up to first-order serial correlation by construction, since  $\eta_t$  and  $\epsilon_t$  may be correlated.

Equation (5) is a linear GMM regression, with valid moment conditions of the form:

$$\mathbb{E} [(\pi_t - \lambda s_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1}) \bar{Z}_t] = 0 \quad (6)$$

where  $\bar{Z}_t$  is a vector of instruments. The choice of admissible instruments  $\bar{Z}_t$  depends on the properties of the GMM residual  $\epsilon_t - \gamma_f \eta_{t+1}$ . Rational expectations together with  $\mathbb{E}(\epsilon_t | \mathcal{F}_{t-1}) = 0$  imply that any  $\bar{Z}_t \in \mathcal{F}_{t-1}$  is admissible. If we further assume that  $s_t$  is exogenous, namely  $\mathbb{E}(\epsilon_t s_t) = 0$ , then  $s_t$  is also an admissible instrument.

The estimator used by GG is a 2-step GMM estimator with a 12-lag Newey and West (1987) Heteroscedasticity and Autocorrelation Consistent (HAC) estimate of the variance of the moment conditions. Their instrument set includes only lagged variables, which means that they implicitly treat  $s_t$  as endogenous.<sup>2</sup> All the empirical results in this paper are based on the original data set of GG. The data is quarterly, and the sample spans from 1960:Q1 to 1997:Q4. The variable definitions and measurement are given in the appendix.

## Concepts of identification

Although the linear GMM regression (5) differs from a prototype linear IV regression model, the latter provides useful insights into the relevant identification issues. Identification depends on the correlation between the endogenous regressors and the instruments: under-identification occurs when this correlation is zero and weak identification arises when it is small. Also, the strength of identification can be studied using the so-called concentration parameter, which is defined in Appendix A.2.

The concentration parameter is a matrix of dimension equal to the number of endogenous variables in the model and can be thought of as a multivariate signal-noise ratio in the first-stage regression of the endogenous variables on the instruments. This matrix characterizes the finite-sample distributions of IV estimators and test statistics and their departure from the first-order asymptotic approximations (see Staiger and Stock 1997). Moreover, as argued by several authors, the minimum eigenvalue of the concentration parameter matrix can serve as a scalar unitless measure of identification, or even used to develop a test for weak instruments (Stock and Yogo 2003). Hereafter, we will refer to the concentration parameter as the minimum eigenvalue of the conventional concentration matrix, denoted  $\mu^2$ .

The usual rank condition for identification is equivalent to  $\mu^2 > 0$ , and when satisfied, we say that the model is *generically* identified. In contrast, when  $\mu^2 = 0$ , the model is *partially* or *under-*identified. As the weak instruments literature has shown, generic identification is not sufficient for GMM estimation and inference to be reliable. Identification is deemed *weak* when inferences based on conventional normal approximating distributions are misleading (Stock, Wright, and Yogo 2002).

To be more precise, it is useful to have a working definition of weak identification. Stock and Yogo (2003) offer such a definition in the context of the prototype IV regression, where weak identification is synonymous to weak instruments. Instruments are deemed weak if  $\mu^2$  is below some threshold, which varies according to the chosen tolerance criteria for expected bias of any given IV estimator or maximum size distortion of tests based on it. Typically, a value of  $\mu^2$  less than 10 would be considered indicative of weak identification.<sup>3</sup>

## 3 Identification analysis

To examine the identification of the forward-looking model (4) using GMM, we need to look at the first-stage regression for the endogenous regressors,  $\pi_{t+1}$  and possibly  $s_t$ , given a set of instruments  $\bar{Z}_t$ . The nature of the endogenous regressors and the instruments means that the first-stage regression can be derived upon knowledge of the reduced form of the system  $(\pi_t, s_t)$ . It is well-known that the RE assumption in model (4) implies restrictions on this reduced form. To derive the restricted reduced form, we need to specify a completing model for the forcing variables and solve the model. This approach will be referred to as *structural* identification analysis.

Alternatively, we can perform a pretest of identification based on the unrestricted first-stage regression, that is, ignoring the restrictions implied by the structural model. This approach is standard in IV regressions. However, such pretests can be misleading in the case of the forward-looking model (4), because they have power even when the model is spuriously identified through mis-specification, as we explain below.

In contrast, by conditioning on the structural model being correctly specified, structural identification analysis avoids conflating identification and mis-specification issues and therefore, it is more reliable than standard identification pretests.

### 3.1 Under-identification

We start by examining the pathological cases of under-identification, and offer a result that can be used to derive a simple test of the under-identification hypothesis.

For the analysis of this section, we consider a generalized version of the baseline equation (4), allowing for more dynamics of  $\pi_t$  and  $s_t$  in the model, namely

$$\pi_t = \sum_{j=0}^n \lambda_j s_{t-j} + \gamma_f \mathbb{E}(\pi_{t+1} | \mathcal{F}_t) + \sum_{i=1}^m \gamma_i \pi_{t-i} + \epsilon_t \quad (7)$$

To discuss identification, we need a completing model for  $s_t$ . We consider the following linear specification:

$$s_t = \sum_{i=1}^p \rho_i s_{t-i} + \sum_{j=1}^q \varphi_j \pi_{t-j} + v_t \quad (8)$$

where  $v_t$  is a mean innovation process w.r.t.  $\mathcal{F}_{t-1}$ , with  $\mathbb{E}(v_t^2) = \sigma_v^2$  and  $\mathbb{E}(v_t \epsilon_t) = \sigma_{v\epsilon}$ . The parameters of the completing process  $\{\rho_i\}_{i=1}^p$ ,  $\{\varphi_j\}_{j=1}^q$ ,  $\sigma_v^2$  and  $\sigma_{v\epsilon}$ , are the nuisance or auxiliary parameters.

Appendix A.3 discusses the conditions for existence and uniqueness of a solution and derives the backward and forward solutions to the model. The latter takes the form

$$\pi_t = \sum_{i=1}^{\kappa_\pi} \delta_i \pi_{t-i} + \sum_{j=0}^{\kappa_s} \alpha_j s_{t-j} + \alpha_\epsilon \epsilon_t \quad (9)$$

where  $\kappa_\pi = \max(q-1, m)$  and  $\kappa_s = \max(p-1, n)$ . The coefficients  $\{\delta_i\}$ ,  $\{\alpha_j\}$  and  $\alpha_\epsilon$ , are functions of the structural and nuisance parameters (see Appendix A.3). The condition for uniqueness of the solution (determinacy) depends on the values of the structural and nuisance parameters. Thus, the parameter space can be partitioned into a determinacy region, in which the forward solution (9) is the unique solution to the model (7); and an indeterminacy region in which several other solution are possible. These other solutions, which we refer to as backward solutions, are characterized by richer dynamics and the possibility of sunspot shocks (see Appendix A.3).

We focus on the forward solution for various reasons. First, because it is more relevant for many important applications of the forward-looking model (7), like the New Phillips curve example that we consider here, but also many other linear quadratic adjustment cost models (e.g., the inventory models in West and Wilcox 1996), where the underlying economic theory restricts the parameters of the model to lie in the determinacy region *a priori*. Second, because the objective of this study is to expose pathological situations of weak identification, and these are more likely to occur when the model has a forward solution (see Mavroeidis 2004). Third, backward solutions make identification analysis more involved since it is necessary to address the issue of the indeterminacy of the reduced form.

The following result characterizes pathological cases of under-identification, and generalizes Pesaran (1987, Proposition 6.2).



**Proposition 3.1.** *When the forcing variable follows (8) and the reduced form of the structural model (7) is given by the forward solution (9), the structural parameters are under-identified if  $\rho_i = 0$  for all  $i > n + 1$  and  $\varphi_j = 0$  for all  $j > m + 1$ .*

This result follows by noting that, if the dynamics of the forcing variable are limited by  $q \leq m + 1$  and  $p \leq n + 1$  in (8), then  $\kappa_\pi = m$  and  $\kappa_s = n$  in the reduced form (9). So the reduced form is *nested* within the structural model (7), and therefore, there are more structural than reduced-form parameters. Since the latter determine the data generating process (DGP), this means there are infinitely many observationally equivalent structures, i.e. there are infinitely many values of the structural parameters consistent with the same DGP.<sup>4</sup>

Note that Proposition 3.1 applies not only to cases in which the structural model has a unique (determinate) solution, but also in the case of indeterminacy, provided the reduced form is characterized by a forward solution.<sup>5</sup> This is arguably a special case that rules out sunspots. Under indeterminacy, the possibility of sunspot equilibria introduces additional considerations for the identification analysis of the model, which are beyond the scope of the present study. Interested readers are referred to Pesaran (1987) and the recent important contribution of Lubik and Schorfheide (2004) who propose a test for indeterminacy and discuss some relevant identification issues.

Proposition 3.1 suggests a simple test of the null hypothesis that the structural model (7) is under-identified, based on the completing model (8). However, it must be emphasized that the converse of this proposition does not hold, i.e., higher order dynamics in  $s_t$  are necessary but not sufficient for the generic identification of the structural model. A simple counter-example will be given in the context of the GG model below.

### 3.1.1 Application to the baseline model

The result of the previous section can be applied to the GG model (4), which is a special case of (7) with  $m = 1$  and  $n = 0$ . Proposition 3.1 suggests the under-identification hypothesis

$$\begin{aligned} H_0 : \quad & \rho_i = \varphi_j = 0, \text{ for all } i > 1 \text{ and } j > 2, \text{ against} \\ H_1 : \quad & \rho_i \neq 0, \text{ for some } i > 1 \text{ or } \varphi_j \neq 0 \text{ for some } j > 2. \end{aligned} \tag{10}$$

Provided the residuals in the completing model (8) are homoscedastic and serially uncorrelated, the above hypothesis can be tested by a standard F-test of exclusion restrictions in that regression. Upon estimating model (8) with  $p = q = 4$  by OLS, standard diagnostic tests on  $\hat{v}_t$  show no evidence of heteroscedasticity, serial correlation, or non-normality (details are in the unpublished additional appendix). Hence, the standard F-test is appropriate.

Using  $p = q = 4$  the F-statistic does not reject  $H_0$  even at the 70% level of significance. This means that if model (4) has a forward solution, as it is conventionally believed (see GG section 4.4), it will be under-identified.<sup>6</sup> This conclusion is robust to using more lags of  $\pi_t$  and  $s_t$  under the alternative, i.e., varying  $p$  and  $q$  from 5 to 8. In fact, even the more restrictive hypothesis that  $s_t$  follows a simple first-order autoregression, AR(1), cannot be rejected at the same level of significance. Thus, the completing model (8) admits a very simple parsimonious representation in this case.

### 3.2 Empirical identification

Even if the previous under-identification hypothesis (10) were rejected, this would be insufficient evidence to conclude the model is empirically identified on that information set. This is because the strength of identification still remains an issue that can be studied using the concentration parameter.

Equations (8) and (9) constitute the reduced form of the system  $(\pi_t, s_t)$ , that can be used to derive the first-stage regression for the endogenous variables  $(\pi_{t+1}, s_t)$  in the structural model (4). We conform with the common practice of treating the forcing variable  $s_t$  as endogenous.

For tractability and clarity, we consider the simplest completing model for  $s_t$  that could yield generic identification, namely

$$s_t = \rho_1 s_{t-1} + \rho_2 s_{t-2} + v_t. \quad (11)$$

The forward solution to the model is a special case of (9)

$$\pi_t = \alpha_0 s_t + \alpha_1 s_{t-1} + \delta_1 \pi_{t-1} + \alpha_\epsilon \epsilon_t \quad (12)$$

where the reduced-form parameters  $(\alpha_0, \alpha_1, \delta_1, \alpha_\epsilon)$  are functions of the structural and nuisance parameters (see Appendix A.4).

Leading (12) one period and taking expectations conditional on  $\mathcal{F}_{t-1}$ , the forecasting equation for  $\pi_{t+1}$  is

$$\pi_{t+1} = \tilde{\alpha}_1 s_{t-1} + \tilde{\alpha}_2 s_{t-2} + \delta_1^2 \pi_{t-1} + \tilde{\eta}_{t+1} \quad (13)$$

where  $\tilde{\alpha}_1, \tilde{\alpha}_2$  and  $\tilde{\eta}_{t+1}$  are given in Appendix A.4.

The GMM regression (5) consists of the two endogenous regressors  $Y_t = (\pi_{t+1}, s_t)'$  and one exogenous regressor  $X_t = \pi_{t-1}$ . The first-stage regression for the endogenous regressors  $Y_t$  consists of equations (13) and (11). It can be written in the generic form

$$Y_t = \Pi' Z_t + \Phi' X_t + V_t \quad (14)$$

where, in this case, the only relevant instruments are  $Z_t = (s_{t-1}, s_{t-2})'$  and  $X_t = \pi_{t-1}$ . The coefficient matrix  $\Pi$  is given in Appendix A.5, while  $\Phi = (\delta_1^2, 0)$ , and the residual is  $V_t = (\tilde{\eta}_{t+1}, v_t)'$ .

Let us now turn to the concentration parameter  $\mu^2$ , in order to investigate the potential sources of weak identification. A description of  $\mu^2$  is given in Appendix A.5, where it can be seen that  $\mu^2$  is analytically intractable even in this simple case. Some insight about  $\mu^2$  can be gained if we look at a restricted version of the model, where an analytical expression for  $\mu^2$  is available, see equation (27) in Appendix A.5. In order to understand the interplay of the structural and nuisance parameters in  $\mu^2$ , we point out the following properties.

First,  $\mu^2$  is a function not only of the nuisance parameters  $\rho_1, \rho_2, \sigma_v, \sigma_{v\epsilon}$  but also of the structural ones  $\lambda, \gamma_f, \gamma_b$  and  $\sigma_\epsilon$ . By definition,  $\mu^2$  is a function of the reduced-form parameters  $\alpha_0, \alpha_1, \delta_1, \alpha_\epsilon$  (see Appendix A.5), and these are linked to the structural parameters through the solution to the model (see equation (24) in the appendix). Contrast this with a model that has no forward-looking components, e.g., a traditional backward-looking Phillips curve, with  $\gamma_f = 0$  in (4). In that case, the only endogenous variable would be  $s_t$ , and the first-stage regression would coincide with the completing model for  $s_t$ , equation (11), whose coefficients are the nuisance parameters. So  $\mu^2$  would not depend on the structural parameters  $\lambda, \gamma_b$  or  $\sigma_\epsilon$ .

Second, the limiting case of under-identification that we discussed in section 3.1 ( $\mu^2 = 0$ ) arises when either  $\rho_2 = 0$  or  $\lambda = 0$  (see Appendix A.5). This shows that  $\rho_2 \neq 0$  is necessary but not sufficient for generic identification. That is, the model will be under-identified when  $\lambda = 0$ , irrespective of the dynamics of the forcing variables. It is obvious that if  $\lambda$  is zero in the model (4), then inflation is driven solely by the structural shock  $\epsilon_t$ , and it is independent of the dynamics of  $s_t$ . Third,  $\mu^2$  is invariant to re-scaling of the data. This means that it only depends on  $\sigma_\epsilon^2, \sigma_v^2$  and  $\sigma_{v\epsilon}$  through  $\sigma_v^2/\sigma_\epsilon^2$  and  $\sigma_{v\epsilon}/\sigma_\epsilon^2$ .

Fourth, since  $\mu^2 = 0$  when  $\rho_2 = 0$ , we expect it to be small when  $\rho_2$  is ‘close’ to zero, when holding all other parameters fixed. However, since  $\rho_2$  is not the only parameter governing the strength of identification, it does not follow that identification will be strong when  $\rho_2$  is significantly different from zero (both in a quantitative, as well as in a statistical sense). In order to illustrate this point, we explore numerically how  $\mu^2$  varies with the nuisance parameters  $\rho_2$  and  $\sigma_v/\sigma_\epsilon$ , holding the structural parameters  $(\lambda, \gamma_f, \gamma_b)$  fixed at the values reported by GG (see GG Table 2) and calibrating the remaining nuisance parameters to the GG data. In particular,  $\rho_1$  is allowed to vary with  $\rho_2$  so as to keep the first autocorrelation of  $s_t$ ,  $\rho_1/(1 - \rho_2)$ , fixed at the value 0.9 which is estimated from the data.<sup>7</sup> This value is typical of many other persistent macroeconomic time series.

The results are given in Table 1. We see that most of the values of the concentration parameter

Table 1: The concentration parameter for the New Keynesian Phillips curve as a function of  $\rho_2$  and  $\sigma_v/\sigma_\epsilon$ .

$\rho_2$	$\sigma_v/\sigma_\epsilon$ :	1/2	1	2	4	8
0.65		0.122	0.484	1.880	6.807	20.375
0.45		0.022	0.086	0.337	1.251	3.818
0.25		0.002	0.007	0.029	0.105	0.316
0.05		0.000	0.000	0.000	0.000	0.000
<b>-0.05</b>		<b>0.000</b>	0.000	0.000	0.000	0.000
-0.25		0.002	0.008	0.029	0.103	0.269
-0.45		0.023	0.092	0.354	1.239	3.187
-0.65		0.135	0.535	2.064	7.250	19.333

Note: This table reports the concentration parameter  $\mu^2$ , which is a measure of empirical identification of a structural model estimated by GMM. Typically,  $\mu^2 < 10$  indicates weak identification. The structural model is  $\pi_t = \lambda s_t + \gamma_f E(\pi_{t+1} | \mathcal{F}_t) + \gamma_b \pi_{t-1} + \epsilon_t$ ; the forcing variable follows  $s_t = 0.9(1 - \rho_2)s_{t-1} + \rho_2 s_{t-2} + v_t$ ;  $(\epsilon_t, v_t)$  are uncorrelated innovations with variances  $\sigma_\epsilon^2$  and  $\sigma_v^2$  respectively. The instruments are  $s_{t-1}, s_{t-2}$  and  $\pi_{t-1}$ . The table shows numerically how  $\mu^2$  varies as a function of  $\rho_2$  and  $\sigma_v/\sigma_\epsilon$ , when the remaining parameters are fixed at  $(\lambda, \gamma_f, \gamma_b) = (0.015, 0.591, 0.378)$ . In bold is the value of  $\mu^2$  that corresponds to the actual parameter estimates  $\hat{\rho}_2 = -0.05$  and  $\hat{\sigma}_v/\hat{\sigma}_\epsilon = 1/2$  from the data.

reported in Table 1 are small. That is, even when  $\rho_2$  is considerably different from zero, identification may still be weak, depending on the value of the remaining parameters. We also see that  $\mu^2$  is increasing in  $\sigma_v/\sigma_\epsilon$ , or, equivalently it is decreasing in the variability of the exogenous inflation

shock  $\sigma_\epsilon^2$ , other things equal.

The value of the concentration parameter that is implied by the estimates from the GG data (highlighted in Table 1) is indistinguishable from zero to 4 decimal places. Nonetheless, it must be emphasized that this is not a formal test of weak identification. Such a test would have to account for the uncertainty in estimating  $\mu^2$ , along the lines of Stock and Yogo (2003). The discussion here aims at highlighting the potential sources of weak identification, and shows that, for plausible values of the model’s parameters, weak identification cannot be ruled out.

Although it would be useful to have a formal test of weak identification, such a test is not yet available for forward-looking models like (4). The test of weak instruments proposed by Stock and Yogo (2003) is not applicable here because it does not account for autocorrelation and possible heteroscedasticity of the residuals in the linear GMM regression (5). A more important complication in deriving such a test arises from the fact that  $\mu^2$  also depends on the structural parameters, which cannot be consistently estimated under the null hypothesis of weak identification.

Finally, it is worth asking whether identification would get stronger if the forcing variables had additional dynamics. To investigate this, we generalized the completing process (11) to a third-order autoregression, by adding the term  $\rho_3 s_{t-3}$ . We set  $\rho_3 = 0.1$  and repeated the analysis of this section to derive  $\mu^2$  for all the cases reported in Table 1. Although for the cases  $\rho_2 > 0$  the concentration parameter increased slightly, for  $\rho_2 < 0$  it decreased substantially (see Appendix A.5 for an explanation). This shows that it is not, in general, true that higher order dynamics in the forcing variables will improve the empirical identification of the forward-looking model.

### 3.3 Implications of weak identification

In this section, we illustrate the devastating implications of weak identification for GMM estimation of the GG model (4) by means of some Monte Carlo experiments. The Monte Carlo setting requires the specification of values for all of the model’s parameters. We perform two experiments. In the first one, we set the parameters of interest  $\theta = (\lambda, \gamma_f, \gamma_b)$  to values reported by GG, (0.015, 0.591, 0.378), which suggest that forward looking behavior is dominant. The remaining parameters are calibrated to the GG data.

We simulate the GMM estimator used by GG for the three parameters of interest  $\theta$ . The results are reported in Table 2. The experiment in the top panel is based on the estimated parameters, and reveals weak identification, whereas the lower panel contrasts the results with a situation of strong identification (corresponding to the bottom right-hand corner of Table 1). Figure 1 compares the simulated distributions under weak and strong identification.

The simulation results reveal the following. When identification is weak, the GMM estimator  $\hat{\theta}$  remains biased even in what would be conventionally thought of as large samples. The estimates of the forward and backward looking coefficients  $\hat{\gamma}_f, \hat{\gamma}_b$  are on average equal to 0.42, but the distribution of  $\hat{\gamma}_f$  exhibits considerably more variability and skewness.

The most important feature of these results is that the dispersion of the estimators does not fall with the sample size when identification is weak. The estimators are biased at all sample sizes, and they show no tendency to converge to the true values of the parameters. This is in sharp contrast to the strongly identified case, where the bias disappears even in samples of moderate size, the GMM estimator is much more accurate and converges to the true value of the parameters at the usual rate  $\sqrt{T}$ .

Table 2: Simulation of the finite sample moments of the 2-step GMM estimator for the New Keynesian Phillips Curve.

(a) Weak identification.

Sample	$\gamma_b$		$\gamma_f$		$\lambda$	
	mean	st. dev.	mean	st. dev.	mean	st. dev.
50	0.410	0.120	0.421	0.330	0.044	0.297
100	0.421	0.099	0.429	0.322	0.039	0.164
150	0.425	0.093	0.428	0.298	0.038	0.119
300	0.428	0.094	0.426	0.305	0.039	0.084
500	0.430	0.092	0.425	0.291	0.037	0.063
1000	0.429	0.094	0.429	0.293	0.036	0.052

(b) Strong identification.

Sample	$\gamma_b$		$\gamma_f$		$\lambda$	
	mean	st. dev.	mean	st. dev.	mean	st. dev.
50	0.379	0.095	0.461	0.209	0.029	0.024
100	0.379	0.060	0.524	0.151	0.021	0.015
150	0.379	0.047	0.552	0.117	0.018	0.011
300	0.378	0.030	0.578	0.074	0.016	0.007
500	0.378	0.022	0.583	0.053	0.015	0.005
1000	0.378	0.015	0.588	0.036	0.015	0.003

Note: This table reports the mean and standard deviation of the GMM estimators for the New Keynesian Phillips curve  $\pi_t = \lambda s_t + \gamma_f E(\pi_{t+1} | \mathcal{F}_t) + \gamma_b \pi_{t-1} + \epsilon_t$ . The instrument set contains four lags of  $\pi_t$  and  $s_t$  and a 12-lag Newey-West (1987) estimate of the covariance matrix was used. The data is generated using  $s_t = 0.9(1 - \rho_2)s_{t-1} + \rho_2 s_{t-2} + v_t$ ;  $(\epsilon_t, v_t)$  are independent normally distributed innovations with variances 0.2 and  $\sigma_v^2$  respectively. The true values of the parameters are  $\gamma_b = 0.378, \gamma_f = 0.591, \lambda = 0.015$ . For weak identification,  $\rho_2 = -0.05, \sigma_v = 0.1$ . For strong identification,  $\rho_2 = -0.65, \sigma_v = 1.6$ .

This simulation appears to indicate relatively modest problems in the estimation of the forward looking coefficient  $\gamma_f$  when the population value of that coefficient is set at the value estimated by GG. However, this interpretation is incorrect. When the model is weakly identified, the estimator would give similar results for any other population value of  $\gamma_f$ . The estimator tends to find a substantially positive  $\gamma_f$ , irrespective of what the true value of  $\gamma_f$  is. The GG estimates are roughly consistent with this tendency of the estimator. This explains why the above simulation underrates the seriousness of the problem. We demonstrate this point by means of a second simulation experiment.

In our second experiment, we compare the behavior of the GMM estimator for different true values of the parameters in model (4) and show that the distribution of the estimator is roughly invariant to changes in the true values of those parameters. In particular, we compare the previous setting, in which the data was generated from a process with dominant forward-looking dynamics, to one in which the true inflation process is purely backward-looking. For simplicity, we refer to the

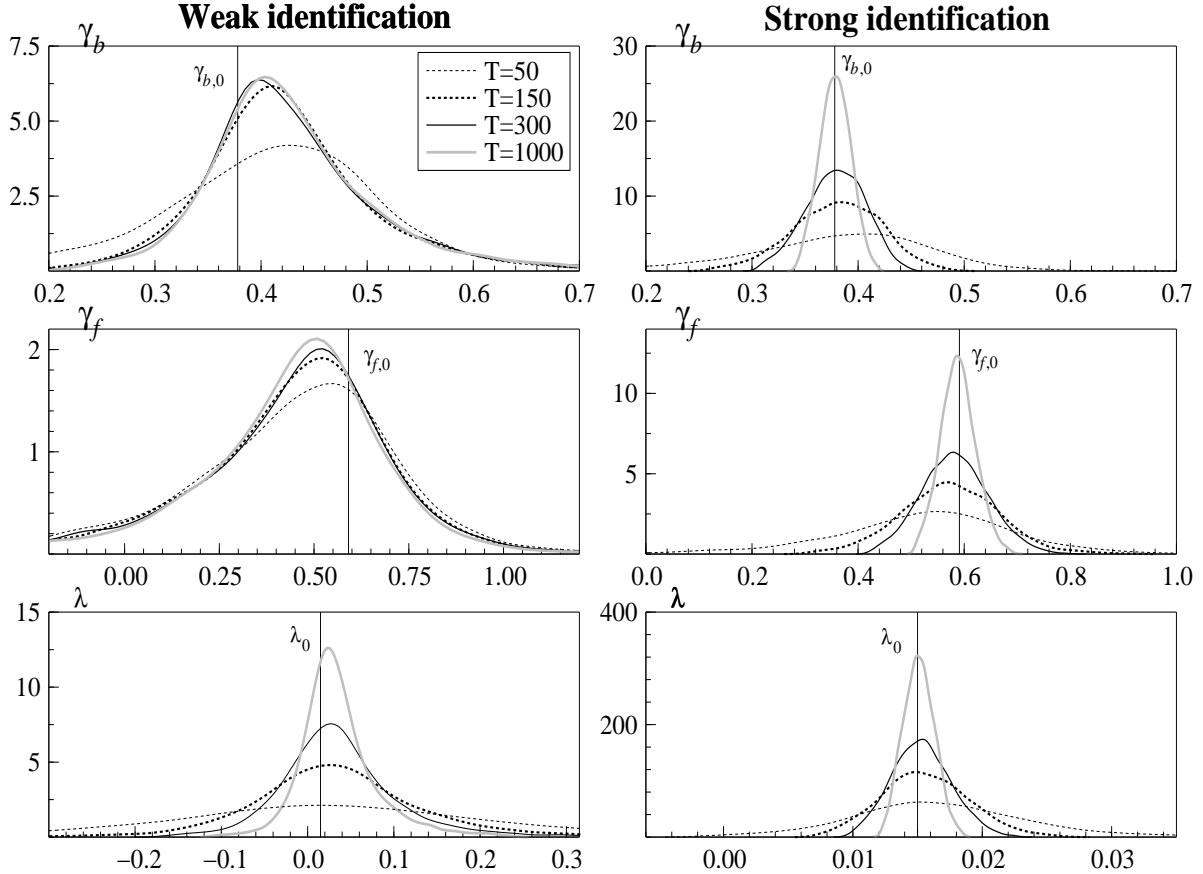


Figure 1: Simulated densities of the 2-step GMM estimator for the parameters of the New Keynesian Phillips Curve  $\pi_t = \lambda s_t + \gamma_f E(\pi_{t+1} | \mathcal{F}_t) + \gamma_b \pi_{t-1} + \epsilon_t$  at various sample sizes (denoted by  $T$ ). The instrument set contains four lags of  $\pi_t$  and  $s_t$  and a 12-lag Newey-West (1987) estimate of the covariance matrix was used. The data is generated using  $s_t = 0.9(1 - \rho_2)s_{t-1} + \rho_2 s_{t-2} + v_t$ ;  $(\epsilon_t, v_t)$  are independent normally distributed innovations with variances 0.04 and  $\sigma_v^2$  respectively. The true parameters are  $\gamma_b = 0.378, \gamma_f = 0.591, \lambda = 0.015$  (shown by vertical lines). For weak identification,  $\rho_2 = -0.05, \sigma_v = 0.1$ . For strong identification,  $\rho_2 = -0.65, \sigma_v = 1.6$ .

two settings as ‘new’ and ‘old’ Phillips curves respectively. The old Phillips curve is nested in (4) with  $\gamma_f = 0$ . For the simulation experiment, we set the remaining parameters such that the reduced form dynamics of inflation would be identical in both settings if model (4) was under-identified (see Appendix A.6).

Figure 2 shows the distribution of the estimators in four different cases. For every parameter estimator, we plot the two densities corresponding to simulations from the new and the old Phillips curves (solid and dashed lines respectively), and we do that both when the model (4) is weakly identified (left panels) and when it is well identified (right panels).

Under weak identification, the distributions of the estimators are almost indistinguishable for both DGPs (this is especially true for the estimator of  $\gamma_f$  in the middle left panel of Figure 2). In view of the discussion in the previous subsection, this should come as no surprise, since the concentration parameter  $\mu^2$  is in the order of  $10^{-5}$  in both settings. When  $\rho_2$  and  $\sigma_v/\sigma_\epsilon$  are calibrated to the data, instruments are so weak that the two different population values for the

parameter vector  $(\lambda, \gamma_f, \gamma_b, \sigma_\epsilon)$  are almost observationally equivalent. As explained in Appendix A.6, these two population values are chosen so as to be exactly observationally equivalent in the limiting case of  $\rho_2 = 0$ , where  $\mu^2$  is identically zero and the model is under-identified. This remarkable similarity of the estimator distributions is therefore not a coincidence, but a consequence of weak identification and is entirely consistent with the findings of the weak instruments literature, see Stock, Wright, and Yogo (2002). Unreported results show that the same conclusions arise for a wide range of the nuisance parameters over which  $\mu^2$  is small, see Table 1.<sup>8</sup>

For the simulations in the right panels of Figure 2 we set  $\rho_2$  and  $\sigma_v/\sigma_\epsilon$  such that  $\mu^2$  is large (around 40), see also Appendix A.6. Note that the second order dynamics and variability of  $s_t$  that are necessary to achieve strong identification are much higher than what is found in the data. In contrast to the case of weak identification, the GMM estimators clearly track the true values of the parameters when the model is well identified.<sup>9</sup>

This experiment demonstrates how unreliable GMM estimators are for all the parameters of the model when identification is weak. When the data is generated from a purely backward-looking Phillips curve,  $\gamma_f$  will be overestimated and  $\gamma_b$  will be underestimated. The direction of the biases is reversed when the true model is the new Phillips curve reported by GG. In both cases, the GMM estimators of  $\gamma_f$  and  $\gamma_b$  are on average equal to 0.42, and the skewness of their distributions implies that the outcome  $\hat{\gamma}_f > \hat{\gamma}_b$  occurs more often than the opposite outcome. Thus, GMM estimation of model (4) is biased in favor of a hybrid specification with apparently dominant forward-looking behavior, *irrespective* of the true nature of forward and backward-looking dynamics of inflation.

Weak identification also renders conventional tests on the parameters completely misleading, as it is well-known from the weak instruments literature (see Stock, Wright, and Yogo 2002). To illustrate this point, we consider the standard t-test on  $\gamma_f$ , which is often interpreted as a test of the significance of forward-looking behavior. The t statistic is the ratio of the GMM estimator to its estimated standard error, and it is assumed to have a standard normal distribution in large samples. However, when the model is weakly identified, the t-test may reject a true null hypothesis much more often than the chosen level of significance. In the present setting, when the data is generated from the old Phillips curve, a 5%-level t-test rejects the null hypothesis that  $\gamma_f = 0$  more than 60% of the time. This profound over-rejection means that the test would report evidence of forward-looking behavior even when this is not warranted by the data.

The above simulations demonstrate why it is dangerous to rely solely on the rank condition for identification, before proceeding with conventional estimation and inference. Even when the forcing variables have enough dynamics which would guarantee generic identification, e.g., a nonzero concentration parameter, the possibility of weak identification remains a serious issue.

### 3.4 Naive pretests of identification

The analysis of section 3.1.1 showed that  $s_t$  follows an AR(1) and so the only relevant instruments must be  $s_{t-1}$  and  $\pi_{t-1}$  if the model is correctly specified. We now re-examine instrument relevance by looking at an unrestricted version of the first stage-regression (14). We consider a  $k$  dimensional instrument vector  $\bar{Z}_t$ , that contains lags of  $s_t$  and  $\pi_t$ , and let the  $k \times 2$  coefficient matrix  $\Pi$  be unrestricted. The null hypothesis of under-identification is  $H_0 : \text{rank}(\Pi) \leq 1$  against the alternative  $H_1 : \text{rank}(\Pi) = 2$ .

In the prototype IV model, where the first-stage regression residuals  $V_t$  are homoscedastic and

serially uncorrelated, this could be tested using, say, the canonical correlations rank test of Anderson (1951). Here,  $V_t$  in equation (14) is serially correlated, and potentially heteroscedastic, so we apply a generalized reduced rank test proposed by Kleibergen and Paap (2003), the rk-statistic, using the Newey and West (1987) estimator to account for the autocorrelation and heteroscedasticity in  $V_t$ . Using four lags of  $\pi_t$  and  $s_t$  as instruments, the value of the test statistic is 29.13, with a p-value of 0.000, showing that the under-identification hypothesis is strongly rejected (details of the test can be found in the unpublished additional appendix).

Thus, the evidence on identifiability is mixed. On the one hand, the structural identification analysis suggested that the model must be partially or at best weakly identified. In other words, when the Phillips curve (4) is correctly specified, lagged values of  $\pi_t$  and  $s_t$  cannot be informative instruments for the endogenous variables. On the other hand, the unrestricted identification pretest suggests that those instruments are relevant. How can we reconcile these two contradictory findings?

The answer lies in the fact that the structural model (4) may be mis-specified, in the sense that the orthogonality conditions (6) are violated, i.e. at least some of the instruments in the information set  $\mathcal{F}_{t-1}$  are not valid. Such invalid instruments could correlate with future inflation, thus giving rise to spurious identification of the model. This possibility is explored in the next section.

## 4 Mis-specification

The objective of this section is twofold: (i) to explain the problem of spurious identification caused by mis-specification, and (ii) to examine why such mis-specification may be hard to detect.

By mis-specification of the model (4) we mean the violation of the moment conditions (6). Recall that the GMM residual in equation (5) is  $e_t = \epsilon_t - \gamma_f \eta_{t+1}$ , where  $\epsilon_t$  is the structural error in the model (4) and  $\eta_{t+1}$  is the forecast error in predicting  $\pi_{t+1}$  at time  $t$ . Hence, violation of the moment conditions admits two interpretations. It could be seen as a failure of rationality if it arises from  $E(\eta_{t+1} \bar{Z}_t) \neq 0$ . Alternatively, when  $E(\epsilon_t \bar{Z}_t) \neq 0$ , it could be thought of as an omitted variable problem. For instance, we may express the assumed structural error as  $\epsilon_t = \varepsilon_t + \xi' w_t$ , where  $\varepsilon_t$  is an innovation w.r.t.  $\mathcal{F}_{t-1}$ , that is,  $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ ,  $w_t$  is a vector of variables that correlate with the instruments and are incorrectly omitted from the model, and  $\xi$  is a vector of coefficients.<sup>10</sup>

Of course, in the absence of independent data on expectations, it is impossible to distinguish between those two interpretations. All we can observe is that the (unrestricted) reduced form dynamics of inflation are inconsistent with the rational expectations solution to the model. In the ensuing discussion, we maintain the assumption of rational expectations, following the approach in the literature, but this is inconsequential for our results.

In particular, we focus on the case where the omitted variables  $w_t$  consist of lags of  $\pi_t$  and  $s_t$ . In this case, the model is dynamically mis-specified, as in the pure forward-looking Phillips curve (2) relative to the hybrid model (4), where  $\xi' w_t = \gamma_b \pi_{t-1}$ . We look at this type of mis-specification for various reasons. First, it is well-suited for raising a number of methodological points of general interest. Second, it follows the approach in the literature, where the hybrid Phillips curve (4) was developed as an extension to the pure forward-looking model (2). Third, it nests other explanations, such as autocorrelated structural errors (e.g., Smets and Wouters 2003).<sup>11</sup> Finally, it is consistent with the data in the present example, and enables us to resolve the conflicting results of the



identification analysis in the previous section.

#### 4.1 Omitted dynamics and spurious identification

To examine the implications of dynamic mis-specification in the Phillips curve, we consider a particular generalization of the GG model (4) that would be consistent with the observed inflation dynamics in the data. This generalization is derived as follows. We first estimate a parsimonious reduced-form model for inflation, of the form (9), and then ‘invert’ this to find a forward-looking specification of the form (7) that has this reduced form as its solution.

The estimated reduced-form equation for inflation is of the form

$$\pi_t = \alpha_0 s_t + \alpha_2 s_{t-2} + \delta_1 \pi_{t-1} + \delta_3 \pi_{t-3} + u_t, \quad (15)$$

(estimates are given in Appendix A.7). Equation (15) together with an AR(1) specification for  $s_t$  enable us to derive  $E(\pi_{t+1}|\mathcal{F}_t) = (\alpha_0 \rho_1) s_t + \alpha_2 s_{t-1} + \delta_1 \pi_t + \delta_3 \pi_{t-2}$ . Then, given any value for the forward-looking parameter  $\gamma_f$ , we can subtract  $\gamma_f / (1 - \gamma_f \delta_1) \times E(\pi_{t+1}|\mathcal{F}_t)$  from both sides of equation (15) to get an isomorphic forward-looking specification

$$\pi_t = \lambda s_t + \lambda_1 s_{t-1} + \lambda_2 s_{t-2} + \gamma_f E(\pi_{t+1}|\mathcal{F}_t) + \gamma_1 \pi_{t-1} + \gamma_2 \pi_{t-2} + \gamma_3 \pi_{t-3} + \varepsilon_t \quad (16)$$

where the structural parameters  $(\lambda, \lambda_1, \lambda_2, \gamma_1, \gamma_2, \gamma_3)$  are functions of the reduced-form parameters (see Appendix A.7). When  $\gamma_f$  is set such that the lag polynomial  $1 - \gamma_f L^{-1} - \gamma_1 L - \gamma_2 L^2 - \gamma_3 L^3$  has exactly one explosive root, equation (15) is the unique solution to model (16). For clarity, we base our discussion on this particular model, but note that the analysis generalizes. We can make the following observations.

First, the generalized forward-looking model (16) is obviously under-identified, since there is an infinity of observationally equivalent values for the structural parameters for different arbitrary choices of  $\gamma_f$ . For instance, (16) is consistent with no ‘forward-looking behavior’  $\gamma_f = 0$ . Consequently, as we discussed in section 3, the parameters in (16) cannot be estimated consistently, and GMM-based inference will be misleading.

In fact, by applying Proposition 3.1 we see that for model (16) to be identified, the labour share  $s_t$  must have at least fourth order autoregressive dynamics, or receive feedback from at least the fifth lag of inflation. This condition is more restrictive than for the simpler specification (4), see section 3.1.1. In other words, the more general the dynamic specification of the forward-looking model, the more restrictive is the condition on the dynamics of the forcing variable for identification.

Second, the generalized specification (16) shows that the GG model (4) is mis-specified because the variables  $s_{t-1}, s_{t-2}, \pi_{t-2}, \pi_{t-3}$  have been incorrectly omitted from that model. These variables remain in the error term of the GMM regression (5), which can be written as

$$e_t = \varepsilon_t - \gamma_f \eta_{t+1} + \gamma_2 \pi_{t-2} + \gamma_3 \pi_{t-3} + \lambda_1 s_{t-1} + \lambda_2 s_{t-2} = \varepsilon_t - \gamma_f \eta_{t+1} + \xi' \bar{Z}_t$$

where  $\bar{Z}_t$  is the instrument vector, that includes the variables  $s_{t-1}, s_{t-2}, \pi_{t-2}, \pi_{t-3}$  and  $\xi$  is a vector of parameters depending on the choice of instruments. For instance, when  $\bar{Z}_t$  consists of four lags of  $\pi_t$  and  $s_t$ ,  $\xi = (0, \gamma_2, \gamma_3, 0, \lambda_1, \lambda_2, 0, 0)'$ . It follows immediately that the GMM moment conditions are violated since  $E(\bar{Z}_t e_t) = E(\bar{Z}_t \bar{Z}_t') \xi \neq 0$  unless  $\xi = 0$ .

Third, the mis-specified model (4) is apparently identified in the sense that the rank condition for identification is satisfied. This can be seen clearly from the forecasting regression for  $\pi_{t+1}$  given  $\mathcal{F}_{t-1}$ :

$$E(\pi_{t+1}|\mathcal{F}_{t-1}) = \tilde{\delta}_1\pi_{t-1} + \tilde{\delta}_2\pi_{t-2} + \tilde{\delta}_3\pi_{t-3} + \tilde{\alpha}_1s_{t-1} + \tilde{\alpha}_2s_{t-2} \quad (17)$$

where  $\tilde{\delta}_i$  and  $\tilde{\alpha}_j$  are given in Appendix A.7. Equation (17) shows that the variables  $s_{t-1}, s_{t-2}, \pi_{t-2}, \pi_{t-3}$  that are incorrectly omitted from the model are relevant instruments for future inflation  $\pi_{t+1}$ . Thus, the rank condition for the identification of model (4) is satisfied (see Appendix A.7 for details).

This explains why the naive under-identification pretest reported in section 3.4 found evidence of identification. When applied to the correctly specified model (16), the result of that identification pretest becomes consistent with the structural identification analysis. Namely, the unrestricted reduced rank test (using the rk-statistic) does not reject the hypothesis that model (16) is under-identified.

The question that naturally arises is: how detectable is this type of mis-specification by a test of over-identifying restrictions? In principle, mis-specification is detectable insofar as the over-identifying restrictions of the model are violated. We can define a scalar measure of detectable mis-specification  $\nu^2$ , as the distance of the over-identifying restrictions from zero (see Appendix A.2). Like the concentration parameter  $\mu^2$ , the mis-specification parameter  $\nu^2$  is a function of the second moments of the data, and can only be computed when the parameters of the DGP are known. As we see below,  $\nu^2$  is useful in understanding the power of tests of over-identifying restrictions, such as Hansen’s (1982)  $J$  test.

Using the GMM estimator and data of GG, the p-value of the  $J$  test for the hybrid model (4) is 0.97. Thus, the mis-specification is not detected.<sup>12</sup> This motivates us to study the finite sample power of the  $J$  in this context, with particular emphasis on the implications of ‘over-instrumenting’ and ‘over-correction’ for serial correlation.

## 4.2 The power of the $J$ test

Under the null hypothesis that the over-identifying restrictions hold ( $\nu^2 = 0$ ), and provided the model is empirically identified ( $\mu^2$  large), the  $J$  test statistic follows approximately a chi-squared distribution with degrees of freedom equal to the degree of over-identification (Hansen 1982). Under the alternative hypothesis ( $\nu^2 > 0$ ), the power of the  $J$  test is driven, in principle, by the degree of mis-specification  $\nu^2$  and the degree of over-identification.

However, in finite samples, the power of the test may also be affected by the type of HAC estimator used to account for serial correlation and heteroscedasticity in the residuals. Different HAC estimators, albeit asymptotically equivalent, can differ substantially in finite samples, thus imparting substantial distortions to GMM-based inference (den Haan and Levin 1997).

We investigate the finite-sample power of the test by means of a Monte Carlo experiment. The DGP is described by the estimated reduced-form model for  $\pi_t$ , equation (15), and a first-order autoregression for  $s_t$ . We compare the power of four different versions of the test statistic: using two different numbers of instruments ( $k=8$  and  $24$ ); and two different types of HAC estimators to correct for serial correlation: MA- $l$  refers to the parametric estimator of West (1997) allowing only up to first-order residual autocorrelation, as suggested by the model, and HAC(12) refers to the nonparametric Newey and West (1987) estimator with lag-truncation parameter 12.

Table 3 reports the rejection frequencies of the test statistic at the 5% level, under this fixed alternative. Notably, the concentration and mis-specification parameters are increasing (linearly) with the sample size, reflecting stronger ‘identification’ and mis-specification of the model. Several conclusions can be drawn from those results.

Table 3: Rejection frequencies of Hansen’s (1982)  $J$  test of over-identifying restrictions for the New Keynesian Phillips Curve at the 5% level.

$T$	$\mu^2$	$\nu^2$	$k=8$			$k=24$		
			HAC(12)	MA- $l$	Asympt.	HAC(12)	MA- $l$	Asympt.
50	4.5	1.444	0.000	0.091	0.124	0.000	0.020	0.081
150	13.5	4.333	0.000	0.349	0.315	0.000	0.142	0.166
300	27.0	8.667	0.308	0.637	0.604	0.000	0.411	0.335
1000	90.0	28.888	0.988	0.994	0.994	0.602	0.947	0.928

This table reports the rejection frequencies of Hansen’s  $J$  test of over-identifying restrictions for the New Keynesian Phillips curve  $\pi_t = \lambda s_t + \gamma_f E(\pi_{t+1} | \mathcal{F}_t) + \gamma_b \pi_{t-1} + \epsilon_t$ . The data generating process is  $\pi_t = 0.68\pi_{t-1} + 0.24\pi_{t-3} + 0.55s_t - 0.48s_{t-2} + 1.4\epsilon_t$ , and  $s_t = 0.9s_{t-1} + v_t$ ;  $(\epsilon_t, v_t)$  are independent normally distributed innovations with variances  $\sigma_\epsilon^2 = 0.04$  and  $\sigma_v^2 = 0.01$ ;  $T$  denotes the sample size;  $\mu^2$  measures instrument relevance;  $\nu^2$  measures the degree to which the over-identifying restrictions are violated;  $k$  is the number of instruments, comprising lags of  $\pi_t$  and  $s_t$ . The  $J$  test is based on the 2-step GMM estimator. HAC(12) denotes the Newey and West (1987) HAC estimator with lag truncation 12; MA- $l$  denotes West’s (1997) HAC estimator. The critical value is the 95th percentile of the  $\chi^2$  distribution with  $k - 3$  degrees of freedom. Asymptotic power is based on a non-central  $\chi^2$  distribution with degrees of freedom  $k - 3$  and non-centrality parameter  $\nu^2$ .  $10^5$  Monte Carlo replications used.

First, the omitted dynamics contribute more towards spurious identification of the model ( $\mu^2 > 0$ ) than towards detectable mis-specification ( $\nu^2 > 0$ ). This is a significant factor explaining the low power of the test. Second, the addition of a large number of irrelevant instruments reduces the power of the test, as anticipated. Intuitively, the same degree of mis-specification  $\nu^2$  is diluted over many instruments, thus reducing the test’s power to detect it. In other words, ‘over-instrumenting’ unambiguously reduces power.

Third, we notice a marked difference in the behavior of the test using different autocorrelation corrections. The MA- $l$ -based test exhibits relatively small deviations from its asymptotic power, even at small samples. In contrast, the HAC(12)-based version suffers a severe downward bias in all cases. This is in line with extensive Monte Carlo evidence on the small sample properties of the various HAC estimators (e.g., den Haan and Levin 1997), and points out the potentially serious costs of ‘over-correction’ for serial correlation. A partial explanation of the poor performance of the HAC(12)-based test may be that the nonparametric HAC(12) estimator converges at a slower rate to the true asymptotic variance of the moment conditions than the parametric MA- $l$  estimator (see West 1997).<sup>13</sup>

## 5 Conclusion

In this paper, we studied the conditions for identification of a forward-looking rational expectations model, which are often overlooked when this model is estimated by limited information methods. Our analysis was based on the New Keynesian Phillips curve model of Galí and Gertler (1999) but the results have wider applicability. Our main findings can be summarized as follows.

The rank condition for identification is not sufficient for a forward-looking model to be reliably estimated. Even when the rank condition is satisfied, identification may be *weak*, depending on the nature of the dynamics of the forcing variables and the true values of the structural parameters. The possibility of weak identification cannot be ruled out *a priori*, and renders conventional GMM estimation completely unreliable. We demonstrated this in the context of the New Keynesian Phillips curve, where we found that when the model is weakly identified, GMM estimation will be biased in favor of a hybrid specification with apparently dominant forward-looking behavior, irrespective of the true nature of the forward and backward-looking dynamics of inflation.

If weak identification could be detected using a formal statistical test, then the problem exposed above would not be that severe. However, we show that existing tests for weak instruments (Stock and Yogo 2003) are not applicable to forward-looking models, and the development of a formal diagnostic test for weak identification in such models remains an important challenge for future research. A partial solution to the problem is to use inferential methods that are robust to weak identification, such as the tests proposed by Kleibergen (2001) and Stock and Wright (2000). These tests yield reliable inference even in the case of weak identification.

Standard identification pretests that are based on the unrestricted covariance between the endogenous regressors and the instruments can be very misleading because they have power even when the model is mis-specified. Dynamic mis-specification generates spurious identification and is only partly detectable by a test of the over-identifying restrictions such as Hansen's (1982)  $J$  test. Moreover, the common practice of using too many instruments and too general corrections for serial correlation seriously impairs the power of the  $J$  test in finite samples and obscures specification problems. Thus, looking at the reduced form of the complete system may prove a valuable alternative to the single-equation approach, as it may help uncover that mis-specification.

These findings have important implications for applied work. GMM should not be applied indiscriminately to forward-looking models and inference on those model should be conducted using methods that are robust to weak identification. Our analysis also suggests that full information methods, such as FIML, may be a preferable alternative to GMM (see Jondeau and Le Bihan 2003), since good use of GMM requires a structural identification analysis that is typical of the full information approach.

Finally, with regards to the estimation of the New Phillips curve of Galí and Gertler (1999) using US data, our analysis reveals the following. On the one hand, if we assume that the model is correctly specified, we find no evidence that it is empirically identified. On the other hand, if we allow for the possibility of dynamic mis-specification, we find evidence of spurious identification. These findings cast some doubts on the reliability of existing results, and suggest they need to be re-examined using methods that are robust to identification problems.

## Notes

<sup>1</sup> Note that the exogenous inflation shock  $\epsilon_t$  has been appended to the original model (GG Equation 26). Absence of this shock is implausible, since, amongst other things, it would imply that the resulting GMM regression residual,  $e_t$  in Equation (5), is serially uncorrelated, which is at odds with the data.

<sup>2</sup> This could be justified by measurement error in  $s_t$ , when the latter is a *proxy* for the true relationship being driven by marginal costs. In fact, the exogeneity status of  $s_t$  is inconsequential for the analysis of generic identification, as we will see below.

<sup>3</sup>Note that this threshold rises with the number of instruments.

<sup>4</sup> Note that the reduced-form parameters in the solution (9) are independent from the covariance between the structural error and the forcing variable,  $\sigma_{ve}$ , which determines whether the latter is endogenous or not in the model (7). Thus, proposition 3.1 holds irrespective of the endogeneity status of  $s_t$ .

<sup>5</sup>An example of this is the minimum state variable (MSV) solution, which has recently received renewed attention because it was shown to be E-stable under learning (see McCallum 2003).

<sup>6</sup>The parameter estimates reported by GG are such that  $\gamma_f, \gamma_b > 0$  and  $\gamma_f + \gamma_b < 1$ , in line with the underlying theory in their paper. These restrictions imply that the forward solution to the model is unique (see Appendix A.4).

<sup>7</sup>We are thankful to Ken West for this suggestion. When, instead,  $\rho_1$  is kept fixed to 0.9 and  $\rho_2$  varies between 0 and -0.8, the values of  $\mu^2$  are very similar to those reported in Table 1.

<sup>8</sup>So these results are not an artefact of  $\mu^2$  being practically zero. For instance, the results in the left panels of Figure 2 would be very similar even when  $\rho_2$  and  $\sigma_v/\sigma_\epsilon$  are set substantially higher so that  $\mu^2 = 1$ , say, instead of  $10^{-5}$ .

<sup>9</sup> The results are shown only for a sample of 300 observations for brevity. They are essentially the same when the sample size is 1000.

<sup>10</sup> Yet another example is parameter instability, which can be formulated as an omitted variable problem in the usual way. For instance, when  $\lambda$  changes over the sample,  $w_t = D_t s_t$  where  $D_t = 1$  in the subsample over which  $\lambda$  is different, and 0 otherwise.

<sup>11</sup> The case of other autocorrelated forcing variables being omitted from the model is similar to an autocorrelated structural error, since this error represents information known to the agents but not to the econometrician.

<sup>12</sup> This is also the case even for the pure forward-looking model (2), despite the fact the analysis in GG suggests that it is dynamically mis-specified, since some degree of backwardness ( $\gamma_b \neq 0$ ) is found to be statistically significant.

<sup>13</sup>Using the MA- $l$  HAC estimator, with five lags of inflation and  $s_t$  as instruments, the  $J$  test applied to the GG model (4) rejects the validity of the over-identifying restrictions at the 5% level of significance. However, this result is not to be taken at face value, since it is not robust to changes in the instrument set.

## A Appendix

### A.1 GG data description

Quarterly inflation rate:  $\pi_t = 100 \times \Delta \log(\text{GDP deflator})$ .

Labor share (in deviation from steady state):  $s_t = c \times 100 \log \frac{\text{unit labour cost}}{\text{unit price}}$ .

$c$  is a correction factor due to Sbordone (2002):  $c = n(m-1)/(m^2-n)$ , where  $n$  is the share of labour in the Cobb Douglas production function  $Y = A K^{1-n} L^n$ , and  $m$  is the average markup of prices over unit costs. GG set  $n = 2/3$  and  $m = 1.1$ , so  $c = 0.12$ .

## A.2 Some definitions

**The concentration parameter** Consider a prototype linear IV regression model  $y_t = \beta'Y_t + \gamma'X_t + u_t$ , where  $Y_t$  and  $X_t$  are endogenous and exogenous regressors respectively, with first-stage regression  $Y_t = \Pi'Z_t + \Phi'X_t + V_t$ . Suppose that  $X_t$ ,  $Z_t$  and  $V_t$  all have finite second moments denoted by  $\Sigma_{XX}$ ,  $\Sigma_{XZ}$ ,  $\Sigma_{ZZ}$  and  $\Sigma_{VV}$  and let  $\Omega = \Sigma_{ZZ} - \Sigma'_{XZ}\Sigma^{-1}_{XX}\Sigma_{XZ}$ . The concentration parameter matrix is  $T\Sigma_{VV}^{-1/2}\Pi'\Omega\Pi\Sigma_{VV}^{-1/2}$ , and its minimum eigenvalue  $\mu^2$  is the smallest root of the polynomial  $\det(T\Pi'\Omega\Pi - z\Sigma_{VV}) = 0$ , where  $\det(A)$  denotes the determinant of a matrix  $A$ .

**The 2-step GMM estimator** Stack the  $T$  observations on the endogenous and exogenous variables in the matrices  $y$ ,  $Y$ ,  $X$ , and  $Z$  of dimensions  $T \times 1$ ,  $T \times n$ ,  $T \times K_1$  and  $T \times K_2$  respectively, with  $K_2 \geq n$ . Let  $\bar{X} = (Y, X)$ ,  $\bar{Z} = (X, Z)$ ,  $\theta = (\beta', \gamma)'$ , and define the  $K = K_1 + K_2$  sample moments  $g_T(\theta) = \bar{Z}'(y - \bar{X}\theta)/T$ .

The 2-step efficient GMM estimator of  $\theta$  is defined as follows. Obtain an initial estimate of  $\theta$  by two stage least squares,  $\hat{\theta}_1 = (\bar{X}'P_{\bar{Z}}\bar{X})^{-1}\bar{X}'P_{\bar{Z}}y$ , where  $P_{\bar{Z}} = \bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'$ . Let  $W_T(\theta)$  denote the inverse of a consistent HAC estimator of the variance of  $g_T(\theta)$ . The 2-step efficient GMM estimator minimizes the objective function  $Q_T(\theta) = g_T(\theta)'W_T g_T(\theta)$ , where  $W_T = W_T(\hat{\theta}_1)$ , and is given by  $\hat{\theta} = (\bar{X}'\bar{Z}W_T\bar{Z}'\bar{X})^{-1}\bar{X}'\bar{Z}W_T\bar{Z}'y$ .

Assume that  $\bar{Z}'y/T \xrightarrow{p} \Sigma_{\bar{Z}y}$  and  $\bar{Z}'\bar{X}/T \xrightarrow{p} \Sigma_{\bar{Z}\bar{X}}$  where  $\xrightarrow{p}$  denotes convergence in probability. (This is satisfied in the GG model when  $\pi_t$  and  $s_t$  are weakly stationary). It follows that  $g_T(\theta) \xrightarrow{p} g(\theta) = \Sigma_{\bar{Z}y} - \Sigma_{\bar{Z}\bar{X}}\theta$ , uniformly in  $\theta$ . Whenever there exists a  $\theta_0$  such that the orthogonality conditions  $g(\theta_0) = 0$  hold, the model is correctly specified, and  $\theta_0$  is the true value of  $\theta$ . Otherwise, if there is no  $\theta$  that solves  $g(\theta) = 0$ , the orthogonality conditions are violated and the model is mis-specified. Obviously, for this to be possible there must be more instruments than endogenous regressors in the model.

**Pseudo-true values** To see what happens to  $\hat{\theta}$  when the model is mis-specified, assume that  $W_T(\theta) \xrightarrow{p} W(\theta)$  uniformly in  $\theta$ , where  $W(\theta)$  is a positive definite matrix. Then, provided  $\Sigma_{\bar{Z}\bar{X}}$  is of full rank, the GMM estimator  $\hat{\theta}$  converges to  $\theta^* = (\Sigma'_{\bar{Z}\bar{X}}W^*\Sigma_{\bar{Z}\bar{X}})^{-1}\Sigma'_{\bar{Z}\bar{X}}W^*\Sigma_{\bar{Z}y}$ , where  $W^* = W(\theta_1^*)$  and  $\theta_1^*$  is defined as  $\theta^*$  with  $W^*$  replaced by  $\Sigma_{\bar{Z}\bar{Z}}^{-1}$ . This  $\theta^*$  is the *pseudo-true* value of  $\theta$ , and differs for different HAC estimators  $W_T$ , or different choice of instruments  $Z$ . In contrast, when the model is correctly specified,  $\theta^* = \theta_1^* = \theta_0$ .

**Mis-specification parameter** Under the above assumptions,  $Q_T(\theta) \xrightarrow{p} Q(\theta) = g(\theta)'W(\theta)g(\theta)$ , uniformly in  $\theta$ . Whenever the moment conditions are not satisfied,  $Q(\theta) > 0$  for all  $\theta$ . Hansen's (1982)  $J$  test of over-identifying restrictions is based on the statistic  $TQ_T(\hat{\theta})$ , which, under the null hypothesis of correct specification, is asymptotically distributed as a chi-squared distribution with degrees of freedom equal to the number of over-identifying restrictions  $K_2 - n$ . Under the alternative  $Q_T(\hat{\theta}) \xrightarrow{p} Q(\theta^*) > 0$ , and the power of the test can be characterized by the scalar  $\nu^2 = TQ(\theta^*)$ , which we will refer to as the *mis-specification parameter*.

### A.3 Solution of model (7)

By a solution to the RE model (7) we mean a non-explosive distribution for  $\pi_t$  that satisfies equation (7). For convenience, we use the lag polynomial notation  $\gamma(L) = 1 - \gamma_1 L - \gamma_2 L^2 - \dots$ ,  $\lambda(L) = \lambda_0 + \lambda_1 L + \lambda_2 L^2 + \dots$ ,  $\rho(L) = 1 - \rho_1 L - \rho_2 L^2 - \dots$ , and  $\varphi(L) = \varphi_1 L + \varphi_2 L^2 + \dots$ , where  $L$  is the lag operator, such that  $Lx_t = x_{t-1}$ , and  $\gamma_i = 0$  for  $i > m$ ,  $\lambda_i = 0$  for  $i > n$ ,  $\rho_i = 0$  for  $i > p$  and  $\varphi_i = 0$  for  $i > q$ .

Recall the forecast error  $\eta_{t+1} = \pi_{t+1} - \pi_{t+1|t}$ , which is a MDS. Substitute  $\pi_{t+1} - \eta_{t+1}$  for  $\pi_{t+1|t}$  in (7), lag one period and then divide through by  $-\gamma_f$  and re-arrange to get

$$\left[1 - \gamma_f^{-1} \gamma(L) L\right] \pi_t = -\gamma_f^{-1} \lambda(L) L s_t - \gamma_f^{-1} \epsilon_{t-1} + \eta_t \quad (18)$$

For this reduced-form equation to be a solution to the RE model (7) it must not be explosive.

The stability conditions can be checked as follows. Premultiply equation (7) by  $\rho(L)$  and substitute  $\rho(L)s_t$  from (8) to get

$$\rho(L) \left[1 - \gamma_f^{-1} \gamma(L) L\right] \pi_t = -\gamma_f^{-1} \lambda(L) L [\varphi(L)\pi_t + v_t] - \gamma_f^{-1} \rho(L) \epsilon_{t-1} + \rho(L) \eta_t$$

Since  $v_t, \epsilon_t$  and  $\eta_t$  are innovation processes, the stability of the system is only determined by the characteristic roots of the autoregressive polynomial

$$\rho(z) (1 - z\gamma(z)/\gamma_f) + z\lambda(z)\varphi(z)/\gamma_f = 0. \quad (19)$$

By Blanchard and Kahn (1980), a solution exists if at most one root of equation (19) is explosive, i.e., lies inside the unit circle. If the polynomial (19) has no explosive roots, the RE model (7) has infinitely many solutions. To characterize this indeterminacy, the MDS  $\eta_t$  in the general solution (18) can be decomposed linearly by projection onto  $v_t$  and  $\epsilon_t$

$$\eta_t = \alpha_v v_t + \alpha_\epsilon \epsilon_t + \zeta_t \quad (20)$$

where  $\alpha_v, \alpha_\epsilon$  are free parameters and  $\zeta_t$  is an indeterminate MDS, uncorrelated with  $v_t$  and  $\epsilon_t$  by construction.  $\zeta_t$  is sometimes referred to as a sunspot shock. Then, using (20) in (18) and substituting for  $v_t \equiv s_t - s_{t|t-1}$  from (8), the solution(s) can be expressed as

$$\underbrace{\left[1 + \alpha_v \varphi(L) - \gamma_f^{-1} \gamma(L) L\right]}_{a(L)} \pi_t = \underbrace{\left[\alpha_v \rho(L) - \gamma_f^{-1} \lambda(L) L\right]}_{b(L)} s_t + \underbrace{(\alpha_\epsilon - \gamma_f^{-1} L)}_{c(L)} \epsilon_t + \zeta_t \quad (21)$$

This is an autoregressive distributed lag model whose residual can be represented as an unrestricted first-order moving average process (see Pesaran 1987). The lag polynomial  $a(L)$  is at most of order  $\max(q, m + 1)$ , since  $\varphi(L)$  is of order  $q$  and  $\gamma(L)$  is of order  $m$ . Similarly, the order of  $b(L)$  cannot exceed  $\max(p, n + 1)$ .

**Backward and forward solutions** We will refer to the class of solutions (21) as *backward solutions*. When the characteristic polynomial (19) has no explosive roots, there is an infinite number of possible backward solutions, since we can choose  $\alpha_\epsilon, \alpha_v$  and the MDS  $\zeta_t$  freely.

A special solution arises if we remove the sunspot shock  $\zeta_t$  and set  $\alpha_\epsilon$  and  $\alpha_v$  such that the lag polynomials  $a(L)$ ,  $b(L)$  and  $c(L)$  in (21) have a common factor. Let  $z_0$  be a root of (19) such that  $\rho(z_0) \neq 0$ . Then, it can be easily checked that

$$\alpha_\epsilon = \frac{z_0}{\gamma_f} \quad \text{and} \quad \alpha_v = \alpha_\epsilon \frac{\lambda(z_0)}{\rho(z_0)}, \quad (22)$$

imply that  $a(z_0) = b(z_0) = c(z_0) = 0$ , i.e., all three polynomials share the factor  $(1 - L/z_0)$ . Hence, we can find lag polynomials  $\delta(L) = 1 - \sum_{i=1}^{\kappa_\pi} \delta_i L^i$  and  $\alpha(L) = \sum_{j=0}^{\kappa_s} \alpha_j L^j$ , of orders  $\kappa_\pi = \max(q - 1, m)$  and  $\kappa_s = \max(p - 1, n)$  respectively, such that  $a(L) = \delta(L)(1 - L/z_0)$ ,  $b(L) = \alpha(L)(1 - L/z_0)$ . Substituting these expressions and  $c(L) = \alpha_\epsilon(1 - L/z_0)$  into (21) together with  $\zeta_t = 0$  and cancelling from both sides the common factor  $(1 - L/z_0)$ , we get the *forward solution* (9). The coefficients of the forward solution are derived by matching coefficients in the identities

$$\begin{aligned} 1 + \left( \alpha_v \varphi_1 - \frac{1}{\gamma_f} \right) L + \sum_{i=2}^{\kappa_\pi+1} \left( \alpha_v \varphi_i + \frac{\gamma_{i-1}}{\gamma_f} \right) L^i &\equiv \left( 1 - \frac{1}{z_0} L \right) \left( 1 - \sum_{i=1}^{\kappa_\pi} \delta_i L^i \right) \\ \alpha_v - \sum_{j=1}^{\kappa_s+1} \left( \alpha_v \rho_j + \frac{\lambda_{j-1}}{\gamma_f} \right) L^j &\equiv \left( 1 - \frac{1}{z_0} L \right) \sum_{j=0}^{\kappa_s} \alpha_j L^j. \end{aligned}$$

After some algebraic manipulations (details are in the unpublished additional appendix), they can be written recursively as follows:

$$\begin{aligned} \delta_1 &= 1/\gamma_f - 1/z_0 - \alpha_v \varphi_1, & \delta_i &= \delta_{i-1}/z_0 - \gamma_{i-1}/\gamma_f - \alpha_v \varphi_i, & i &= 2, \dots, \max(q - 1, m) \\ \alpha_0 &= \alpha_v, & \alpha_j &= \alpha_{j-1}/z_0 - \lambda_{j-1}/\gamma_f - \alpha_v \rho_j, & j &= 1, \dots, \max(p - 1, n). \end{aligned} \quad (23)$$

When the characteristic polynomial (19) has exactly one explosive root  $z_0$ , the forward solution is the *unique* solution to model (7).

#### A.4 Solution of the GG model (4)

When the forcing variable follows (11), the characteristic polynomial (19) simplifies to  $1 - z/\gamma_f - z^2 \gamma_b/\gamma_f = 0$ . Under the restrictions  $\gamma_f, \gamma_b \geq 0$  and  $\gamma_f + \gamma_b \leq 1$  implied by the underlying economic theory in GG, there is at most one explosive root  $z_0 = (1 - \sqrt{1 - 4\gamma_f \gamma_b})/(2\gamma_b)$ . Unless  $\gamma_f + \gamma_b = 1$  and  $\gamma_f > \gamma_b$ ,  $|z_0| < 1$  and the model (4) has the unique forward solution (12) (details are in the unpublished additional appendix).

**Coefficients of equation (12)** The solution parameters  $\alpha_0, \alpha_1, \delta_1$  and  $\alpha_\epsilon$  can be derived by applying the formulae (22) and (23), with  $m = 1, n = 0, p = 2, q = 0$  and  $\gamma_1 = \gamma_b$ . By straightforward algebraic manipulations, they can be written as:

$$\alpha_0 = \frac{\lambda z_0}{\gamma_f \rho(z_0)}, \quad \alpha_1 = \frac{\lambda \rho_2 z_0^2}{\gamma_f \rho(z_0)}, \quad \delta_1 = \frac{1 - \sqrt{1 - 4\gamma_f \gamma_b}}{2\gamma_f}, \quad \alpha_\epsilon = \frac{1}{1 - \delta_1 \gamma_f} \quad (24)$$

where  $z_0 = (1 - \sqrt{1 - 4\gamma_f \gamma_b})/(2\gamma_b)$  is the explosive root, derived above. By (23),  $\delta_1 = 1/\gamma_f - 1/z_0$ , so that  $z_0 = \gamma_f/(1 - \delta_1 \gamma_f)$ , and  $z_0 \neq 0$  for all  $\gamma_f \neq 0$ .



### Coefficients of equation (13)

$$\begin{aligned}\tilde{\alpha}_1 &= (\alpha_0\rho_1 + \alpha_1 + \delta_1\alpha_0)\rho_1 + \alpha_0\rho_2 + \delta_1\alpha_1, & \tilde{\alpha}_2 &= (\alpha_0\rho_1 + \alpha_1 + \delta_1\alpha_0)\rho_2, \\ \text{and } \tilde{\eta}_{t+1} &= \alpha_\epsilon\epsilon_{t+1} + \delta_1\alpha_\epsilon\epsilon_t + \alpha_0v_{t+1} + (\alpha_0\rho_1 + \alpha_1 + \delta_1\alpha_0)v_t\end{aligned}\quad (25)$$

### A.5 The concentration parameter in the GG model

$\mu^2$  is defined in section A.2, above. In the example of section 3.2,  $\Pi$ ,  $\Sigma_{VV}$  and  $\Omega$  are as follows. From equation (14)

$$\Pi = \begin{pmatrix} \tilde{\alpha}_1 & \rho_1 \\ \tilde{\alpha}_2 & \rho_2 \end{pmatrix}. \quad (26)$$

From the definition of  $\tilde{\eta}_{t+1}$  in (25),  $V_t = A(\epsilon_t, v_t)' + B(\epsilon_{t+1}, v_{t+1})'$ , so that  $\Sigma_{VV} = A\Sigma A' + B\Sigma B'$ , where

$$A = \begin{pmatrix} \delta_1\alpha_\epsilon & \alpha_0(\delta_1 + \rho_1) + \alpha_1 \\ 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} \alpha_\epsilon & \alpha_0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_\epsilon^2 & \sigma_{v\epsilon} \\ \sigma_{v\epsilon} & \sigma_v^2 \end{pmatrix}.$$

Finally,  $\Omega$  is defined above as  $\Sigma_{ZZ} - \Sigma_{ZX}\Sigma_{XX}^{-1}\Sigma'_{ZX}$ , and since  $Z_t = (s_{t-1}, s_{t-2})'$  and  $X_t = \pi_{t-1}$ , and assuming  $\pi_t, s_t$  are weakly stationary,

$$\Sigma_{ZZ} = \begin{pmatrix} E(s_t^2) & E(s_t s_{t-1}) \\ E(s_t s_{t-1}) & E(s_{t-1}^2) \end{pmatrix} \quad \Sigma_{ZX} = \begin{pmatrix} E(s_t \pi_t) \\ E(s_{t-1} \pi_t) \end{pmatrix} \quad \text{and} \quad \Sigma_{XX} = E(\pi_t^2).$$

$\mu^2 = 0$  occurs if and only if  $\Pi$  is of reduced rank.  $\Pi$  is of reduced rank if and only if its determinant is zero. From equation (26),  $\det(\Pi) = \tilde{\alpha}_1\rho_2 - \rho_1\tilde{\alpha}_2 = \alpha_1\rho_2 + \delta_1\alpha_2 = \lambda\rho_2 z_0(1 + \delta_1 z_0)/(\gamma_f \rho(z_0))$ , where the second equality follows by substituting for  $\tilde{\alpha}_1$  and  $\tilde{\alpha}_2$  from equation (25), and the third by substituting for  $\alpha_0$  and  $\alpha_1$  from equation (24). Also,  $1 + \delta_1 z_0 = 1 + \delta_1 \gamma_f / (1 - \delta_1 \gamma_f) = 1/(1 - \delta_1 \gamma_f) \neq 0$ . So,  $\det(\Pi) = 0$  if and only if  $\rho_2 = 0$  or  $\lambda = 0$ .

**The concentration parameter in a model without lagged inflation** If  $\gamma_b$  is set to 0 in (5), and  $s_t$  is assumed to be exogenous, then  $\mu^2$  can be solved analytically:

$$\mu^2 = \frac{a_2^2 \sigma_v^2 / \sigma_\epsilon^2}{(1 - \rho_2^2) (a_1^2 \sigma_v^2 / \sigma_\epsilon^2 + 2a_1 \sigma_{v\epsilon} / \sigma_\epsilon^2 + 1)}, \quad (27)$$

where  $a_1 = \lambda(\rho_1 + \gamma_f \rho_2)/(1 - \gamma_f \rho_1 - \gamma_f^2 \rho_2)$  and  $a_2 = \lambda\rho_2/(1 - \gamma_f \rho_1 - \gamma_f^2 \rho_2)$  [adapted from Mavroeidis 2004, Equation (18)]. It follows that  $\mu^2 = 0$  if and only if  $\lambda = 0$  or  $\rho_2 = 0$ ,  $\mu^2$  is homogeneous in  $\sigma_\epsilon^2, \sigma_v^2, \sigma_{v\epsilon}$ , and is increasing in  $\sigma_v/\sigma_\epsilon$  when  $\sigma_{v\epsilon} = 0$ .

**Effect of higher order dynamics on  $\mu^2$**  If the completing model (11) is generalized to an AR(3), then  $s_{t-3}$  becomes a relevant instrument. The direct impact of this on  $\mu^2$  is positive. However, the change of  $\rho_3$  will alter the dynamics of the system through the solution to the model. The matrices  $\Pi$ ,  $\Sigma_{ZZ}$  and  $\Sigma_{ZX}$  will be altered, and it is possible that the contributions of the other two instruments  $s_{t-1}, s_{t-2}$  to  $\mu^2$  is reduced. Hence, the overall effect of the change in  $\rho_3$  is ambiguous.

## A.6 Simulation settings

If  $\rho_2 = 0$  in (11), the model (4) is under-identified (Proposition 3.1). The reduced form is given by (12) with  $\delta_1$  and  $\alpha_\epsilon$  as in (24) but  $\alpha_1 = 0$  and  $\alpha_0 = \lambda/(1 - \gamma_f(\delta_1 + \rho_1))$ . The reduced form parameters are  $(\alpha_0, \delta_1, \sigma_u)$  where  $\sigma_u^2$  denotes the variance of the reduced form residual  $u_t = \alpha_\epsilon \epsilon_t$ . There are infinite possible combinations of  $(\lambda, \gamma_f, \gamma_b, \sigma_\epsilon)$  consistent with the same  $(\alpha_0, \delta_1, \sigma_u)$ . If we treat  $\alpha_0, \delta_1$  and  $\sigma_u$  as given, for any choice of  $\gamma_f$ , the remaining structural parameters must satisfy  $\gamma_b = \delta_1(1 - \gamma_f\delta_1)$ ,  $\lambda = \alpha_0(1 - \gamma_f(\delta_1 + \rho_1))$  and  $\sigma_\epsilon = (1 - \gamma_f\delta_1)\sigma_u$ . It is easy to check that the following two values  $(\lambda, \gamma_f, \gamma_b, \sigma_\epsilon) = (0.015, 0.59, 0.38, 0.2)$  and  $(0.114, 0, 0.57, 0.3)$ , yield identical reduced form dynamics for  $\pi_t$ .

For the simulations reported in Figure 2, the nuisance parameters  $\rho_2, \sigma_v, \sigma_{v\epsilon}$  are set as follows. For weak identification, they are calibrated to the GG data. For strong identification, for setting 1 (new Phillips curve), they are set as in Table 2 and Figure 1. For setting 2 (old Phillips curve), they are set slightly differently so as to ensure a similar degree of empirical identification ( $\mu^2$ ) and similar variability in the estimators to facilitate the comparison.

## A.7 Models of section 4.1

**OLS estimates for equation (15)**  $\hat{\alpha}_0 = 0.55$  (0.16),  $\hat{\alpha}_2 = 0.48$  (0.16),  $\hat{\delta}_1 = 0.68$  (0.06),  $\hat{\delta}_3 = 0.24$  (0.06), and  $\hat{\sigma}_u = 0.26$ . (standard errors in parentheses)

**Coefficients of equation (16)**  $\lambda = \alpha_0[1 - \gamma_f(\delta_1 + \rho_1)]$ ,  $\lambda_1 = -\alpha_2\gamma_f$ ,  $\lambda_2 = \alpha_2(1 - \gamma_f\delta_1)$ ,  $\gamma_1 = \delta_1(1 - \gamma_f\delta_1)$ ,  $\gamma_2 = -\delta_3\gamma_f$ ,  $\gamma_3 = \delta_3(1 - \gamma_f\delta_1)$  and  $\epsilon_t = (1 - \gamma_f\delta_1)u_t$ .

**Derivation of equation (17)** This is derived from Equation (15) and  $s_t = \rho_1 s_{t-1} + v_t$ . The coefficients are  $\tilde{\delta}_1 = \delta_1^2$ ,  $\tilde{\delta}_2 = \delta_3$ ,  $\tilde{\delta}_3 = \delta_1\delta_3$ ,  $\tilde{\alpha}_1 = (\alpha_0(\rho_1 + \delta_1) + \alpha_2)\rho_1$ ,  $\tilde{\alpha}_2 = \delta_1\alpha_2$ .

**Spurious identification** Equation (17) together with  $s_t = \rho_1 s_{t-1} + v_t$  suffice to derive the first-stage regression for the model (4), in the form (14). Apart from the exogenous regressor  $X_t = \pi_{t-1}$ , the other relevant instruments are  $Z_t = (\pi_{t-2}, \pi_{t-3}, s_{t-1}, s_{t-2})$ . The coefficient matrix

$$\Pi' = \begin{pmatrix} \tilde{\delta}_2 & \tilde{\delta}_3 & \tilde{\alpha}_1 & \tilde{\alpha}_2 \\ 0 & 0 & \rho_1 & 0 \end{pmatrix}$$

is clearly of full rank 2, so the rank condition for identification is satisfied. This would not be true if the model (4) was correctly specified, because in that case  $\tilde{\delta}_2 = \tilde{\delta}_3 = \tilde{\alpha}_2 = 0$ .

## References

- Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Statistics* 22, 327–351.
- Batini, N., B. Jackson, and S. Nickell (2000). Inflation dynamics and the labour share in the UK. Discussion paper 2, External MPC Unit, Bank of England, UK.

- Binder, M. and M. H. Pesaran (1995). Multivariate rational expectations models: A review and some new results. In M. H. Pesaran and M. R. Wickens (Eds.), *Handbook of applied econometrics*, Volume Macroeconomics, pp. 139–187. Oxford: Blackwell.
- 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.
- Clarida, R., J. Galí, and M. Gertler (1998). Monetary policy rules in practice: Some international evidence. *European Economic Review* 42, 1033–67.
- 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.
- Fuhrer, J. C. and G. R. Moore (1995). Inflation persistence. *Quarterly Journal of Economics* 110, 127–159.
- Galí, J. and M. Gertler (1999). Inflation dynamics: a structural econometric analysis. *Journal of Monetary Economics* 44, 195–222.
- Galí, J., M. Gertler, and J. D. López-Salido (2001). European inflation dynamics. *European Economic Review* 45, 1237–1270.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.
- Hansen, L. P. and T. J. Sargent (1991). *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. (2001). Testing parameters in GMM without assuming that they are identified. Discussion paper, Tinbergen Institute, Amsterdam.
- Kleibergen, F. and R. Paap (2003). Generalized reduced rank tests using the Singular Value Decomposition. Discussion paper 003/4, Tinbergen Institute, Amsterdam.
- Lubik, T. A. and F. Schorfheide (2004). Testing for indeterminacy: An application to U.S. monetary policy. *American Economic Review* 94(1), 190–216.
- Lucas, R. E. J. (1976). Econometric policy evaluation: a critique. In K. Brunner and A. Meltzer (Eds.), *The Phillips Curve and Labor Markets.*, Carnegie-Rochester Conference Series on Public Policy. Amsterdam: North-Holland.
- Mavroeidis, S. (2004). Weak identification of forward-looking models in monetary economics. *Oxford Bulletin of Economics and Statistics* 66(Supplement), 609–635.
- McCallum, B. T. (2003). The unique Minimum State Variable RE solution is E-stable in all well formulated linear models. NBER working paper 9960, NBER, USA.
- 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. (1981). Identification of rational expectations models. *Journal of Econometrics* 16(3), 375–398.
- Pesaran, M. H. (1987). *The limits to Rational Expectations*. Oxford: Blackwell Publishers.
- Sbordone, A. M. (2002). Prices and unit labor costs: a new test of price stickiness. *Journal of Monetary Economics* 49, 265–292.
- Smets, F. and R. Wouters (2003). An estimated Dynamic Stochastic General Equilibrium model for the Euro area. Working paper, European Central Bank.
- Staiger, D. and J. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.
- 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.

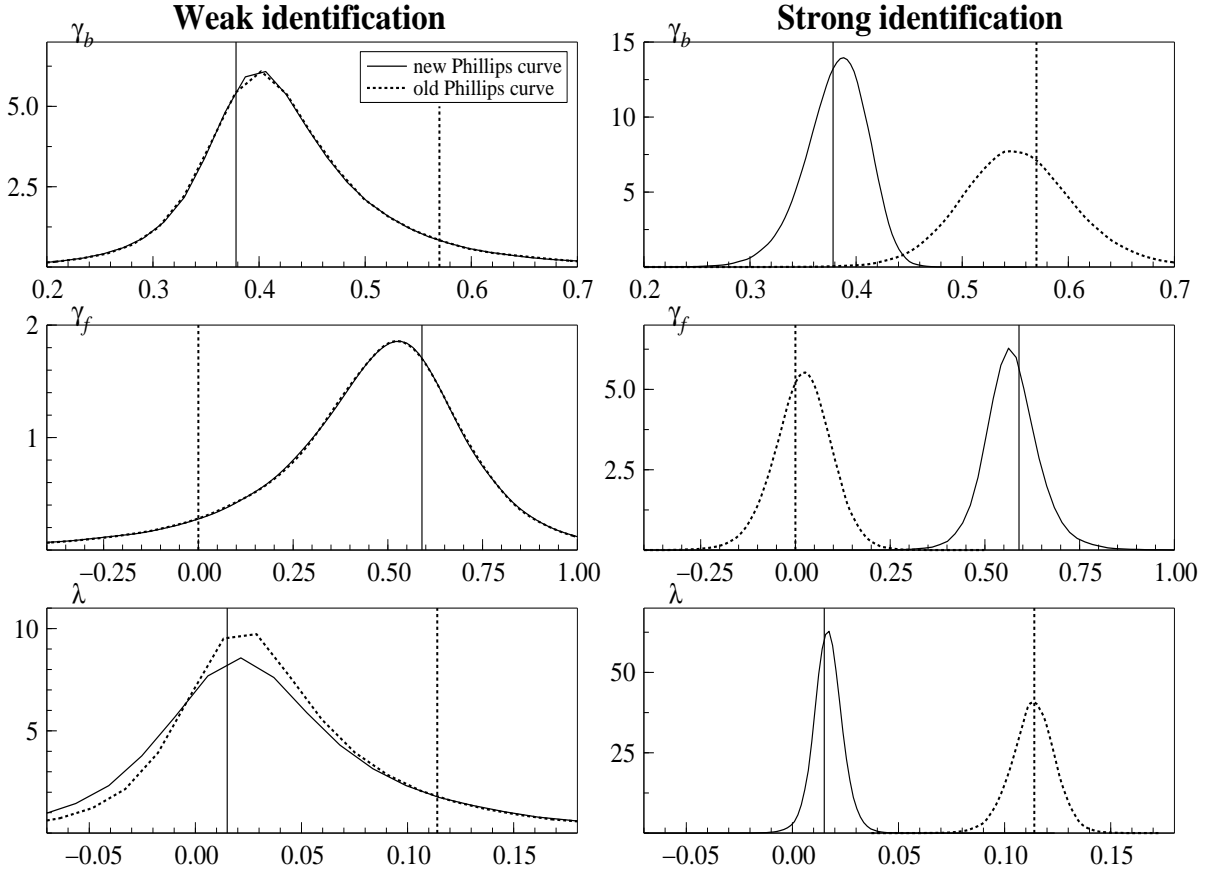


Figure 2: Comparison of the finite sample densities of the 2-step GMM estimator for the parameters of the New Keynesian Phillips Curve  $\pi_t = \lambda s_t + \gamma_f \mathbf{E}(\pi_{t+1} | \mathcal{F}_t) + \gamma_b \pi_{t-1} + \epsilon_t$ . The sample size is 300, the instrument set contains four lags of  $\pi_t$  and  $s_t$ , and a 12-lag Newey-West (1987) estimate of the covariance matrix was used. Data is generated using  $s_t = 0.9(1 - \rho_2)s_{t-1} + \rho_2 s_{t-2} + v_t$ ;  $(\epsilon_t, v_t)$  are normally distributed innovations with variances  $\sigma_\epsilon^2$  and  $\sigma_v^2$  respectively and covariance  $\sigma_{v\epsilon}$ . Each panel plots two densities for the respective parameter estimator when data is simulated from two different settings. Setting 1 (solid curves) is a new Phillips curve with  $(\lambda, \gamma_f, \gamma_b, \sigma_\epsilon) = (0.015, 0.59, 0.38, 0.2)$  (these replicate the  $T = 300$  curves in Figure 1). setting 2 (dotted curves) is an ‘old’ Phillips curve with  $(\lambda, \gamma_f, \gamma_b, \sigma_\epsilon) = (0.11, 0, 0.57, 0.3)$ . Left and right panels differ in the specification of  $(\rho_2, \sigma_v, \sigma_{v\epsilon})$ : left panels (weak identification)  $(-0.05, 0.1, 0)$ ; right panels (strong identification)  $(-0.65, 1.6, 0)$  for setting 1 and  $(0.65, 4.8, 0.14)$  for setting 2. Vertical lines show the true values of the parameters in each setting.