

# Weak instrument robust tests in GMM and the new Keynesian Phillips curve

Frank Kleibergen\*  
Brown University

Sophocles Mavroeidis†  
Brown University

September 26, 2008

## Abstract

We discuss weak instrument robust statistics in GMM for testing hypotheses on the full parameter vector or on subsets of the parameters. We use these test procedures to re-examine the evidence on the new Keynesian Phillips curve model. We find that US postwar data are consistent with the view that inflation dynamics are predominantly forward-looking, but we cannot rule out the presence of considerable backward-looking dynamics. Moreover, the Phillips curve has become flatter recently, and this is an important factor contributing to its weak identification.

## 1 Introduction

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

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

where  $\pi_t$  denotes inflation,  $x_t$  is some proxy for marginal costs,  $E_t$  denotes the expectation conditional on information up to time  $t$ , and  $u_t$  is an unobserved cost-push shock. This model can be derived from microfoundations in a dynamic general equilibrium framework with price stickiness, a la Calvo (1983) and indexation, see Woodford (2003). Variants of equation (1) appear in many studies on macroeconomic dynamics and monetary policy, see *e.g.*, Lubik and Schorfheide (2004) and Christiano *et. al.* (2005). Equation (1) is usually referred to as the ‘semi-structural’ specification corresponding to a deeper microfounded structural model. The various structural specifications proposed in the literature share essentially the same semi-structural form, but their underlying deep parameters, which are functions of  $\lambda$ ,  $\gamma_f$  and  $\gamma_b$ , differ. Hence, we chose to focus our discussion mainly on the

---

\*Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, United States. Email: Frank\_Kleibergen@brown.edu. Homepage: [http://www.econ.brown.edu/fac/Frank\\_Kleibergen](http://www.econ.brown.edu/fac/Frank_Kleibergen).

†Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, United States. Email: Sophocles\_Mavroeidis@brown.edu. Homepage: [http://www.econ.brown.edu/fac/Sophocles\\_Mavroeidis](http://www.econ.brown.edu/fac/Sophocles_Mavroeidis).

semi-structural specification, which is more general, but we also present empirical results for a popular structural version of the model.

In a seminal paper, Galí and Gertler (1999) estimated a version of this model in which the forcing variable  $x_t$  is the labor share and the parameters  $\lambda, \gamma_f, \gamma_b$  are functions of three key structural parameters: the fraction of backward-looking price-setters, the average duration an individual price is fixed (the degree of price stickiness) and a discount factor. Using postwar data on the U.S., Galí and Gertler (1999) reported that real marginal costs are statistically significant and inflation dynamics are predominantly forward-looking. They found  $\gamma_b$  to be statistically significant but quantitatively small relative to  $\gamma_f$ .

Several authors have argued that the above results are unreliable because they are derived using methods that are not robust to identification problems, also known as weak instrument problems, see Canova and Sala (2005), Mavroeidis (2005) and Nason and Smith (2005). As we explain below, weak instrument problems arise if marginal costs have limited dynamics or if their coefficient is close to zero, i.e., when the NKPC is flat, since in those cases the requisite exogenous variation in inflation forecasts is limited. Moreover, the weak instruments literature (cf. Andrews and Stock (2005), Dufour (2003), Kleibergen (2007) and Stock *et al* (2002)) has shown that using conventional inference methods after pretesting for identification is both unreliable and unnecessary. It is unreliable because the size of such two-step testing procedures cannot be controlled. It is unnecessary because there are identification-robust methods that are as powerful as the non-robust methods when instruments are strong and more powerful than the aforementioned two-step procedures when instruments are weak. For example, we show that when the instruments are weak, the commonly used pre-test rule advocated by Stock and Yogo (2005) to only use two stage least squares t statistics when the first-stage F statistic exceeds ten has less power than the identification-robust statistic of Anderson and Rubin (1949).

Unfortunately, the use of identification-robust methods has not yet become the norm in this literature. To the best of our knowledge, the studies that did use identification-robust methods are Ma (2002), Kichian *et al* (2006), Nason and Smith (2005) and Martins and Gabriel (2006), and their results suggest that the NKPC is weakly identified. In particular, they could not find evidence in favor of the view that forward-looking dynamics are dominant, nor that the labor share is a significant driver of inflation. These results seem to confirm the criticism of Mavroeidis (2005) regarding the poor identifiability of the NKPC.

In this paper, we discuss the various identification-robust methods that can be used to conduct inference on the parameters of the NKPC, with particular emphasis on the problem of inference on subsets of the parameters. Our discussion of the theoretical econometrics literature, as well as the results of the Monte Carlo simulations that we conducted for the NKPC, lead us to recommend as the preferred method of inference the generalized method of moments (GMM) extension of the conditional likelihood ratio (CLR) statistic. The CLR statistic was proposed by Moreira (2003) for the linear instrumental variables regression model with one included endogenous variable and it was later extended to GMM by Kleibergen (2005). We refer to this extension as the MQLR statistic. The MQLR is at least as powerful as any of the other tests, and yields the smallest confidence sets in our empirical application. We also provide an efficient method for computing a version of the MQLR statistic which has some appealing numerical properties.

Our empirical analysis is based on quarterly postwar US data from 1960 to 2007. We obtain one and two dimensional confidence sets derived by inverting each of the identification-robust statistics

for the key parameters of the NKPC, and our results for the full sample can be summarized as follows. In accordance with Galí and Gertler (1999), we find evidence that forward-looking dynamics in inflation are statistically significant and that they dominate backward-looking dynamics. However, the confidence intervals are so wide as to be consistent both with no backward-looking dynamics, as well as very substantial backward-looking behavior. Moreover, we cannot reject the null hypothesis that the coefficient on the labor share is zero. We also test the deeper structural parameters in the model of Galí and Gertler (1999), which characterize the degree of price stickiness and the fraction of backward-looking price setters. With regards to price stickiness, the 95% level confidence interval, though unbounded from above, is still informative, because it suggests that prices remain fixed for a least two quarters, thus uncovering evidence of significant price rigidity. However, regarding the fraction of backward-looking price setters, the 95% confidence interval is very wide, suggesting that the data are consistent both with the view that price setting behavior is purely forward-looking, as well as with the opposite view that it is predominantly backward-looking.

Finally, we conduct an identification-robust structural stability test proposed by Caner (2007) and find evidence of instability before 1984, but no such evidence thereafter. This shows that the model is not immune to the Lucas (1976) critique. When we split the sample in half to pre and post 1984 periods, we find that the slope of the NKPC is markedly smaller in the second half of the sample, lending some support to the view that the Phillips curve is now flatter than it used to be. This finding may have useful implications for the conduct of monetary policy.

The structure of the paper is as follows. In Section 2, we introduce the model and discuss the main identification issues. Section 3 presents the relevant econometric theory for the identification-robust tests. Monte Carlo simulations on the size and power of those tests for the NKPC are reported in Section 4, and the results of the empirical analysis are given in Section 5. Section 6 discusses some directions for future research, and Section 7 concludes. Analytical derivations are provided in an Appendix at the end.

Throughout the paper, we use the following notation:  $I_m$  is the  $m \times m$  identity matrix,  $P_A = A(A'A)^{-1}A'$  for a full rank  $n \times m$  matrix  $A$  and  $M_A = I_n - P_A$ . Furthermore, “ $\xrightarrow{p}$ ” stands for convergence in probability, “ $\xrightarrow{d}$ ” for convergence in distribution,  $E$  is the expectation operator and  $E_t$  denotes expectations conditional on information available at time  $t$ .

## 2 Identification of the New Keynesian Phillips Curve

The parameters of the NKPC model are typically estimated using a limited information method that replaces the unobserved term  $E_t(\pi_{t+1})$  in equation (1) by  $\pi_{t+1} - \eta_{t+1}$ , where  $\eta_t$  is the one-step ahead forecast error in  $\pi_t$ , to obtain the estimable equation

$$\pi_t = \lambda x_t + \gamma_f \pi_{t+1} + \gamma_b \pi_{t-1} + e_t, \quad (2)$$

with  $e_t = u_t - \gamma_f \eta_{t+1}$ . The assumption  $E_{t-1}(u_t) = 0$  implies that  $E_{t-1}(e_t) = 0$ , so equation (2) can be estimated by GMM using any predetermined variables  $Z_t$  as instruments. The moment conditions are given by  $E(f_t(\theta)) = 0$ , where  $f_t(\theta) = Z_t(\pi_t - \lambda x_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1})$  and  $\theta = (\lambda, \gamma_f, \gamma_b)$ . Note that since  $e_t$  is not adapted to the information at time  $t$ , it does not follow from the above assumptions that  $E(e_t e_{t-1}) = 0$ , and hence,  $e_t$  may exhibit first-order autocorrelation without contradicting the

model. Moreover,  $e_t$  may also be heteroskedastic. Thus equation (2) does not fit into the framework of the linear instrumental variables (IV) regression model with independently and identically distributed (iid) data. Nonetheless, the tools used to study identification in the linear IV regression model can be used to discuss the relevant identification issues for the NKPC.

We first consider the rank condition for identification of the parameters in (2). The parameters  $\theta$  are identified if the Jacobian of the moment conditions  $E(\frac{\partial f_t(\theta)}{\partial \theta'})$  is of full rank. Equation (2) has two endogenous regressors,  $\pi_{t+1}$  and  $x_t$ , and one exogenous regressor  $\pi_{t-1}$ . If  $\Pi_1 \pi_{t-1} + \Pi_2 Z_{2,t}$  denotes the linear projection of the endogenous regressors  $(\pi_{t+1}, x_t)$  on the instruments  $Z_t = (\pi_{t-1}, Z'_{2,t})'$ , the rank condition for identification is that the rank of  $\Pi_2$  is equal to two since  $\pi_{t-1}$  is included as an exogenous regressor in equation (2).

To study the rank condition for identification, we need to model the reduced-form dynamics of  $\pi_t$  and  $x_t$ . This can be done by postulating a model for the forcing variable,  $x_t$ , and solving the resulting system of equations to obtain the restricted reduced-form model for the joint law of motion of  $(\pi_t, x_t)$ . This enables us to express the coefficient matrix  $\Pi_2$  in the projection of the endogenous regressors  $(\pi_{t+1}, x_t)$  on the instruments  $Z_{2,t}$  as a function of the structural parameters. Because the rank condition is intractable in general, see Pesaran (1987, ch. 6), it is instructive to study a leading special case in which analytical derivations are straightforward, and which suffices to provide the main insights into the identification problems.

Consider the purely forward-looking version of the model (1) with  $\gamma_b = 0$ , and assume that  $x_t$  is stationary and follows a second-order autoregression. The complete system is given by the equations:

$$\pi_t = \lambda x_t + \gamma_f E_t(\pi_{t+1}) + u_t \quad (3)$$

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + v_t. \quad (4)$$

Solving equation (3) forward, we obtain

$$\pi_t = \lambda \sum_{j=0}^{\infty} \gamma_f^j E_t(x_{t+j}) + u_t = \alpha_0 x_t + \alpha_1 x_{t-1} + u_t, \quad (5)$$

with  $\alpha_0 = \lambda / [1 - \gamma_f (\rho_1 + \gamma_f \rho_2)]$  and  $\alpha_1 = \lambda \gamma_f \rho_2 / [1 - \gamma_f (\rho_1 + \gamma_f \rho_2)]$ .<sup>1</sup> Substituting for  $x_t$  in (5) using (4), and evaluating the resulting expression at time  $t + 1$ , we obtain the first-stage regression for the endogenous regressors  $\pi_{t+1}, x_t$

$$\begin{pmatrix} E_{t-1}(\pi_{t+1}) \\ E_{t-1}(x_t) \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_0 ((\rho_1 + \rho_2 \gamma_f) \rho_1 + \rho_2) & \alpha_0 (\rho_1 + \rho_2 \gamma_f) \rho_2 \\ \rho_1 & \rho_2 \end{pmatrix}}_{\Pi} \underbrace{\begin{pmatrix} x_{t-1} \\ x_{t-2} \end{pmatrix}}_{Z_t}. \quad (6)$$

It is then straightforward to show that the determinant of the coefficient matrix  $\Pi$  in the above expression is proportional to  $\alpha_0 \rho_2^2$ , and hence, since  $\alpha_0$  is proportional to  $\lambda$ , the rank condition for identification is satisfied if and only if  $\lambda \neq 0$  and  $\rho_2 \neq 0$ . Thus, identification requires the presence of second order dynamics and that  $\lambda$  exceeds zero (since economic theory implies  $\lambda \geq 0$ ).

The rank condition is, however, not even sufficient for reliable estimation and inference because

---

<sup>1</sup>The solution can be verified by computing  $E_t(\pi_{t+1})$ , substituting into the model (1) and matching the coefficients on  $x_t$  and  $x_{t-1}$  in the resulting equation to  $\alpha_0$  and  $\alpha_1$  in (5).

of the problem of weak instruments, see *e.g.* Stock *et. al.* (2002). Even when the rank condition is satisfied, instruments can be weak for a wide range of empirically relevant values of the parameters. Loosely speaking, instruments are weak whenever there is a linear combination of the endogenous regressors whose correlation with the instruments is small relative to the sample size. More precisely, the strength of the instruments is characterized in linear IV regression models by a unitless measure known as the concentration parameter, see *e.g.* Phillips (1983) and Rothenberg (1984). The concentration parameter is a measure of the variation of the endogenous regressors that is explained by the instruments, after controlling for any exogenous regressors, relative to the variance of the residuals in the first-stage regression, i.e., a multivariate signal-to-noise ratio in the first-stage regression. For the first-stage regression given in equation (6), the concentration parameter can be written as  $T\Sigma_{VV}^{-1/2}\Pi'E(Z_tZ_t')\Pi\Sigma_{VV}^{-1/2}$ , where  $\Sigma_{VV}$  is the variance of first-stage regression residuals and  $T$  is the sample size. This is a symmetric positive semi-definite matrix whose dimension is equal to the number of endogenous regressors. The interpretation of the concentration parameter is most easily given in a model with a single endogenous regressor in terms of the so-called first-stage  $F$  statistic that tests the rank condition for identification. Under the null hypothesis of no identification, i.e., when the instruments are completely irrelevant, the expected value of the first-stage  $F$  statistic is equal to 1, while under the alternative it is greater than 1. The concentration parameter, divided by the number of instruments,  $\mu^2/k$ , is then approximately equal to  $E(F) - 1$ . In the case of  $m$  endogenous regressors, the strength of the instruments can be measured by the smallest eigenvalue of the concentration matrix, which we shall also denote by  $\mu^2$ , to economize on notation.

Even in the above simple model which has only two endogenous regressors,  $\pi_{t+1}$  and  $x_t$ , the analytical derivation of the smallest eigenvalue of the concentration matrix,  $\mu^2$ , is impractical. Thus, we shall consider a special case in which the model has a single endogenous regressor, so that  $\mu^2$  can be derived analytically.<sup>2</sup> This special case arises when we assume that  $E(v_t u_t) = E(x_t u_t) = 0$ , so that the regressor  $x_t$  in equation (3) becomes exogenous, and the only endogenous regressor is  $\pi_{t+1}$ . Hence, the first-stage regression is given by  $E_t(\pi_{t+1})$ . From the law of motion of  $\pi_t, x_t$  given by equations (5) and (4), we obtain that  $E_t(\pi_{t+1}) = \alpha_0 E_t(x_{t+1}) + \alpha_1 x_t = \alpha_0 \rho_1 x_t + \alpha_0 \rho_2 x_{t-1} + \alpha_1 x_t$ , and hence, the first-stage regression can be written equivalently as

$$\pi_{t+1} = (\alpha_0 \rho_1 + \alpha_1) x_t + \alpha_0 \rho_2 x_{t-1} + \eta_{t+1}, \quad (7)$$

where  $\eta_t = u_t + \alpha_0 v_t$  is the one-step innovation in  $\pi_t$ . To simplify the derivations, assume further that  $u_t, v_t$  are jointly normally distributed and homoskedastic. Since  $E(u_t v_t) = 0$ , it follows that  $E(v_t \eta_t) \equiv \rho_{\eta v} \sigma_v \sigma_\eta = \alpha_0 \sigma_v^2$ , or  $\alpha_0 = \rho_{\eta v} \sigma_\eta / \sigma_v$ , where  $\sigma_\eta^2 = E(\eta_t^2)$ ,  $\sigma_v^2 = E(v_t^2)$  and  $\rho_{\eta v} = E(\eta_t v_t) / (\sigma_v \sigma_\eta)$ . Recall that  $x_t$  is the exogenous regressor, and  $x_{t-1}$  is the additional (optimal) instrument. The parameter that governs identification is therefore the coefficient of  $x_{t-1}$  in the first-stage regression (7), i.e.,  $\alpha_0 \rho_2$  or  $\rho_{\eta v} \rho_2 \sigma_\eta / \sigma_v$ . The expression of the concentration parameter corresponds to  $T(\alpha_0 \rho_2)^2 E(x_{t-1}^2 | x_t) / \sigma_\eta^2$ . Now, since  $x_t$  and  $x_{t-1}$  are jointly normal random variables with zero means, it follows that  $E(x_{t-1}^2 | x_t) = \text{var}(x_{t-1} | x_t)$  is independent of  $x_t$  and is equal to

---

<sup>2</sup>This special case suffices for the points we raise in this section. For the Monte Carlo experiments reported in section 4, where we consider the general case with two endogenous regressors, the concentration parameter is computed by simulation.

$\sigma_v^2 / (1 - \rho_2^2)$ .<sup>3</sup> Hence, the concentration parameter  $\mu^2$  is:

$$\mu^2 = \frac{T \rho_2^2 \rho_{\eta v}^2}{1 - \rho_2^2} = \frac{T \lambda^2 \rho_2^2 \sigma_v^2}{(1 - \rho_2^2) (1 - \gamma_f (\rho_1 + \gamma_f \rho_2))^2 \sigma_\eta^2}, \quad (8)$$

since  $\lambda = \rho_{\eta v} \sigma_\eta / \sigma_v (1 - \gamma_f (\rho_1 + \gamma_f \rho_2))$ .

It is interesting to note that, for a fixed value of  $\lambda$ , the concentration parameter varies non-monotonically with  $\gamma_f$  and this has implications for the power of a test of  $H_0 : \gamma_f = \gamma_{f,0}$  against an alternative  $H_1 : \gamma_f = \gamma_{f,1}$ . To illustrate this, Figure 1 plots  $\mu^2$  as a function of  $\gamma_f$  when  $\lambda = 0.5$ , for different values of  $\rho$ , the first-order autocorrelation coefficient of  $x_t$ , and  $\rho_2$ . When  $\gamma_{f,0} = 0.5$ , Figure 1 shows that  $\mu^2$  is increasing in  $\gamma_f$  when  $\rho > 0$ , and decreasing in  $\gamma_f$  when  $\rho < 0$ . Thus, when  $\rho > 0$ , tests of  $H_0 : \gamma_{f,0} = 0.5$  have more power against alternatives  $H_1 : \gamma_f = \gamma_{f,1}$  for which  $\gamma_{f,1} > 0.5$  than against alternatives  $H_1 : \gamma_f = \gamma_{f,1}$  for which  $\gamma_{f,1} < 0.5$  and vice versa when  $\rho < 0$ . Depending on the value of  $\rho$  and  $\rho_2$ , the variation in the quality of identification under the alternative can be very large. For example when  $\rho = 0.9$  and  $\rho_2 = -0.8$ , the concentration parameter goes from rather low (about 5) for  $\gamma_f = 0$ , to very high (150) for  $\gamma_f = 1$ . This fact has implications for the study of the power function of tests on  $\gamma_f$ . We explore this issue further in section 4 below.

The above discussion focused on a special case of the model in which the concentration parameter is analytically tractable. In more complicated situations, the concentration parameter can be computed numerically. Mavroeidis (2005) computed it for the model given by equations (2) and (4) for different values of the parameters  $\lambda$ ,  $\rho_2$  and the relative size of the shocks  $\sigma_v / \sigma_u$  and found that it is very small (in the order of  $10^{-4}$ ) for typical values of the parameters reported in the literature. The concentration parameter remains small except for extreme departures of the parameters  $\rho_2$ ,  $\lambda$  and  $\sigma_v / \sigma_u$  from their estimated values. This shows that we should avoid doing inference on the parameters of the NKPC model using procedures that are not robust to weak instruments, since they give unreliable results.

### 3 Weak instrument robust tests for the NKPC

We employ weak instrument robust tests to conduct inference on the parameters of the NKPC model. These tests are defined for GMM so we start their discussion with a brief outline of Hansen's (1982) GMM.

GMM provides a framework for inference on a  $p$ -dimensional vector of parameters  $\theta$  for which the  $k$  dimensional moment equation

$$E(f_t(\theta)) = 0, \quad t = 1, \dots, T, \quad (9)$$

holds. For the NKPC, the moment vector reads

$$f_t(\theta) = Z_t (\pi_t - \lambda x_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1}) \quad (10)$$

as defined in equations (1)-(2). We assume that the moment equation in (9) is uniquely satisfied at  $\theta_0$ . The weak instrument robust statistics are based on the objective function for the continuous updating

---

<sup>3</sup>Since  $E(x_t) = 0$ ,  $\text{var}(x_{t-1}|x_t) = E(x_{t-1}^2) - E(x_t x_{t-1})^2 / E(x_t^2) = \sigma_v^2 (1 - \rho^2) / ((1 - \rho_2^2) (1 - \rho^2)) = \sigma_v^2 / (1 - \rho_2^2)$ , where  $\rho$  is the first autocorrelation of  $x_t$ .

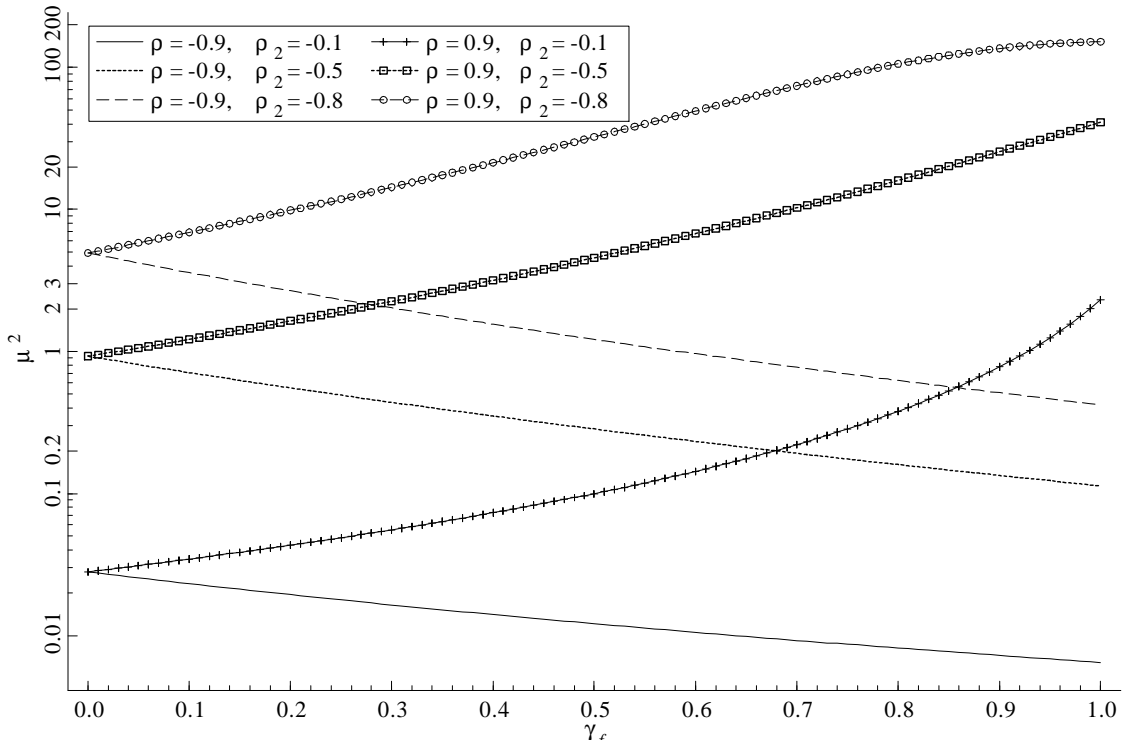


Figure 1: Concentration parameter  $\mu^2$  as a function of  $\gamma_f$  for different values of  $\rho$  and  $\rho_2$ , holding  $\lambda$  fixed at 0.5 and  $\sigma_\eta = 3$ .

estimator (CUE) of Hansen *et. al.* (1996):

$$Q(\theta) = T f_T(\theta)' \hat{V}_{ff}(\theta)^{-1} f_T(\theta), \quad (11)$$

with  $f_T(\theta) = \frac{1}{T} \sum_{t=1}^T f_t(\theta)$ . The  $k \times k$  dimensional covariance matrix estimator  $\hat{V}_{ff}(\theta)$  that we use in (11) is a consistent estimator of the covariance matrix  $V_{ff}(\theta)$  of the moment vector. Besides the moment vector  $f_t(\theta)$ , we use its derivative with respect to  $\theta$  which for the NKPC model reads

$$q_t(\theta) = \text{vec} \left( \frac{\partial f_t(\theta)}{\partial \theta'} \right) = - \left( \begin{pmatrix} x_t \\ \pi_{t+1} \\ \pi_{t-1} \end{pmatrix} \otimes Z_t \right), \quad (12)$$

and  $q_T(\theta) = \frac{1}{T} \sum_{t=1}^T q_t(\theta)$ . To obtain the limiting distributions of the weak instrument robust statistics, we assume that the sample average of the moment vector and its derivative follow a normal random process, see Kleibergen and Mavroeidis (2008).

**Assumption 1.** *The large sample behavior of  $\bar{f}_t(\theta) = f_t(\theta) - E(f_t(\theta))$  and  $\bar{q}_t(\theta) = q_t(\theta) - E(q_t(\theta))$  satisfies*

$$\psi_T(\theta) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{pmatrix} \bar{f}_t(\theta) \\ \bar{q}_t(\theta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_f(\theta) \\ \psi_\theta(\theta) \end{pmatrix} \quad (13)$$

where  $\psi(\theta) = \begin{pmatrix} \psi_f(\theta) \\ \psi_\theta(\theta) \end{pmatrix}$  is a  $k(p+1)$  dimensional Normal distributed random process with mean zero and positive semi-definite  $k(p+1) \times k(p+1)$  dimensional covariance matrix

$$V(\theta) = \begin{pmatrix} V_{ff}(\theta) & V_{f\theta}(\theta) \\ V_{\theta f}(\theta) & V_{\theta\theta}(\theta) \end{pmatrix} \quad (14)$$

with  $V_{\theta f}(\theta) = V_{f\theta}(\theta)' = (V_{\theta f,1}(\theta)' \dots V_{\theta f,p}(\theta)')'$ ,  $V_{\theta\theta}(\theta) = V_{\theta\theta,ij}(\theta)$ ,  $i, j = 1, \dots, p$  and  $V_{ff}(\theta)$ ,  $V_{\theta f,i}(\theta)$ ,  $V_{\theta\theta,ij}(\theta)$  are  $k_f \times k_f$  dimensional matrices for  $i, j = 1, \dots, p$ , and

$$V(\theta) = \lim_{T \rightarrow \infty} \text{var} \left[ \sqrt{T} \begin{pmatrix} f_T(\theta) \\ q_T(\theta) \end{pmatrix} \right]. \quad (15)$$

From the specification of  $q_t(\theta)$  in (12), it seems that  $q_t(\theta)$  does not depend on  $\theta$  in which case, since  $f_t(\theta)$  is linear in  $\theta$ , Assumption 1 is basically just a central limit theorem for  $f_t(\theta_0)$  and  $q_t(\theta_0)$  which holds under mild conditions like the  $r$ -th moments of the absolute values of  $\bar{f}_t(\theta_0)$  and  $\bar{q}_t(\theta_0)$  are finite for some  $r > 2$ , the  $\varphi$  or  $\alpha$ -mixing coefficients of  $\bar{f}_t(\theta_0)$  and  $\bar{q}_t(\theta_0)$ , are of size  $s/(s-1)$  with  $s > 1$ , see *e.g.* White (1984, Th. 5.19).

Assumption 1 differs from the traditional assumptions that are made to obtain the limiting distributions of estimators and test statistics in GMM, see *e.g.* Hansen (1982) and Newey and McFadden (1994). These assumptions consist of a normal random process assumption for the limiting behavior of  $\frac{1}{\sqrt{T}} \sum_{t=1}^T \bar{f}_t(\theta)$  and a full rank assumption for the expected value of the Jacobian,  $J(\theta) = E(\lim_{T \rightarrow \infty} \frac{\partial}{\partial \theta'} f_T(\theta))$ . Since Assumption 1 also makes a normal random process assumption for the limiting behavior of  $\frac{1}{\sqrt{T}} \sum_{t=1}^T \bar{f}_t(\theta)$ , the difference between the traditional assumption and Assumption 1 is that the full rank assumption for  $J(\theta)$  is replaced by a normal random process as-



sumption for the limiting behavior of  $\frac{1}{\sqrt{T}} \sum_{t=1}^T \bar{q}_t(\theta)$ . Since  $\bar{q}_t(\theta)$  is a mean zero random variable, the normal random process assumption for its scaled sum is, as explained above, a mild assumption. The full rank assumption of  $J(\theta)$  is, however, an assumption on a nuisance parameter and therefore difficult to verify. The limiting distributions of statistics that result under Assumption 1 therefore hold more generally than the limiting distributions of statistics that result under the traditional assumptions.

To estimate the covariance matrix, we use the covariance matrix estimator  $\hat{V}(\theta)$  which consists of  $\hat{V}_{ff}(\theta) : k_f \times k_f$ ,  $\hat{V}_{\theta f}(\theta) : k_\theta \times k_f$  and  $\hat{V}_{\theta\theta}(\theta) : k_\theta \times k_\theta$ ,  $k_\theta = pk$ . We assume that the covariance matrix estimator is a consistent one for every value of  $\theta$ . Because of the specific functional form of the derivative (12) for the NKPC, the assumption on the convergence of the derivative of  $\hat{V}_{ff}(\theta)$  that is made in Kleibergen (2005) is automatically satisfied.

The derivative estimator  $q_T(\theta)$  is correlated with the average moment vector  $f_T(\theta)$  since  $V_{\theta f}(\theta) \neq 0$ . The weak instrument robust statistics therefore use an alternative estimator of the derivative of the unconditional expectation of the Jacobian,

$$\hat{D}_T(\theta_0) = [q_{1,T}(\theta_0) - \hat{V}_{\theta f,1}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0) \dots q_{p,T}(\theta_0) - \hat{V}_{\theta f,p}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0)], \quad (16)$$

where  $\hat{V}_{\theta f,i}(\theta)$  are  $k_f \times k_f$  dimensional estimators of the covariance matrices  $V_{\theta f,i}(\theta)$ ,  $i = 1, \dots, p$ ,  $\hat{V}_{\theta f}(\theta) = (\hat{V}_{\theta f,1}(\theta)' \dots \hat{V}_{\theta f,p}(\theta)')'$  and  $q_T(\theta_0) = (q'_{1,T}(\theta_0) \dots q'_{p,T}(\theta_0))'$ . Since  $\hat{V}_{\theta f,j}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0)$  is the projection of  $q_{j,T}(\theta_0)$  onto  $f_T(\theta_0)$ , for  $j = 1, \dots, p$ , it holds that  $\hat{D}_T(\theta_0)$  is asymptotically uncorrelated with  $f_T(\theta_0)$ . Thus, when Assumption 1 and  $H_0 : \theta = \theta_0$  hold,  $\hat{D}_T(\theta_0)$  is an estimator of the expected value of the Jacobian which is in large samples independent of the average moment vector  $f_T(\theta_0)$ .

The expression of the CUE objective function in (11) is such that both the average moment vector  $f_T(\theta)$  and the covariance matrix estimator  $\hat{V}_{ff}(\theta)$  are functions of  $\theta$ . In the construction of the derivative of the CUE objective function with respect to  $\theta$ , the derivative of  $\hat{V}_{ff}(\theta)$  is typically ignored because it is of a lower order in the sample size when the Jacobian has a full rank value, i.e., when the instruments are strong. However, when the instruments are weak, the contribution of the derivative of  $\hat{V}_{ff}(\theta)$  with respect to  $\theta$  to the derivative of the CUE objective function is no longer negligible, and hence, the expression that ignores it is incorrect. When we incorporate the derivatives of both  $f_T(\theta)$  and  $\hat{V}_{ff}(\theta)$  with respect to  $\theta$ , the derivative of the CUE objective function with respect to  $\theta$  reads, see Kleibergen (2005),

$$\frac{1}{2} \frac{\partial Q(\theta)}{\partial \theta'} = T f_T(\theta)' \hat{V}_{ff}(\theta)^{-1} \hat{D}_T(\theta). \quad (17)$$

The independence of  $f_T(\theta)$  and  $\hat{D}_T(\theta)$  in large samples and the property of the derivative of the CUE objective function that it equals the (weighted) product of  $f_T(\theta)$  and  $\hat{D}_T(\theta)$  in (17) imply that we can construct statistics based on the CUE objective function whose limiting distributions are robust to weak instruments. We provide the definition of four of these statistics. The first is the S-statistic of Stock and Wright (2000) which equals the CUE objective function (11). The second statistic is a score or Lagrange Multiplier statistics that is equal to a quadratic form of (17) with respect to  $(\hat{D}_T(\theta)' \hat{V}_{ff}(\theta)^{-1} \hat{D}_T(\theta))^{-1}$  which can be considered as the inverse of the (conditional) information matrix, see Kleibergen (2007). We refer to this statistic as the KLM statistic. The third statistic is

an over-identification statistic that equals the difference between the S and KLM statistics. We refer to this statistic as the JKLM statistic. The fourth and last statistic is an extension of the conditional likelihood ratio statistic of Moreira (2003) towards GMM, see Kleibergen (2005). We refer to this statistic as the MQLR statistic.

**Definition 1.** Let  $\theta = (\alpha' : \beta')'$ , with  $\alpha$  and  $\beta$  being  $p_\alpha$  and  $p_\beta$  dimensional vectors, respectively, such that  $p_\alpha + p_\beta = p$ . To simplify the notation, we denote  $Q(\theta)$  evaluated at  $\theta = (\alpha' : \beta')'$  by  $Q(\alpha, \beta)$  and use the same notation for all other functions of  $\theta$ . Four statistics that test the hypothesis  $H_0 : \beta = \beta_0$  are:

1. The subset S-statistic of Stock and Wright (2000):

$$S(\beta_0) = Q(\tilde{\alpha}(\beta_0), \beta_0), \quad (18)$$

where  $\tilde{\alpha}(\beta_0)$  is the CUE of  $\alpha$  given that  $\beta = \beta_0$ .

2. The subset KLM statistic of Kleibergen (2005):

$$KLM(\beta_0) = f_T(\tilde{\alpha}(\beta_0), \beta_0)' \hat{V}_{ff}(\tilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} P_{\hat{V}_{ff}(\tilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} \hat{D}_T(\tilde{\alpha}(\beta_0), \beta_0)} \hat{V}_{ff}(\tilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} f_T(\tilde{\alpha}(\beta_0), \beta_0). \quad (19)$$

3. The subset JKLM overidentification statistic:

$$JKLM(\beta_0) = S(\beta_0) - KLM(\beta_0). \quad (20)$$

4. The subset extension of the conditional likelihood ratio statistic of Moreira (2003) applied in a GMM setting:

$$MQLR(\beta_0) = \frac{1}{2} [KLM(\beta_0) + JKLM(\beta_0) - rk(\beta_0) + \sqrt{\{KLM(\beta_0) + JKLM(\beta_0) + rk(\beta_0)\}^2 - 4JKLM(\beta_0)rk(\beta_0)}], \quad (21)$$

where  $rk(\beta_0)$  is a statistic that tests for a lower rank value of  $J(\tilde{\alpha}(\beta_0), \beta_0)$  and is a function of  $\hat{D}_T(\tilde{\alpha}(\beta_0), \beta_0)$  and  $\hat{V}_{\theta\theta.f}(\tilde{\alpha}(\beta_0), \beta_0) = \hat{V}_{\theta\theta}(\tilde{\alpha}(\beta_0), \beta_0) - \hat{V}_{\theta f}(\tilde{\alpha}(\beta_0), \beta_0) \hat{V}_{ff}(\tilde{\alpha}(\beta_0), \beta_0)^{-1} \hat{V}_{f\theta}(\tilde{\alpha}(\beta_0), \beta_0)$ :

$$rk(\beta_0) = \min_{\varphi \in \mathbb{R}^{p-1}} T\left(\begin{pmatrix} 1 \\ \varphi \end{pmatrix}\right)' \hat{D}_T(\tilde{\alpha}(\beta_0), \beta_0)' \left[ \left( \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \otimes I_k \right)' \hat{V}_{\theta\theta.f}(\tilde{\alpha}(\beta_0), \beta_0) \left( \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \otimes I_k \right) \right]^{-1} \hat{D}_T(\tilde{\alpha}(\beta_0), \beta_0) \begin{pmatrix} 1 \\ \varphi \end{pmatrix}. \quad (22)$$

The specifications of the weak instrument robust statistics in Definition 1 apply both to tests of hypotheses on subsets of the parameters as well as on the full vector of parameters, in which case  $\beta$  coincides with  $\theta$  and  $\alpha$  drops out of the expressions.

It is useful to view the  $rk(\beta_0)$  statistic in the expression of the MQLR statistic, see equation (22), as the result of minimizing another CUE objective function. Specifically, define the  $k \times (p + 1)$  matrix

$F_T$  as

$$F_T = \frac{1}{T} \sum_{t=1}^T F_t, \quad \text{where } F_t = Z_t \begin{pmatrix} \pi_t \\ x_t \\ \pi_{t+1} \\ \pi_{t-1} \end{pmatrix}', \quad (23)$$

and let  $\hat{W}$  denote a consistent estimator of the asymptotic variance of  $\sqrt{T}\text{vec}(F_T)$ . Then,  $f_t(\theta)$  in (10) can be alternatively specified as  $f_t(\theta) = F_t \begin{pmatrix} 1 \\ -\theta \end{pmatrix}$ , and  $\hat{V}_{ff}(\theta)$  as  $\left[ \begin{pmatrix} 1 \\ -\theta \end{pmatrix} \otimes I_k \right]' \hat{W} \left[ \begin{pmatrix} 1 \\ -\theta \end{pmatrix} \otimes I_k \right]$ . Hence, the CUE objective function in (11) can be specified in the same form as the minimand on the right-hand side of the  $rk(\beta_0)$  statistic in (22), and so  $rk(\beta_0)$  results from minimizing another CUE GMM objective function. Conversely, the CUE objective function (11) corresponds to a statistic that tests that the rank of the  $k \times (p+1)$  matrix  $E(\lim_{T \rightarrow \infty} F_T)$  is equal to  $p$ .<sup>4</sup>

The specification of  $rk(\beta_0)$  in (22) is an alternative specification of the Cragg and Donald (1997) rank statistic that tests if the rank of the Jacobian evaluated at  $(\tilde{\alpha}(\beta_0), \beta_0)$  is equal to  $p-1$ .<sup>5</sup> The specification in (22) is more attractive than the one originally proposed by Cragg and Donald (1997) since it requires optimization over a  $(p-1)$ -dimensional space while the statistic proposed by Cragg and Donald (1997) requires optimization over a  $(k+1)(p-1)$ -dimensional space. Other rank statistics, like those proposed by Lewbel (1991), Robin and Smith (2000) or Kleibergen and Paap (2006), can be used as well for  $rk(\beta_0)$  but the specification in (22) is convenient since it guarantees that  $\text{rk}(\tilde{\beta}) > S(\tilde{\beta})$  which is necessary for  $\text{MQLR}(\beta)$  to be equal to zero at the CUE, see Kleibergen (2007).<sup>6</sup> When  $p=1$ , the definition of  $rk(\beta_0)$  is unambiguous and reads

$$rk(\beta_0) = T \hat{D}_T(\beta_0)' \hat{V}_{\theta\theta, f}(\beta_0)^{-1} \hat{D}_T(\beta_0). \quad (24)$$

The robustness of the statistics stated in Definition 1 to weak instruments was until recently only guaranteed for tests on the full parameter vector, see *e.g.* Stock and Wright (2000), Kleibergen (2002, 2005) and Moreira (2003). Recent results by Kleibergen (2008) and Kleibergen and Mavroeidis (2008) show that the robustness to weak instruments extends to tests on a subset of the parameters when the unrestricted parameters are estimated using the CUE under the hypothesis of interest.

To get the intuition behind the above result, which is formally stated in Theorem 1 below, it is useful to interpret the CUE objective function as a measure of the distance of a matrix from a reduced-rank value. We just showed that the CUE objective function corresponds to a statistic that tests that the rank of the matrix  $E(\lim_{T \rightarrow \infty} F_T)$ , see equation (23), is reduced by one, i.e., it is equal to  $p$ . Indeed, when the moment conditions hold, the rank of the above matrix is at most  $p$ . Now, when the parameter  $\alpha$  that is partialled out is well identified, the first  $p_\alpha$  columns of the Jacobian of the moment conditions  $J(\theta)$  constitute a full rank matrix. The (conditional) limiting distributions of the subset statistics for the parameter  $\beta$  are then equal to those of the corresponding statistics that test a hypothesis on the full parameter vector after appropriately correcting the degrees of freedom parameter of the (conditional) limiting distributions. In contrast, when  $\alpha$  is not well-identified, the

<sup>4</sup> A more detailed explanation of this point is given in the Appendix.

<sup>5</sup> We evaluate the Jacobian at  $(\tilde{\alpha}(\beta_0), \beta_0)$  since we use an estimate of  $\alpha_0$ ,  $\tilde{\alpha}(\beta_0)$ , in the expressions of the different statistics in Definition 1. Of course, when the moments conditions are linear, as it is the case for the NKPC model,  $J(\theta)$  is independent of  $\theta$ , so  $\tilde{\alpha}(\beta_0)$  does not appear in the expression of the Jacobian.

<sup>6</sup> Both the equality of (22) to the Cragg and Donald (1997) rank statistic and that  $S(\tilde{\beta}) < rk(\tilde{\beta})$  are proven in the Appendix.

first  $p_\alpha$  columns of the Jacobian are relatively close to a reduced rank value. But since  $J(\theta)$  is a submatrix of the matrix  $F_T$  that appears in the CUE objective function, weak identification of  $\alpha$  implies that  $F_T$  will be close to a matrix whose rank is reduced even further, relative to the case when  $\alpha$  is well-identified. This explains why the limiting distribution of the CUE objective function for weakly identified values of  $\alpha$  is bounded from above by the limiting distribution that results for well identified values of  $\alpha$ . The CUE objective function corresponds to the  $S$  statistic, and since the subset statistics in Definition 1 are all based on the  $S$  statistic, the above argument extends to the other subset statistics as well, see Kleibergen and Mavroeidis (2008). The (conditional) limiting distributions of the subset statistics under well identified values of  $\alpha$  therefore provide upper bounds on the (conditional) limiting distributions in general.

**Theorem 1.** *Under Assumption 1 and  $H_0 : \beta = \beta_0$ , the (conditional) limiting distributions of the subset  $S$ ,  $KLM$ ,  $JKLM$  and  $MQLR$  statistics given in Definition 1 are such that*

$$\begin{aligned} S(\beta_0) &\underset{a}{\preceq} \varphi_{p_\beta} + \varphi_{k-p} \\ KLM(\beta_0) &\underset{a}{\preceq} \varphi_{p_\beta} \\ JKLM(\beta_0) &\underset{a}{\preceq} \varphi_{k-p} \\ MQLR(\beta_0)|rk(\beta_0) &\underset{a}{\preceq} \frac{1}{2} \left[ \varphi_{p_\beta} + \varphi_{k-p} - rk(\beta_0) + \sqrt{\{\varphi_{p_\beta} + \varphi_{k-p} + rk(\beta_0)\}^2 - 4\varphi_{k-p}rk(\beta_0)} \right] \end{aligned} \quad (25)$$

where “ $\underset{a}{\preceq}$ ” indicates that the limiting distribution of the statistic on the left hand side of the  $\underset{a}{\preceq}$  sign is bounded by the random variable on the right hand side and  $\varphi_{p_\beta}$  and  $\varphi_{k-p}$  are independent  $\chi^2$  distributed random variables with  $p_\beta$  and  $k-p$  degrees of freedom, respectively.

**Proof.** see Kleibergen and Mavroeidis (2008). ■

The bounding distribution of  $MQLR(\beta_0)$  in Theorem 1 is conditional on the value of  $rk(\beta_0)$ . Definition 1 shows that  $rk(\beta_0)$  is a rank statistic that tests the rank of the Jacobian evaluated at  $(\tilde{\alpha}(\beta_0), \beta_0)$ . The rank of the Jacobian is a measure of identification. The bounding distribution of  $MQLR(\beta_0)$  therefore depends on the degree of identification. When  $\alpha$  and/or  $\beta$  are not well identified,  $rk(\beta_0)$  is small and the bounding distribution of  $MQLR(\beta_0)$  is similar to that of  $S(\beta_0)$ . When  $\alpha$  and  $\beta$  are well identified,  $rk(\beta_0)$  is large and the bounding distribution of  $MQLR(\beta_0)$  is similar to that of  $KLM(\beta_0)$ . Since  $rk(\beta_0)$  is a function of  $\hat{D}_T(\tilde{\alpha}(\beta_0), \beta_0)$ , it is in large samples independent of  $S(\beta_0)$  and  $KLM(\beta_0)$  so  $rk(\beta_0)$  does not influence the bounding distributions of  $S(\beta_0)$  and  $KLM(\beta_0)$ .

Theorem 1 is important for practical purposes because it implies that usage of critical values that result from the bounding distributions leads to tests that have the correct size in large samples. This holds because these statistics have the random variables on the right hand side of the bounding sign as their limiting distributions when  $\alpha$  is well identified, see Stock and Wright (2000) and Kleibergen (2005). Hence, there are values of the nuisance parameters for which the limiting distributions of these statistics coincide with the bounding distributions so the maximum rejection frequency of the tests over all possible values of the nuisance parameters coincides with the size of the test. In the cases when  $\alpha$  is weakly identified, these critical values lead to conservative tests since the rejection frequency is less than the significance level of the test.

In addition to cases when  $\alpha$  is well-identified, the size of the weak instrument robust statistics is

also equal to the significance level of the test when  $\beta$  coincides with the full parameter vector.

Confidence sets for the parameter  $\beta$  with confidence level  $(1 - \mu) \times 100\%$  can be obtained by inverting any of the identification robust statistics, see *e.g.* Zivot *et al* (1998). For example, a 95%-level MQLR-based confidence set is obtained by collecting all the values of  $\beta_0$  for which the MQLR test of the hypothesis  $H_0 : \beta = \beta_0$  does not reject  $H_0$  at the 5% level of significance.

Theorem 1 suggests a testing procedure that controls the size of tests on subsets of the parameters. Other testing procedures that control the size of subset tests are the projection-based approach, see *e.g.* Dufour (1997), Dufour and Jasiak (2001) and Dufour and Taamouti (2005,2007), and an extension of the Robins (2004) test, see Chaudhuri (2007) and Chaudhuri *et. al.* (2007).

**Projection-based testing** Projection-based tests do not reject  $H_0 : \beta = \beta_0$  when tests of the joint hypothesis  $H^* : \beta = \beta_0, \alpha = \alpha_0$  do not reject  $H^*$  for some values of  $\alpha_0$ . When the limiting distribution of the statistic used to conduct the joint test does not depend on nuisance parameters, the maximal value of the rejection probability of the projection-based test over all possible values of the nuisance parameters cannot exceed the size of the joint test. Hence, the projection-based tests control the size of the tests of  $H_0$ .

Since the weak instrument robust statistics in Definition 1 that test  $H_0 : \beta = \beta_0$  coincide with their expressions that test  $H^* : \beta = \beta_0, \alpha = \tilde{\alpha}(\beta_0)$ , the projection based statistics do not reject  $H_0$  when the subset weak instrument robust statistics do not reject it either. This holds since the critical values that are applied by the projection based approach are strictly larger than the critical values that result from Theorem 1 since the projection based tests use the critical values of the joint test. Hence, whenever a subset weak instrument robust statistic does not reject, its projection based counterpart does not reject either. Thus projection based tests are conservative and their power is strictly less than the power of the subset weak instrument robust statistics, see Kleibergen (2008) and Kleibergen and Mavroeidis (2008).

**Robins (2004) test** Chaudhuri (2007) and Chaudhuri *et. al.* (2007) propose a refinement of the projection based approach by using the Robins (2004) test. This approach decomposes the joint statistic used by the projection based test into two statistics, one which tests  $\alpha$  given  $\beta_0$  and one which tests  $\beta_0$  given  $\alpha$ . The first of these statistics is used to construct a  $(1 - \nu) \times 100\%$  level confidence set for  $\alpha$  given that  $\beta = \beta_0$  while the second one is used to test the hypothesis  $H^* : \beta = \beta_0, \alpha = \alpha_0$  with  $\mu \times 100\%$  significance for every value of  $\alpha_0$  that is in the confidence set of  $\alpha$  that results from the first statistic. The hypothesis of interest  $H_0$  is rejected whenever the confidence set for  $\alpha$  is empty or when the second test rejects for all values of  $\alpha$  that are in its confidence set. Chaudhuri (2007) and Chaudhuri *et. al.* (2007) show that when the limiting distributions of both statistics do not depend on nuisance parameters the size of the overall testing procedure cannot exceed  $\nu + \mu$ .

If we apply the above decomposition to the KLM and S statistics, the first-stage statistic for both could consist of a KLM statistic that tests  $H_\alpha : \alpha = \alpha_0, \beta = \beta_0$  which is used to construct a  $(1 - \nu) \times 100\%$  confidence set for  $\alpha$  given that  $\beta = \beta_0$ . The second-stage statistics would then be respectively a KLM and S statistic that test  $H_0 : \beta = \beta_0$  with  $\mu \times 100\%$  significance level for every value of  $\alpha$  that lies in its  $(1 - \nu) \times 100\%$  confidence set. Since the value of the first stage KLM statistic is equal to zero at the CUE  $\tilde{\alpha}(\beta_0)$ , the confidence set of  $\alpha$  will not be empty. Whenever the subset KLM and S statistic do not reject  $H_0 : \beta = \beta_0$ , neither will the Robins (2004) based procedure since

the statistics that it uses in the second stage coincide with the subset KLM and S statistics when evaluated at  $\alpha_0 = \tilde{\alpha}(\beta_0)$  which lies in the  $(1 - \nu) \times 100\%$  confidence set of  $\alpha$ . Thus the Robins (2004) based testing procedure is conservative and its power is strictly less than the power of the subset weak instrument robust statistics. The same argument can be applied, in a somewhat more involved manner, to the MQLR statistic as well, but we omit it for reasons of brevity.

**Tests at distant values of the parameters** At distant values of the parameters of interest, the identification robust statistics from Definition 1 correspond with statistics that test the identification of the parameters, see Kleibergen (2007, 2008) and Kleibergen and Mavroeidis (2008). For example, Kleibergen and Mavroeidis (2008) show that for a scalar  $\beta_0$ , the behavior of the S-statistic at large values of  $\beta_0$  is characterized by:

$$\lim_{\beta_0 \rightarrow \infty} S(\beta_0) = \min_{\varphi \in \mathbb{R}^{p-1}} T q'_T \begin{bmatrix} \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \otimes I_k \end{bmatrix} \begin{bmatrix} \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \otimes I_k \end{bmatrix}' \hat{V}_{\theta\theta} \begin{bmatrix} \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \otimes I_k \end{bmatrix} \end{bmatrix}^{-1} \begin{bmatrix} \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \otimes I_k \end{bmatrix}' q_T. \quad (26)$$

where we have used the fact that  $q_t(\theta)$  in (12) and hence also  $\hat{V}_{\theta\theta}$  do not depend on the parameters because the moment conditions are linear. Thus, the specification of (26) is identical for all parameters in the NKPC, and this implies that if  $S(\beta_0)$  is not significant at a distant value of  $\beta_0$  for one specific parameter, it is therefore not significant for any other parameter either. Hence, when a confidence set for a specific parameter is unbounded, it is unbounded for the other parameters as well.<sup>7</sup> Thus the weak identification of one parameter carries over to all the other parameters.

Equation (26) shows that at distant values,  $S(\beta_0)$  has the same functional form as  $rk(\beta_0)$  in (22) which corresponds to the Cragg and Donald (1997) rank statistic. At distant values of  $\beta_0$ ,  $S(\beta_0)$  therefore corresponds to a test of the rank of the Jacobian  $J(\theta)$  using  $q_T$ . The rank of the Jacobian governs the identification of the parameters so, at distant values of  $\beta_0$ ,  $S(\beta_0)$  corresponds to a test of the rank condition for identification of the parameters.

Similar expressions for the other identification robust statistics in Definition 1 result at distant scalar values of  $\beta_0$ . It is, for example, shown in Kleibergen (2008) that when Assumption 1 holds, at distant values of  $\beta_0$  the conditioning statistic  $rk(\beta_0)$  has a  $\chi^2(k - p_\alpha)$  limiting distribution. This implies that  $rk(\beta_0)$  has a relatively small value at distant values of  $\beta_0$  so  $MQLR(\beta_0)$  corresponds to  $S(\beta_0)$  at distant values of  $\beta_0$ .

For tests on the full parameter vector, the statistics in Definition 1 also correspond with statistics that test the rank condition for identification. In the linear instrumental variables regression model with one included endogenous variable, Kleibergen (2007) shows that the S-statistic, which then corresponds to the Anderson-Rubin (AR) statistic, see Anderson and Rubin (1949), is identical to the first-stage F statistic at distant values of the parameter of interest. This indicates that the AR or S-statistic is, in case of weak instruments, more powerful than the pretest-based two stage least squares t statistic which is commonly used in practice. In Stock and Yogo (2005), it is shown that the two stage least squares t statistic can be used in a reliable manner when the first-stage F statistic exceeds ten. When it is less than ten, this two-step approach implicitly yields an unbounded confidence set for the parameters. However, a value of ten for the first-stage F statistic is highly significant and

<sup>7</sup>Dufour (1997) shows that statistics that test hypotheses on locally non-identified parameters and whose limiting distributions do not depend on nuisance parameters have a non-zero probability for a unbounded confidence set.

immediately implies that the parameters are not completely unidentified, since it already excludes large values of the parameters from confidence sets that are based on the AR or S statistic. Hence, the AR statistic has power at values of the first-stage F statistic for which the two stage least squares t statistic cannot be used in a reliable manner. Since the MQLR statistic is more powerful than the S statistic, this conclusion extends to the use of the MQLR statistic as well. Thus, the pre-test based two stage least squares t statistic is inferior to the MQLR statistic both because it cannot control size accurately, and because it is less powerful when instruments are weak.

The above shows that the presence of a weakly identified parameter leads to unbounded confidence sets for all other parameters even for those that are well identified, *i.e.* those parameters whose corresponding columns in the Jacobian of the moment conditions differ from zero. A weakly identified parameter therefore contaminates the analysis of all other parameters. It might therefore be worthwhile to remove such a parameter from the model. The resulting model will obviously be misspecified. We consider this trade-off between misspecification and the desire to have bounded confidence sets an important topic for further research.

**Stability tests** Besides testing for a fixed value of the parameters, the weak instrument robust statistics can be used to test for changes in the parameters over time, as well, *i.e.*, to obtain identification-robust tests for structural change. Caner (2007) derives such tests based on the S and KLM statistics. If we define the average moment vectors for two consecutive time periods,

$$\begin{aligned} f_{\pi T}(\theta) &= \frac{1}{T} \sum_{t=1}^{[\pi T]} f_t(\theta) \\ f_{(1-\pi)T}(\theta) &= \frac{1}{T} \sum_{t=[\pi T]+1}^T f_t(\theta) \end{aligned} \quad (27)$$

with  $[\cdot]$  the Entier or integer-part function, and  $\hat{V}_{\pi, ff}(\theta_0)$  and  $\hat{V}_{1-\pi, ff}(\theta_0)$  are the covariance estimators of the moment vectors for each time period, Caner's (2007) sup-S statistic to test for structural change reads

$$S_{\text{change}} = \sup_{a \leq \pi \leq b} \min_{\theta} T \left[ \pi f_{\pi T}(\theta)' \hat{V}_{\pi, ff}(\theta)^{-1} f_{\pi T}(\theta) + (1 - \pi) f_{(1-\pi)T}(\theta)' \hat{V}_{1-\pi, ff}(\theta)^{-1} f_{(1-\pi)T}(\theta) \right], \quad (28)$$

with  $0 < a < b < 1$ . Under Assumption 1<sup>8</sup> and no structural change, Caner (2007, Theorem 1) shows that the limiting distribution of the structural change S-statistic (28) is bounded by

$$\sup_{a \leq \pi \leq b} \left[ \frac{1}{\pi} W_k(\pi)' W_k(\pi) + \frac{1}{1-\pi} (W_k(1) - W_k(\pi))' (W_k(1) - W_k(\pi)) \right], \quad (29)$$

with  $W_k(t)$  a  $k$ -dimensional standard Brownian motion defined on the unit interval. Alongside the S statistic, Caner (2007) shows that the KLM statistic can be extended to test for structural change in a similar manner.

The null hypothesis of no structural change (*i.e.*, structural stability) is a hypothesis on a subset of the parameters of the two-period model. The unrestricted parameter here is the vector  $\theta$  which is the same in the two periods under the null hypothesis  $H_0^{SC}$  of no structural change. The bounding distributions of Caner (2007) are the limiting distributions of the statistics that test the null hypothesis of no structural change  $H_0^{SC}$  jointly with the hypothesis  $H_0^\theta : \theta = \theta_0$ . Because the structural change

---

<sup>8</sup>Strictly speaking, we need to replace Assumption 1 by a functional central limit theorem for the partial sums of the moment conditions and their Jacobian.

statistic (28) is evaluated at the CUE of  $\theta$ , it does not reject  $H_0^{SC}$  whenever the corresponding test of the joint null hypothesis  $H_0^{SC} \cap H_0^\theta$  does not reject for some values of  $\theta_0$ . Thus the identification-robust tests proposed by Caner (2007) are projection-based tests of no structural change, and, as Caner (2007) observes, they are conservative whenever  $\theta$  is well-identified.

The bounding distribution for the sup-S test given in expression (29) can also be written as

$$\sup_{a \leq \pi \leq b} \left[ \frac{(W_k(\pi) - \pi W_k(1))'(W_k(\pi) - \pi W_k(1))}{\pi(1-\pi)} \right] + \chi_k^2. \quad (30)$$

When  $\theta$  is well-identified, Caner (2007, Theorem 3) shows that this bound can be sharpened to

$$\sup_{a \leq \pi \leq b} \left[ \frac{(W_k(\pi) - \pi W_k(1))'(W_k(\pi) - \pi W_k(1))}{\pi(1-\pi)} \right] + \chi_{k-p}^2, \quad (31)$$

where  $p$  is the dimension of  $\theta$ . Usage of the critical values associated with the distribution in (31) results in a subset version of the sup-S test, which is clearly more powerful than the projection-based version obtained from the bounding distribution (30), since the value of the test statistic is the same in both cases and only the critical value changes. It is therefore of interest to study whether the bounding results for subset statistics from Kleibergen and Mavroeidis (2008) extend to structural stability tests that are based on identification-robust statistics. This is an important topic for further research on which we are currently working.

Because of the prevalence of the Lucas (1976) critique, it is important to test the stability of the parameters of the NKPC model. The statistics proposed by Caner (2007) are well suited for this purpose since their limiting distributions are robust to weak instruments.

## 4 Simulations

To illustrate the properties of the above statistics, we conduct a simulation experiment that studies the size and power properties of tests of  $\gamma_f$ , the coefficient of the forward-looking term in the NKPC, see equation (2). The data generating process is given by equations (7) and (4), where  $\eta_t$  and  $v_t$  are jointly normally distributed with unit variances and correlation  $\rho_{\eta v}$ . The sample size that we use is 1000 and we use 10000 simulations to construct the power curves.<sup>9</sup> We calibrate  $\rho_{\eta v}$  to the US data on inflation and the labor share over the period 1960 to the present and this gives  $\rho_{\eta v} = 0.2$ .

In Section 2, we showed that, in the special case in which the regressor  $x_t$  is exogenous, the concentration parameter  $\mu^2$  varies with  $\gamma_f$  when  $\lambda$  is kept fixed, see equation (8) and Figure 1. In the present setting, we treat  $x_t$  as an endogenous regressor, so the formula given in equation (8) does not apply, as we need to measure the strength of the instruments  $\mu^2$  by the smallest eigenvalue of the concentration matrix. By numerical computation, it can be shown that this eigenvalue also varies with  $\gamma_f$  when  $\lambda$  is fixed, in a way that is very similar to the case when  $x_t$  is exogenous, shown in Figure 1. When we construct power curves for tests on  $\gamma_f$ , the dependence of  $\mu^2$  on the value of  $\gamma_f$  makes the power curves difficult to interpret. This is because we cannot attribute a change in the rejection frequency to either the difference between the actual value of  $\gamma_f$  and the hypothesized one or a change in the quality of the identification. In the construction of the power curves for tests on  $\gamma_f$ , we therefore keep the smallest eigenvalue of the concentration parameter  $\mu^2$  constant when

---

<sup>9</sup>Since we construct the power curves for a fixed value of the concentration parameter, the large value of the sample size is only used to reduce the sampling variability.



we vary  $\gamma_f$ . We achieve this by allowing  $\lambda$  to change when we vary  $\gamma_f$  according to the equation  $\lambda = \rho_{\eta v} \sigma_{\eta} / \sigma_v (1 - \gamma_f (\rho_1 + \gamma_f \rho_2))$ .

Since the identification of the structural parameters depends on  $\rho_2$ , we use it to vary the quality of the instruments. The other reduced-form parameter  $\rho_1$  is set equal to  $(1 - \rho_2) \rho$ , so as to keep the first-order autocorrelation coefficient of  $x_t$ ,  $\rho$ , fixed at the value estimated from US data on the labor share, as in Mavroeidis (2005). The moment conditions are given by (10) with  $\gamma_b = 0$  and the instrument set is  $Z_t = (\pi_{t-1}, \pi_{t-2}, \pi_{t-3}, x_{t-1}, x_{t-2}, x_{t-3})'$ .

We compute the size and power of testing  $H_0 : \gamma_f = 1/2$  against  $H_1 : \gamma_f \neq 1/2$  at the 5% significance level using a two-step Wald statistic and the subset  $S$ , KLM, JKLM and MQLR statistics. The latter statistics all use the CUE for  $\lambda$ . Table 1 reports the rejection frequencies under the null hypothesis and Figures 2 and 3 show the resulting power curves of 5% significance level tests of the null hypothesis  $H_0 : \gamma_f = 1/2$  against  $H_1 : \gamma_f = \gamma_{f,1}$  for values of  $\gamma_{f,1}$  between zero and one. The left-hand sides correspond to weak instruments (the smallest eigenvalue of the concentration matrix divided by the number of instruments,  $\mu^2/k$ , is equal to 1), while the right-hand sides corresponds to strong instruments ( $\mu^2/k = 30$ ). The power curves reported in the Figures are for the case  $\rho_{\eta v} = 0.2$ , for which the associated null rejection frequencies are given in the left two columns of Table 1.

Table 1 also reports the rejection frequencies of the tests with identical values of  $\mu^2/k$  but with a different value of the correlation coefficient of the errors,  $\rho_{\eta v} = 0.99$ . The results show that the Wald statistic is size distorted in the case of weak instruments, while the size of all other statistics is around or below 5%. Table 1 also shows that the projection and Robins based  $S$  tests are conservative.

Figure 2 shows that under weak instruments the Wald statistic is severely size distorted while the power of all identification robust tests is similar and small. Under strong instruments, the power curves that result from the KLM and MQLR statistics are indistinguishable and the  $S$  test is less powerful. The size of the Wald test is improved relative to the weak instruments case, but it is still higher than the nominal level.

Figure 3 compares the power of the subset  $S$  test with the projection based version  $S_{pr}$  as well as the Robins version,  $S_{rob}$  the latter relying on a 2% level  $KLM$  pretest for  $\lambda$ . The subset  $S$  test dominates both of the other two versions.

In the linear instrumental variables regression model with homoscedastic errors and one included endogenous variable the MQLR statistic coincides with the CLR statistic of Moreira (2003). In that model, Andrews *et al* (2006) show that the MQLR statistic is the most powerful statistic for testing two-sided hypotheses. They obtain this result by constructing the power envelope, which results from point optimal test statistics, and showing that the power curve that results from the MQLR statistic coincides with the power envelope. Figure 2 shows that the MQLR statistic is the preferred statistic for testing hypotheses on subsets of the parameters in GMM in our simulation experiment as well. An extension of this result to the general case of testing hypotheses on subsets of the parameters in GMM has not been established. We consider it an important topic for further research.

## 5 Estimation results

We estimate the NKPC model using quarterly data for the US economy over the period 1960 quarter 1 to 2007 quarter 4. Inflation is measured by 100 times  $\Delta \ln(P_t)$ , where  $P_t$  is the GDP deflator, obtained from the FRED database of the St. Louis Fed. Following Galí and Gertler (1999), we use the labor

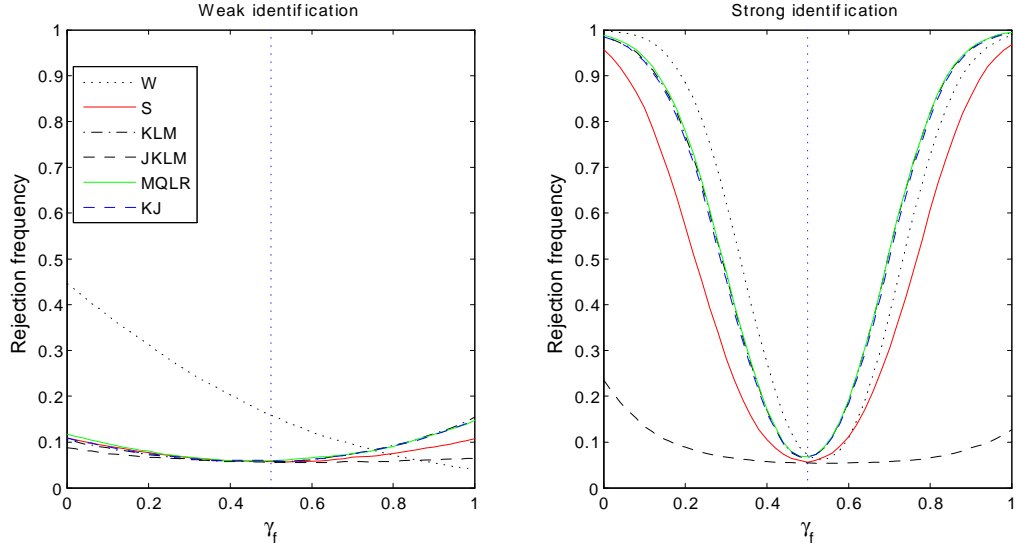


Figure 2: Power curves of 5% level tests for  $H_0 : \gamma_f = 0.5$  against  $H_1 : \gamma_f \neq 0.5$ . The sample size is 1000 and the number of MC replications is 10000.

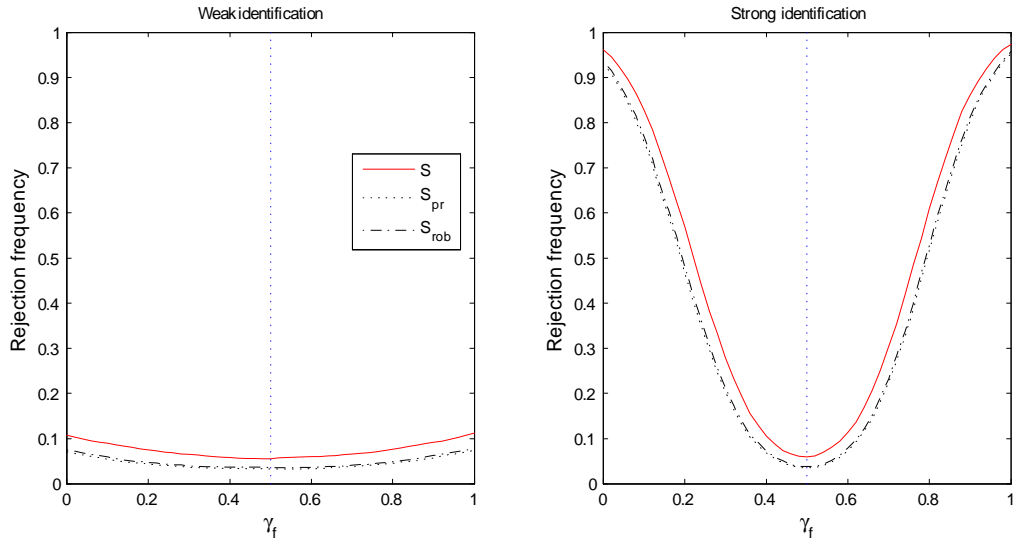


Figure 3: Comparison of the power curves of 5% level subset, projection and Robins (2004)  $S$  tests for  $H_0 : \gamma_f = 0.5$  against  $H_1 : \gamma_f \neq 0.5$ . The sample size is 1000 and the number of MC replications is 10000.

Table 1: Null rejection frequencies of 5% level tests of the hypothesis  $\gamma_f = 0.5$  against a two-sided alternative in the NKPC

	$\rho_{\eta v} = 0.2$		$\rho_{\eta v} = 0.99$	
	weak	strong	weak	strong
$W$	0.159	0.072	0.485	0.076
$S$	0.057	0.056	0.058	0.061
$KLM$	0.057	0.067	0.055	0.055
$JKLM$	0.055	0.054	0.057	0.058
$MQLR$	0.059	0.067	0.059	0.056
$KJ$	0.059	0.068	0.056	0.054
$S_{pr}$	0.033	0.034	0.034	0.034
$S_{rob}$	0.035	0.037	0.037	0.037

The model is  $E[Z_t(\pi_t - \lambda x_t - \gamma_f \pi_{t+1})] = 0$ , where  $Z_t$  includes first three lags of  $\pi_t, x_t$ . Newey and West (1987) Weight matrix. The smallest eigenvalue of the concentration matrix per instrument ( $\mu^2/k$ ) is 1 for weak and 30 for strong identification. 10000 Monte Carlo replications.

share as a proxy for marginal costs. Use of alternative measures, such as the CBO’s estimate of the output gap, detrended output or output growth, produce similar results (point estimates of  $\lambda$  are different, notably negative when output gap is used, but the confidence intervals are very similar in all cases). The data for the labor share were obtained from the Bureau of Labor Statistics.<sup>10</sup> To ensure our estimates of the slope coefficient  $\lambda$  are comparable to those reported in Galí and Gertler (1999), we scale the log of the labor share by a constant factor of 0.123, as they do. This factor depends on two unidentifiable structural parameters, and only affects the interpretation of the coefficient  $\lambda$ .

We estimate the NKPC model (2) by the CUE, using three lags of inflation and the labor share as instruments and the Newey and West (1987) HAC estimator of the optimal weight matrix. The point estimates as well as the bounds of 95% confidence sets derived by inverting the subset MQLR statistic are reported in the first column of Table 2. We also report the result of the Hansen (1982) test of over-identifying restrictions, which is correctly sized when evaluated at the CUE, as explained above. The  $p$  value for the Hansen test is about 0.5, showing no evidence against the validity of the moment restrictions at conventional significance levels.<sup>11</sup>

The point estimates we obtain are comparable to those found in many other studies that use a similar limited information approach. The slope of the Phillips curve is estimated to be positive but small, and notably insignificantly different from zero. The forward-looking coefficient  $\gamma_f$  is about 3/4, and dominates the backward-looking coefficient, which is about 1/4. However, the confidence intervals are relatively wide, and notably wider than the Wald-based confidence intervals reported in most other studies. Thus, we cannot reject at the 5% level the pure NKPC model which arises when  $\gamma_b = 0$ . This seems counter to the conventional view, e.g., Galí and Gertler (1999), that some degree of ‘intrinsic’ persistence is necessary to match the observed inflation dynamics. We shall investigate later the robustness of these results to changes in the instruments and estimation period.

<sup>10</sup>BLS, series ID: PRS85006173.

<sup>11</sup>These results are not sensitive to the choice of HAC estimator. For example, they are almost the same if we use the quadratic spectral kernel estimator proposed by Andrews and Monahan (1992) or the MA- $l$  estimator proposed by West, K.D. (1997).

Many studies of the hybrid NKPC impose the restriction that the forward and backward coefficients sum to one, i.e.,  $\gamma_f + \gamma_b = 1$ , see Buiter and Jewitt (1989), Fuhrer and Moore (1995), Christiano *et al* (2005) and Rudd and Whelan (2006). Even though formal theoretical justification for this restriction can be provided from microfoundations, e.g., Galí and Gertler (1999), Woodford (2003, Chapter 2) and Christiano *et al* (2005), the motivation for it has largely been empirical. Indeed, our estimates reported in Table 2 indicate that  $\gamma_f + \gamma_b$  is not significantly different from one, in line with most other studies. To shed further light on this, we report  $1 - p$  value plots for various identification-robust (subset) tests on the parameter  $\delta = \gamma_f + \gamma_b - 1$  in Figure 4. The null hypothesis  $\gamma_f + \gamma_b = 1$  is not rejected at conventional significance levels by any of the tests. It is also noteworthy that the parameter  $\gamma_f + \gamma_b - 1$  is very accurately estimated, since the 95% level MQLR confidence interval reported in Table 2 is very tight around zero.

It may be thought that imposing the restriction  $\gamma_f + \gamma_b = 1$  will improve the identifiability of the structural parameters  $\lambda, \gamma_f$ . This argument was made, for example, in Jondeau and Le Bihan (2008). To investigate this possibility, we re-estimate the NKPC under the restriction  $\gamma_f + \gamma_b = 1$ , and the results are reported in the second column of Table 2. It is clear that the fit of the model, as indicated by Hansen's (1982) over-identification statistic, is essentially not altered. Moreover, by comparing the confidence intervals for  $\lambda$  and  $\gamma_f$  reported in the two columns of Table 2, we notice that they shrink very little when the restriction is imposed. Intuitively, imposing restrictions can help identification mainly when the restrictions are placed in directions of the parameter space where identification is weak. The reason why the restriction  $\gamma_f + \gamma_b = 1$  does not improve the identifiability of  $\lambda$  and  $\gamma_f$  is because the parameter  $\gamma_f + \gamma_b$  is rather well-identified.

In fact, all of our subsequent results are essentially unaffected by whether we impose the restriction  $\gamma_f + \gamma_b = 1$  or not, so, for simplicity, we shall report only the results of the restricted model. There is one other complication that we need to account for when we estimate the restricted model, which relates to the persistence in the observed data. When  $\gamma_f \leq 1/2$ , the unique stable solution for  $\pi_t$  in the restricted model is (see Rudd and Whelan (2006)):

$$\Delta\pi_t = \frac{\lambda}{1 - \gamma_f} \sum_{j=0}^{\infty} \left( \frac{\gamma_f}{1 - \gamma_f} \right)^j E_t(x_{t+j}) + u_t.$$

When the labor share,  $x_t$ , is stationary and not Granger-caused by  $\pi_t$ , as equation (4) above, it follows that  $\pi_t$  has a unit root. Thus, using lags of  $\pi_t$  as instruments would violate the conditions for the asymptotic theory given in Section 3. In other words, the identification-robust tests need not control size when the instruments are non-stationary. To avoid this problem, we use lags of  $\Delta\pi_t$  (instead of lags of  $\pi_t$ ) as instruments.<sup>12</sup> However, we should point out, for completeness, that the restricted model does not necessarily imply that inflation has a unit root. When  $x_t$  is Granger-caused by  $\pi_t$ , the dynamics of  $\pi_t$  are determined by a vector autoregression in  $(\pi_t, x_t)$ , and stationarity depends on the roots of the characteristic polynomial. Moreover, even in the case when  $x_t$  is not Granger-caused by  $\pi_t$ , when  $\gamma_f > 1/2$ , the solution of the model is of the form (see Rudd and Whelan 2006):

$$\pi_t = \left( \frac{1 - \gamma_f}{\gamma_f} \right) \pi_{t-1} + \frac{\lambda}{\gamma_f} \sum_{j=0}^{\infty} E_t(x_{t+j}) + u_t$$

---

<sup>12</sup>This is also done in Rudd and Whelan (2006). See also Mavroeidis (2006) for further details on this issue.

Table 2: Estimates of the NKPC

Parameter	unrestricted	restricted ( $\gamma_f + \gamma_b = 1$ )
$\lambda$	0.035 [-0.053, 0.170]	0.039 [-0.049, 0.167]
$\gamma_f$	0.773 [0.531, 1.091]	0.770 [0.556, 1.053]
$\gamma_b$	0.230 [-0.062, 0.451]	
$\gamma_f + \gamma_b - 1$	0.003 [-0.046, 0.059]	
Hansen test	2.486	2.492
$p$ value	0.478	0.477

The model is  $E[Z_t(\pi_t - c - \lambda x_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1})] = 0$ . Instruments include a constant and 3 lags of  $\pi_t, x_t$  (lags of  $\pi_t$  are replaced by lags of  $\Delta\pi_t$  in the restricted model). Point estimates are derived using CUE-GMM with Newey and West (1987) Weight matrix, Square brackets contain 95% confidence intervals based on the subset MQLR test. The estimation sample is 1960 quarter 1 to 2007 quarter 3.

which is stationary, whenever  $x_t$  is stationary. However, use of lags of  $\Delta\pi_t$  as instruments is robust across all values of  $\gamma_f$ , and independently of whether  $x_t$  is Granger-caused by  $\pi_t$ , so this is what we do hereafter. It is interesting to note that the empirical results do not depend much on whether lags of  $\pi_t$  or  $\Delta\pi_t$  are used in the instrument set.<sup>13</sup>

## 5.1 One-dimensional confidence sets

Under the restriction  $\gamma_f + \gamma_b = 1$ , model (2) can be re-written as

$$\Delta\pi_t = \lambda x_t + \gamma_f (\pi_{t+1} - \pi_{t-1}) + e_t. \quad (32)$$

Figure 5 reports one minus the  $p$  value plots associated with the subset tests on each of the parameters  $\lambda$  and  $\gamma_f$ . The 95% MQLR confidence bounds reported in the second column of Table 2 coincide with the intersections of the  $1 - p$  value plot for the MQLR statistic with the 0.95 line. We observe that the MQLR plot is almost indistinguishable from the KLM plot, leading to identical confidence sets. The confidence sets derived from the  $S$  test are wider, but shorter than the projection-based  $S_{pr}$  test, as expected. The JKLM test indicates no violation of the moment conditions for any value of the parameters that are within the MQLR and KLM 90% and 95% level confidence sets.

The following conclusions can be drawn from the above results. First, even though the slope of the Phillips curve is estimated to be positive, it is not significantly different from zero at the 35% level according to any of the tests. This conclusion is consistent with the findings of Rudd and Whelan (2006), and it is robust to using additional instruments, as in Galí and Gertler (1999) and Rudd and

<sup>13</sup>The confidence sets we report would be slightly wider if we used  $\pi_{t-1}, \pi_{t-2}$  and  $\pi_{t-3}$  instead of  $\Delta\pi_{t-1}$  and  $\Delta\pi_{t-2}$  as instruments. None of our conclusions is affected by this choice of instruments.

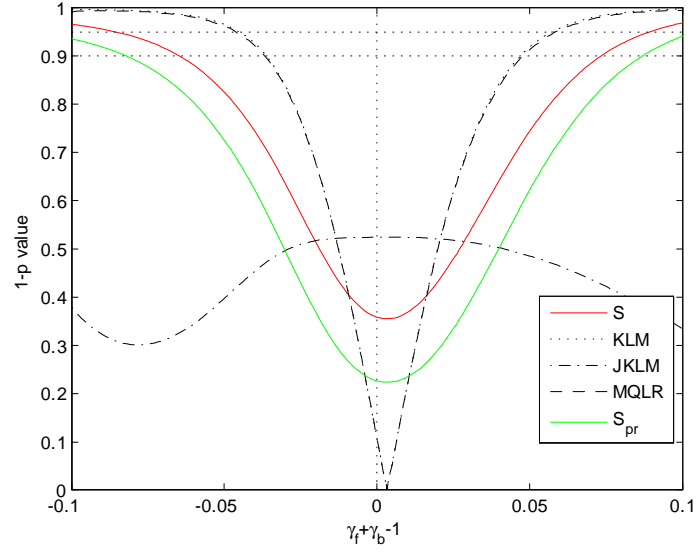


Figure 4: Identification-robust tests of the hypothesis  $\gamma_f + \gamma_b = 1$ . The figure reports one minus p-value plots for different values of coefficient  $\gamma_f + \gamma_b - 1$ .

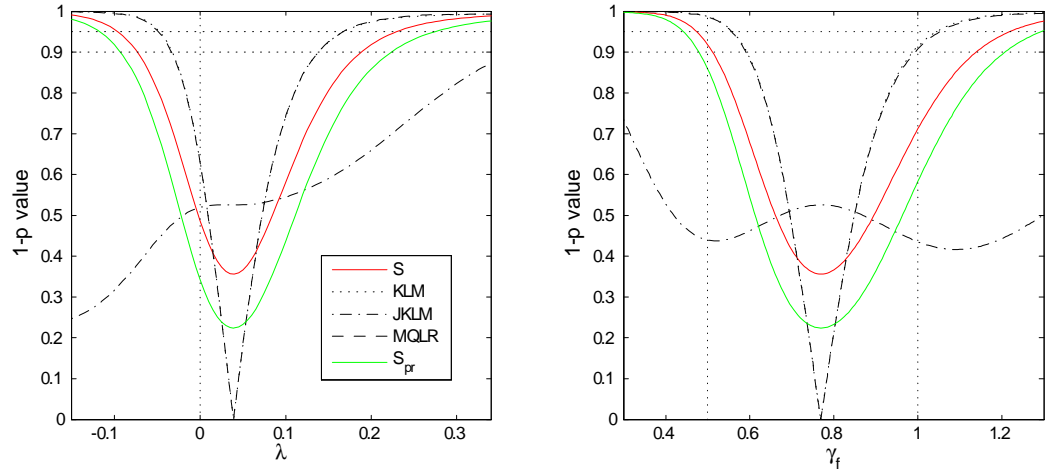


Figure 5: One minus p-value plots for the coefficients  $(\lambda, \gamma_f)$  in the NKPC model, under the restriction  $\gamma_f + \gamma_b = 1$ .

Whelan (2006). One interpretation is that the labor share is not a relevant determinant of inflation.<sup>14</sup> This result remains unchanged when we replace the labor share with other measures of marginal costs, such as the output gap or real output growth.

Second, the coefficient  $\gamma_f$  is not very accurately estimable. This is not surprising, given our earlier discussion about the effects of a small value of  $\lambda$  on the identifiability of  $\gamma_f$ . It is important to note, however, that  $\gamma_f$  is not completely unidentified. Specifically, most of the 95%-level confidence sets exclude values of  $\gamma_f$  close to zero. According to Galí and Gertler (1999), this can be interpreted as evidence of forward-looking behavior in price setting, though this interpretation is not uncontroversial (cf. Rudd and Whelan (2005) or Mavroeidis (2005)). The smallest 95% level confidence interval of  $\gamma_f$  is obtained by inverting the MQLR test, and it includes the value  $\gamma_f = 1$ , which, in view of the restriction  $\gamma_b = 1 - \gamma_f$ , suggests that the pure NKPC model fits the data, at the 5% level, though not at the 10% level. Nonetheless, the evidence is also consistent with large backward-looking dynamics ( $\gamma_f = 0.6$  is also in the 90% MQLR confidence set). With regards to the relative importance of forward versus backward looking behavior, i.e., whether  $\gamma_f > 1/2$  or  $\gamma_f < 1/2$ , using the subset MQLR test, we can infer with 95% confidence that forward-looking dynamics dominate.

These results are relatively insensitive to alternative choices of HAC estimators of the optimal weight matrix, or the choice of lag-truncation parameter in the Newey-West HAC estimator. A higher lag-truncation parameter leads to wider confidence intervals for  $\lambda$  and shifts the confidence intervals for  $\gamma_f$  slightly to the left, but the main conclusions, that  $\lambda$  is not significantly different from zero while  $\gamma_f$  is, remain unchanged.

Finally, the above results that the Phillips curve is weakly identified and relatively flat are consistent with evidence reported in the related literature on inflation forecasting, in the sense that variation in inflation forecasts is found to be predominantly captured by lagged inflation, and activity indicators are found not to matter a lot in inflation forecasts, see Stock and Watson (2008).

## 5.2 Two-dimensional confidence sets

Figure 6 reports joint 90% and 95% confidence sets for the coefficients  $\lambda$  and  $\gamma_f$  derived by inverting the S, MQLR, KLM and JKLM tests. These are the 0.9 and 0.95 contours of the graph of the function  $1 - p(\lambda, \gamma_f)$ , where  $p(\lambda, \gamma_f)$  is the  $p$  value of a test of a (joint) null hypothesis on  $\lambda$  and  $\gamma_f$ . Projection-based confidence sets can be inferred straightforwardly from these joint confidence sets. To facilitate the comparison of the projection-based confidence sets with their one-dimensional subset counterparts shown above, we superimpose the 95% level subset confidence sets for each statistic on the joint confidence sets in Figure 6 using straight lines. Notice that the subset confidence intervals are always smaller than the corresponding projection-based confidence intervals (the latter are not plotted because they can be easily inferred from the shaded areas). In the case of the S test, the 95% subset confidence intervals correspond almost exactly to 90% projection-based confidence intervals (this is also shown in Figure 5, where we also plot the  $p$  value of the projection  $S$  test). For the MQLR and KLM tests, the subset 95% intervals are even smaller than their 90% projection-based counterpart.

---

<sup>14</sup>This interpretation is not uncontroversial. Kuester, Mueller, and Stoelting (2007) argue that the baseline NKPC yields downwardly biased estimates of  $\lambda$  (the sensitivity of inflation to marginal costs) due to the omission of persistent cost-push shocks.

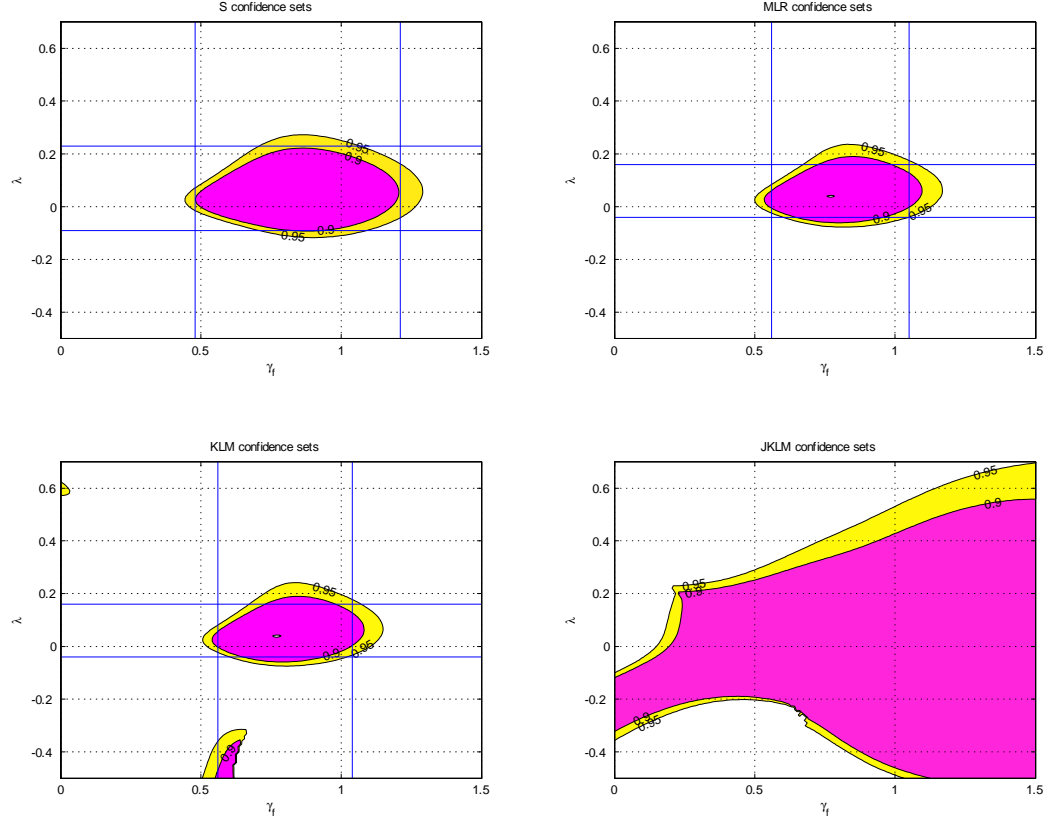


Figure 6: Confidence sets on  $(\lambda, \gamma_f)$  in the NKPC model. The shaded areas contain joint 90% and 95% confidence sets. The straight blue lines denote the bounds of the 95% subset confidence intervals for each parameter (not plotted for the JKLM statistic).



### 5.3 Structural parameters

As we pointed out in the Introduction, the semi-structural parameters  $\lambda$ ,  $\gamma_f$  and  $\gamma_b$  can be expressed as functions of some ‘deep’ structural parameters that relate to the microfoundations used to derive the model. Several different approaches are available in the literature, which give rise to different structural parameterizations. Since an exhaustive account of the different alternatives is beyond the scope of this paper, and because the results do not differ substantially across specifications, we shall focus here only on the parameterization proposed by Galí and Gertler (1999), namely,

$$\lambda = \frac{(1 - \omega)(1 - \alpha)(1 - \beta\alpha)}{\alpha + \omega[1 - \alpha(1 - \beta)]} \quad (33)$$

$$\gamma_f = \frac{\beta\alpha}{\alpha + \omega[1 - \alpha(1 - \beta)]} \quad (34)$$

$$\gamma_b = \frac{\omega}{\alpha + \omega[1 - \alpha(1 - \beta)]} \quad (35)$$

where  $\beta$  is the discount factor,  $\alpha$  is the probability that prices remain fixed in each period, and  $\omega$  is the fraction of backward-looking price setters. As Galí and Gertler (1999) explain, the restriction  $\beta = 1$  yields  $\gamma_f + \gamma_b = 1$ , and, letting  $\theta = (\alpha, \omega)'$ , the GMM moment vector can be written as:<sup>15</sup>

$$f_t(\theta) = Z_t \left[ \omega \Delta \pi_t - \alpha \Delta \pi_{t+1} - (1 - \omega)(1 - \alpha)^2 x_t \right].$$

We note that since the structural parameters are probabilities, the parameter space is bounded. This has implications for obtaining subset confidence intervals for  $\alpha$  and  $\omega$ , because the restricted estimates often fall on the boundary, and this violates the conditions needed to justify using subset tests, (it is well-known that this invalidates conventional GMM tests as well even when identification is strong, see Andrews (2002)). Thus, we only report projection-based confidence intervals for  $\alpha$  and  $\omega$ .

Table 3 reports the estimates and 95% confidence intervals for the structural parameters  $\alpha$  and  $\omega$ . The confidence intervals are obtained by projecting the two-dimensional MQLR-based confidence sets reported in Figure 7. The figure also reports confidence sets based on the S, KLM and JKLM statistics. We notice that although the point estimates reported in Table 3 are quite similar to those reported by Galí and Gertler (1999), the confidence intervals are much wider than the non-robust Wald-based confidence intervals reported by them. Specifically, the identification-robust confidence intervals are completely uninformative about the fraction of backward-looking price setters, suggesting that the data is consistent with the view that price setting behavior is predominantly forward-looking ( $\omega$  not significantly different from zero), as well as with the opposite view that it is backward-looking ( $\omega = 1$ ). This helps explain the conflicting results reported on this issue in the literature (e.g., Fuhrer and Moore (1995)). Regarding price stickiness, the 95% confidence interval on  $\alpha$  suggests that prices remain fixed for at least 2 quarters, which is consistent with micro evidence, but, in line with Figure 6, cannot rule out the possibility that the Phillips curve is completely flat ( $\alpha = 1$ , leading to  $\lambda = 0$ ).

<sup>15</sup>Note that, unlike 2-step or iterative GMM, the CUE GMM objective function is invariant to one-to-one transformations of the parameters.

Table 3: Estimates of the structural NKPC parameters

Parameter	CUE	95% conf. interval
$\alpha$	0.77	[0.56, 1]
$\omega$	0.23	[0, 1]

The confidence intervals are derived from the joint MQLR confidence set (see Figure 7) by projection.

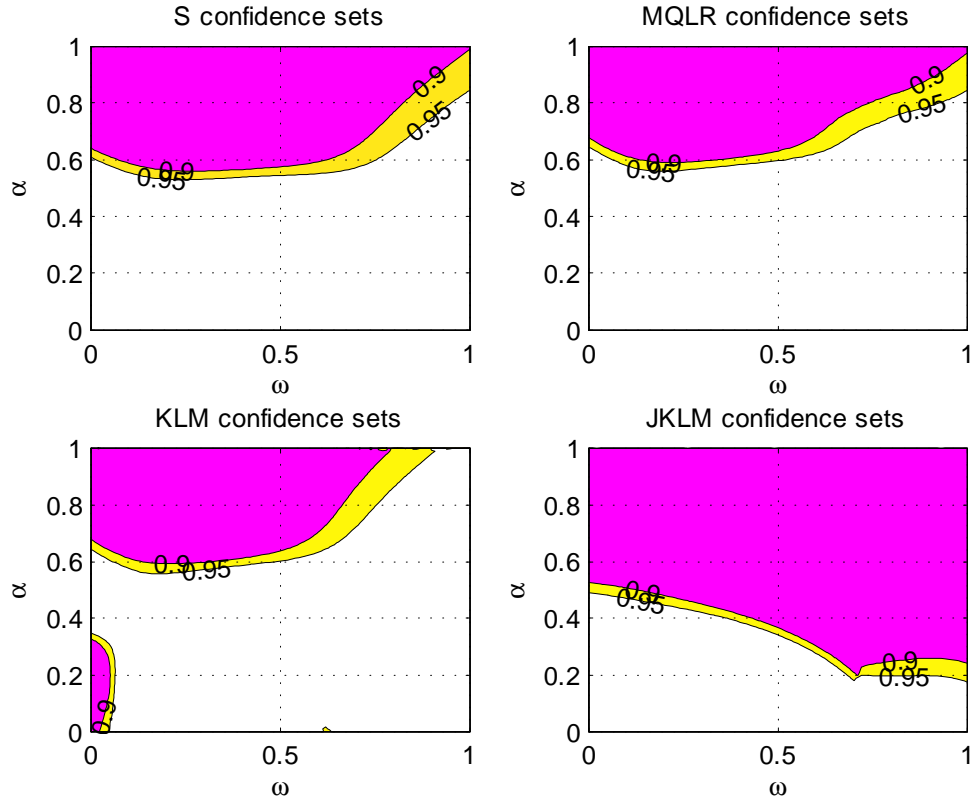


Figure 7: Confidence sets on the structural parameters  $(\alpha, \omega)$  in the NKPC model. The shaded areas contain joint 90% and 95% confidence sets.

Table 4: Subsample estimates of the NKPC

Parameter	1960 (1) - 1983 (4)	1984 (1) - 2007 (3)
$\lambda$	0.208 [-0.011, 0.620]	0.011 [-0.119, 0.231]
$\gamma_f$	0.805 [0.455, 1.570]	0.789 [0.566, 1.108]
Hansen test	1.315	1.754
$p$ value	0.725	0.625
sup $-S$ test	34.674	12.494
$p$ value	0.000	0.555
Proj. $p$ value	0.006	0.887

The model is  $E[Z_t(\Delta\pi_t - c - \lambda x_t - \gamma_f \Delta_2 \pi_{t+1})] = 0$ . Instruments are a constant and  $\Delta\pi_{t-1}, \Delta\pi_{t-2}, s_{t-1}, s_{t-2}, s_{t-3}$ . Point estimates derived using CUE-GMM with Newey and West (1987) Weight matrix. Square brackets contain 95% confidence intervals based on the subset MQLR test. sup- $S$  is the structural stability test of Caner (2007).

## 5.4 Structural stability

Stability of the parameters over time is essential for the model to be immune to the Lucas (1976) critique. Empirical studies typically address this point by estimating the model over subsamples. If the model is to be immune to the Lucas critique, it has to remain invariant to changes that occur elsewhere in the economy. Indeed, there is considerable evidence of changes in monetary policy (Clarida, Galí, and Gertler (2000)), and the nature of macroeconomic fluctuations (Justiniano and Primiceri (2006)). There seems to be an emerging consensus that the US economy has become more stable after the early 1980s, even though the sources of this stability are still under debate. This motivates us to estimate the NKPC over two subsamples of roughly the same size: 1960 quarter 1 to 1983 quarter 4 and 1984 quarter 1 to 2007 quarter 3.<sup>16</sup> The results are reported in Table 4.

The subsample estimates reveal that the coefficient  $\lambda$  is higher before 1984 and lower thereafter. However, the associated confidence intervals are so wide that we cannot conclude that  $\lambda$  differs between the two subsamples. The parameter  $\gamma_f$  is fairly stable across the two samples, though more precisely estimated after 1984. The hypothesis that the Phillips curve is flat ( $\lambda = 0$ ), can be rejected in the pre-1984 sample with greater confidence than for the post-1984 sample (the  $p$  values associated with this hypothesis are 0.06 and 0.85, respectively).

We test the hypothesis of no structural change using Caner's (2007) sup- $S$  statistic that tests for structural change at an unknown break date. The value of the structural change sup- $S$  statistic over all possible break dates in the middle 70% of the sample is 22.376, and the  $p$  value according to Caner's (2007) conservative (projection-based) bounding distribution (30) is 0.22, so we do not find evidence of structural change at conventional significance levels. However, if we were willing to assume that the model (i.e.,  $\lambda$  and  $\gamma_f$ ) is identified, the associated  $p$  value for the sup- $S$  test based

<sup>16</sup>Our choice of break point is motivated by the estimates of the break in output volatility reported by McConnell and Perez-Quiros (2000), see also Justiniano and Primiceri (2006). However, we note that the results are fairly robust to other choices of break date around 1984.

on the distribution given in (31) would be 0.039, so we would reject the null hypothesis of structural stability at the 5% level.<sup>17</sup> As we discussed in Section 3 above, it is presently unclear whether this last conclusion is robust to weak instruments. Thus, this is an important topic for future research, since the power advantages from using subset tests for structural change seem quite substantial.

Concerns over structural stability have led some researchers to focus their empirical analysis on the post-1984 sample, see for example Krause *et al* (2008). To assess the validity of this approach, we conduct structural stability sup- $S$  tests separately for the pre- and post-1984 samples and the results are reported in Table 4. In line with the above view, we find strong evidence of instability before 1984, and no evidence of instability thereafter.

## 6 Directions for future research

A natural response to the current finding that the NKPC is not well-identified based on a limited set of identifying restrictions is to look for more information. By a limited set of identifying restrictions we mean those implied by the assumption that (rational) forecast errors must be uncorrelated with the information that is available at the time expectations are formed. This assumption gives rise to a set of conditional moment restrictions that need to be converted to unconditional ones in order to use GMM inferential procedures, and this is typically done using the first few lags of a handful of macroeconomic variables as instruments. Hence, an obvious response to the identification problem is to include more variables in the instrument set, but there are limits to how many instruments one can use. This is because of the so called ‘many instruments’ problem, which biases GMM in the direction of Least Squares, see Newey and Windmeijer (2005). Recently, Andrews and Stock (2007) showed that, provided the number of instruments is not too large relative to the sample size, the identification-robust statistics remain size-correct in the instrumental variables regression model with many weak instruments, while Newey and Windmeijer (2005) obtained a similar result for the KLM statistic in GMM. Apart from the caveat that the number of instruments must grow sufficiently slowly relative to the sample size, substantial size distortion can arise in finite samples also through the use of a HAC estimator for the optimal GMM weight matrix. Unreported simulation results for the NKPC show that the size of the identification-robust statistics becomes sensitive to the number of instruments when a HAC estimator is used, and that the size distortions can be rather large for typical sample sizes. This leads us to caution against indiscriminate use of many instruments in the estimation of the NKPC. It also shows that further improvement of HAC estimators is an important topic for further research.

The previous discussion begs the question of how to select instruments out of the infinite set implied by the conditional moment restrictions. The standard optimality criterion for selecting instruments has been the asymptotic efficiency of the resulting GMM estimator, see e.g., Hansen and Sargent (1982) for linear rational expectations models like the NKPC. Since this criterion relies on the identification condition it is arguably problematic when instruments are potentially arbitrarily weak.<sup>18</sup> Alternative optimality criteria that do not require identification could be based on the power of the identification-robust tests. To the best of our knowledge, there are currently no results in this direction.

<sup>17</sup>As pointed out earlier, these results are robust to using the unrestricted specification (2) instead of (32).

<sup>18</sup>Bai and Ng (2008b) propose instrument selection methods that handle the case when the number of instruments is larger than the sample size. Their approach also requires identification.

Given the aforementioned difficulties with choosing a small number of instruments out of a very large information set, a practical alternative is to summarize the information contained in large datasets using principal components or factors, see Bai and Ng (2008a, 2008b) and Kapetanios and Marcellino (2006). This approach was used for inference on the NKPC by Beyer *et al* (2007). The reason why we think factors may be help to address the identification problem for the NKPC is that they have been found to be useful in forecasting inflation, see Stock and Watson (1999), and this is precisely what is needed for instruments to be relevant for the NKPC. Bai and Ng (2008a) showed that the use of estimated factors as instruments does not pose a generated instrument problem when the number of variables from which the factors are extracted is large, and therefore, the identification-robust procedures that we use in the paper can be applied without modification. It is therefore possible to check whether factors improve the identification of the NKPC for the US by repeating the analysis of Section 5 using a small number of factors as additional instruments.

Without imposing any further restrictions, an additional source of information that can be exploited for inference on the structural parameters is the structural stability of the model's parameters. This idea is proposed by Mavroeidis and Magnusson (2008), who show how the information contained in stability restrictions can be extracted using appropriately modified versions of standard structural stability tests. When evaluated under the null, these statistics have asymptotic distributions that are free from nuisance parameters even when identification is arbitrarily weak, and hence, they can be inverted to produce confidence sets with correct coverage. Such confidence sets have the appealing interpretation that they contain all the values of the structural parameters (i.e., all possible structures) that satisfy the moment conditions of the model and at the same time are immune to the Lucas critique. Mavroeidis and Magnusson (2008) show that the use of stability restrictions can improve significantly the identification of the parameters of the NKPC.

The above two approaches represent attempts to improve the usage of the identifying restrictions implied by the NKPC under the assumption of rational expectations. Since the NKPC is part of a macroeconomic system of equations, placing more structure on that system can yield additional identifying restrictions for the parameters of the model.

There are several ways in which this has been done in the literature. One approach, based on the methodology of Campbell and Shiller (1987), is to postulate a reduced-form VAR model for the endogenous variables, and use the restrictions on the reduced-form coefficients implied by the NKPC to estimate the structural parameters via minimum distance.<sup>19</sup> The VAR assumption is motivated by the long tradition of using vector autoregressions to model macroeconomic dynamics. To check whether this additional structure improves the identification of the NKPC, it is necessary to use identification-robust methods.

Another approach is to use structural equations for each of the endogenous variables in the model, e.g., to postulate a dynamic stochastic general equilibrium (DSGE) model, as in Beyer *et al* (2007). The DSGE model typically implies tight restrictions on the reduced-form dynamics of the endogenous variables. It is useful to think about these restrictions in the special case where the model is equivalent to a structural VAR. This makes it relatively easy to see that identification is achieved through two types of restrictions: (i) restrictions on the dynamics and (ii) restrictions on the covariance matrix of the structural shocks. Covariance restrictions are tantamount to using the identified structural errors

---

<sup>19</sup>This approach is used by Sbordone (2002). An alternative to minimum distance is maximum likelihood, see Fanelli (2008).

from the other equations of the system, e.g., monetary policy or demand shocks, as instruments for the NKPC equation. This can only be done when the model is completely specified, i.e., by using a full-information approach. When we are interested in the estimation of a single equation, e.g., the NKPC, a limited-information approach can only make use of restrictions on the dynamics, because in order to impose covariance restrictions one needs to be able to identify the structural shocks. In fact, as was pointed out by Lubik and Schorfheide (2004), there are plausible situations in which the model may *only* be identifiable through covariance restrictions.

So far, there has been little work on developing and implementing identification-robust methods for inference on systems of structural equations. Dufour *et al* (2007) propose a multi-equation generalization of the AR statistic and apply it to the NKPC as part of a small-scale DSGE model. Their method consists of stacking the moment conditions of the NKPC together with those implied by each of the other equations in the DSGE model, and computing the AR statistic for a point null hypothesis on all of the parameters. This statistic can be inverted to obtain a joint confidence set on all the parameters. The method of Dufour *et al* (2007) cannot be used to impose covariance restrictions on the structural shocks of the DSGE model. Doing so requires a method for extracting the shocks from the observed data. Developing and implementing such a method is an important topic for future research.

The asymptotic theory for the identification robust tests given in Section 3 relies on the assumption that the moment conditions and their Jacobian satisfy a normal central limit theorem, which places some restrictions on the dependence of the data. This assumption is not innocuous, since macroeconomic time series are highly persistent and potentially trending, and the usual de-trending procedures, such as the Hodrick-Prescott filter, may fail to remove the underlying trends. Gorodnichenko and Ng (2007) study the problems that arise when the detrending procedure used is misspecified, and propose methods that are robust to near unit roots in the data. Since Gorodnichenko and Ng (2007) work under the assumption that the model is identified, it is important to see if their results can be extended to inference procedures that are robust to identification failure, as well. In fact, it is straightforward to obtain a version of the S statistic that is robust to near unit roots,<sup>20</sup> but it is nontrivial to obtain similar results for the KLM and MQLR statistics.

Finally, several authors have expressed concerns about the specification of the NKPC. Some argue that the standard hybrid NKPC is misspecified due to omitted dynamics (e.g., Rudd and Whelan (2007)) while others allow for autocorrelated structural shocks (e.g., Smets and Wouters (2007)). Mavroeidis (2005) showed that the Hansen test of overidentifying restrictions has low power against such misspecification, and this can lead to spurious identification. Since the S and JKLM statistics have power against violation of the moment conditions, the confidence sets based on inverting those statistics may turn out to be smaller than the confidence sets based on the KLM and MQLR statistics. It is therefore interesting to study whether the latter are more robust to mild violations of the moment conditions due to omitted dynamics or autocorrelated structural shocks.

## 7 Conclusions

In this paper, we discussed identification-robust inference procedures for inference on the parameters of the new Keynesian Phillips curve, and applied those methods to postwar US data. Our results showed

---

<sup>20</sup>Mavroeidis, Chevillon, and Massmann (2008) propose such a version in a related model.

that the parameters of the model are weakly identified, and this helps explain the conflicting estimates reported in the literature. However, the use of powerful identification-robust tests revealed that the model is not completely unidentified, thus making it possible to reach useful economic conclusions, such as that forward-looking dynamics dominate backward-looking behavior.

We hope that this paper will help convince applied researchers that the use of identification-robust procedures such as the MQLR statistic entails no cost in terms of sacrificing efficiency for robustness to weak instruments. Hence, this method completely obviates the need to conduct identification analysis or rely on prior identification assumptions and pretests. Therefore, since weak instruments problems are pervasive in this area, we recommend that this procedure be used for inference on the NKPC and other structural macroeconomic models, as well.

## Appendix

The statistic proposed by Cragg and Donald (1997) to test that the rank of a  $k \times p$  matrix  $L$  is equal to  $p - 1$ , with  $\hat{L}$  being a consistent estimator of  $L$  and  $\hat{V}_{\hat{L}}^{-1}$  a consistent estimator of the variance of  $\hat{L}$ , can be written as

$$CD = \min_{A_2 \in \mathbb{R}^{k \times (p-1)}, \varphi \in \mathbb{R}^{p-1}} T \text{vec} \left( \hat{L} - A_2(-\varphi : I_{p-1}) \right)' \hat{V}_{\hat{L}}^{-1} \text{vec} \left( \hat{L} - A_2(-\varphi : I_{p-1}) \right).$$

If used for  $rk(\beta_0)$ , i.e., to test that  $J(\theta) = E \left( \lim_{T \rightarrow \infty} \frac{\partial}{\partial \theta'} f_T(\theta_0) \right)$  is of rank  $p - 1$ , using  $\hat{D}_T = \hat{D}_T(\tilde{\alpha}(\beta_0), \beta_0)$  and  $\hat{V}_{\theta\theta.f} = \hat{V}_{\theta\theta.f}(\tilde{\alpha}(\beta_0), \beta_0)$ , the above statistic reads

$$CD(\beta_0) = \min_{A_2 \in \mathbb{R}^{k \times (p-1)}, \varphi \in \mathbb{R}^{p-1}} T \text{vec} \left( \hat{D}_T - A_2(-\varphi : I_{p-1}) \right)' \hat{V}_{\theta\theta.f}^{-1} \text{vec} \left( \hat{D}_T - A_2(-\varphi : I_{p-1}) \right). \quad (36)$$

This can be equivalently written in terms of the scaled statistic

$$\text{vec} \left( \left[ \hat{D}_T - A_2(-\varphi : I_{p-1}) \right] \begin{pmatrix} 1 & 0 \\ \varphi & I_{p-1} \end{pmatrix} \right) = \begin{pmatrix} \hat{D}_T \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \\ \text{vec} \left( \hat{D}_T \begin{pmatrix} 0 \\ I_{p-1} \end{pmatrix} - A_2 \right) \end{pmatrix},$$

whose asymptotic variance is given by

$$\left( \begin{pmatrix} 1 & 0 \\ \varphi & I_{p-1} \end{pmatrix} \otimes I_k \right)' \hat{V}_{\theta\theta.f} \left( \begin{pmatrix} 1 & 0 \\ \varphi & I_{p-1} \end{pmatrix} \otimes I_k \right) = \begin{pmatrix} A & B \\ B' & C \end{pmatrix}$$

where

$$\begin{aligned} A &= \left( \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \otimes I_k \right)' \hat{V}_{\theta\theta.f} \left( \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \otimes I_k \right) \\ B &= \left( \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \otimes I_k \right)' \hat{V}_{\theta\theta.f} \left( \begin{pmatrix} 0 \\ I_{p-1} \end{pmatrix} \otimes I_k \right) \\ C &= \left( \begin{pmatrix} 0 \\ I_{p-1} \end{pmatrix} \otimes I_k \right)' \hat{V}_{\theta\theta.f} \left( \begin{pmatrix} 0 \\ I_{p-1} \end{pmatrix} \otimes I_k \right). \end{aligned}$$

Using the partitioned inverse formula, the inverse of the above expression can be written as

$$\begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} -A^{-1}B \\ I_{k(p-1)} \end{pmatrix} (C - B'A^{-1}B)^{-1} \begin{pmatrix} -A^{-1}B' \\ I_{k(p-1)} \end{pmatrix}'.$$

Hence, the expression for the  $CD(\beta_0)$  statistic (36) can be written as

$$\min_{\varphi \in \mathbb{R}^{p-1}} \left[ Q_1(\varphi) + \min_{A_2 \in \mathbb{R}^{k \times (p-1)}} Q_2(\varphi, A_2) \right],$$



where

$$Q_1(\varphi) = \begin{pmatrix} 1 \\ \varphi \end{pmatrix}' \hat{D}_T' \left[ \left( \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \otimes I_k \right)' \hat{V}_{\theta\theta.f} \left( \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \otimes I_k \right) \right]^{-1} \hat{D}_T \begin{pmatrix} 1 \\ \varphi \end{pmatrix}$$

and

$$\begin{aligned} Q_2(\varphi, A_2) &= g(\varphi, A_2)' (C - B' A^{-1} B)^{-1} g(\varphi, A_2), \\ g(\varphi, A_2) &= \text{vec} \left( \hat{D}_T \begin{pmatrix} 0 \\ I_{p-1} \end{pmatrix} - B' A^{-1} \hat{D}_T \begin{pmatrix} 1 \\ \varphi \end{pmatrix} - A_2 \right). \end{aligned}$$

Now, since  $g(\varphi, A_2)$  can be made equal to zero by solving for  $A_2$ ,  $\min_{A_2 \in \mathbb{R}^{k \times (p-1)}} Q_2(\varphi, A_2) = 0$ , and hence,  $CD(\beta_0) = \min_{\varphi \in \mathbb{R}^{p-1}} Q_1(\varphi)$ , which is identical to the expression for  $rk(\beta_0)$  in (22). Note, that this expression is akin to a continuously updated linear GMM objective function for the unknown parameter vector  $\varphi$ .

The above specification of  $CD(\beta_0)$  can also be used to show that  $S(\tilde{\beta}) < rk(\tilde{\beta})$ . This holds since, in case of linear moment equations, like, for example,  $f_t(\theta) = \left( \begin{pmatrix} 1 \\ -\theta \end{pmatrix} \otimes I_k \right) (Y_t \otimes Z_t)$ , with  $Z_t : k \times 1$  and  $Y_t = \begin{pmatrix} y_t \\ X_t \end{pmatrix} : (p+1) \times 1$ ,  $X_t : p \times 1$ ; the CUE objective function  $Q(\theta)$  (11) results from a reduced rank objective function as well, see Kleibergen (2007),

$$Q(\theta) = \min_{A_1 \in \mathbb{R}^{k \times p}} T \text{vec} \left[ F_T - A_1(\theta : I_p) \right]' \hat{W}^{-1} \text{vec} \left[ F_T - A_1(\theta : I_p) \right],$$

with  $F_T = \frac{1}{T} \sum_{t=1}^T Z_t Y_t'$  and  $W = \text{var} \left[ \sqrt{T} \text{vec}(F_T) \right]$ ,  $\hat{W}$  is a consistent estimator of  $W$ . The estimator of  $A_1$  that results from the above specification corresponds to  $\hat{D}_T(\theta)$ . This can be shown by using the decomposition used for  $CD(\beta_0)$ . The function on the right hand side can then be specified as

$$\begin{aligned} T \text{vec} \left[ F_T - A_1(\theta : I_p) \right]' \hat{W}^{-1} \text{vec} \left[ F_T - A_1(\theta : I_p) \right] &= \\ T f_T(\theta)' \hat{V}_{ff}(\theta)^{-1} f_T(\theta) + T \text{vec} \left[ \hat{D}_T(\theta) - A_1 \right]' \hat{V}_{\theta\theta.f}(\theta)^{-1} \text{vec} \left[ \hat{D}_T(\theta) - A_1 \right], \end{aligned}$$

which uses the fact that  $\hat{V}_{ff}(\theta) = \left( \begin{pmatrix} 1 \\ -\theta \end{pmatrix} \otimes I_k \right)' \hat{W} \left( \begin{pmatrix} 1 \\ -\theta \end{pmatrix} \otimes I_k \right)$  and  $\hat{V}_{\theta\theta.f}(\theta)$  results from  $\hat{W}$  in a similar fashion.

The specification of the reduced rank objective function coincides with that of the Cragg and Donald (1997) rank statistic that we stated above so  $Q(\tilde{\theta})$  measures the rank reduction of  $E(F_T)$  from rank  $p+1$  to  $p$  while  $rk(\tilde{\theta})$  measures the rank reduction of  $E(F_T)$  from rank  $p$  to  $p-1$ . When we reduce the rank of a matrix, we first remove from its base the column that has the smallest contribution to an objective function that measures its relative distance from a lower rank matrix, like, for example, the Cragg and Donald (1997) rank statistic. This implies that there is a sequential ordering of the values of rank statistics that measure the decline in rank of a matrix. The value of the statistic that measures a decline in the rank of a matrix from  $p+1$  to  $p$  is exceeded by the value of the statistic that measures a decline in rank of the same matrix from  $p$  to  $p-1$ . Since  $S(\tilde{\theta})$  corresponds to the Cragg and Donald (1997) rank statistic that measures a decline in the rank of  $E(F_T)$  from  $p+1$  to  $p$  and  $rk(\tilde{\theta})$  corresponds to the Cragg and Donald (1997) rank statistic that measures a decline in rank value of  $E(F_T)$  from  $p$  to  $p-1$ , it therefore holds that  $rk(\tilde{\theta}) > S(\tilde{\theta})$ .

## References

- Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statistics* 20, 46–63.
- Andrews, D. W. K. (2002). Generalized Method of Moments Estimation When a Parameter Is on a Boundary. *Journal of Business & Economic Statistics* 20(4), 530–44.
- Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60(4), 953–966.
- Andrews, D.W.K., M.J. Moreira and J.H. Stock (2006). Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression. *Econometrica* 74, 715–752.
- Andrews, D.W.K., and J.H. Stock (2005). Inference with weak instruments. NBER Technical Working Papers 0313, National Bureau of Economic Research, Inc.
- Andrews, D.W.K., and J.H. Stock (2007). Testing with many weak instruments. *Journal of Econometrics* 127(1), 24–46.
- Bai, J. and S. Ng (2008a). Instrumental variable estimation in a data rich environment. mimeo, NYU.
- Bai, J. and S. Ng (2008b). Selecting instrumental variables in a data rich environment. mimeo, NYU.
- Beyer, A., R. E. A. Farmer, J. Henry, and M. Marcellino (2007). Factor analysis in a model with rational expectations. NBER Working Papers 13404, National Bureau of Economic Research, Inc.
- Buiter, W. and I. Jewitt (1989). Staggered wage setting with real wage relativities: Variations on a theme of Taylor. In W. Buiter (Ed.), *Macroeconomic Theory and Stabilization Policy*, pp. 183–199. Ann Arbor: University of Michigan Press.
- Calvo, G.A. (1983). Staggered Prices in a Utility-Maximizing Framework. *Journal of Monetary Economics* 12, 383–398.
- Campbell, J. Y. and R. J. Shiller (1987). Cointegration and tests of present value models. *Journal of Political Economy* 95, 1062–1088.
- Caner, M. (2007). Boundedly pivotal structural change tests in continuous updating GMM with strong, weak identification and completely unidentified cases. *Journal of Econometrics* 137, 28–67.
- Canova, F. and L. Sala (2005, May). Back to square one: identification issues in dsge models. Economics Working Papers 927, Department of Economics and Business, Universitat Pompeu Fabra.
- Chaudhuri, S. (2007). Testing of Hypotheses for Subsets of Parameters. Technical report, Department of Economics, University of Washington. Working Paper.
- Chaudhuri, S., Richardson, T., Robins, J. and E. Zivot (2007). Split-Sample Score Tests in Linear Instrumental Variables Regression. Working Paper UWEC-2007-10, Dept. of Economics, University of Washington.

- Christiano, L. J., M. Eichenbaum, and C. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *J. Political Economy* 113, 1–45.
- Clarida, R., J. Galí, and M. Gertler (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics* 115, 147–180.
- Cragg, J.C. and S.G. Donald (1997). Inferring the rank of a matrix. *Journal of Econometrics* 76, 223–250.
- Dufour, J.-M. (1997). Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica* 65, 1365–388.
- Dufour, J.-M. (2003). Identification, weak instruments and statistical inference in econometrics. *Canadian Journal of Economics* 36(4), 767–808. Presidential Address to the Canadian Economics Association.
- Dufour, J.-M. and J. Jasiak (2001). Finite Sample Inference for Simultaneous Equations and Models with Unobserved and Generated Regressors. *International Economic Review* 42, 815–844.
- Dufour, J.-M., L. Khalaf, and M. Kichian (2006). Inflation dynamics and the new keynesian phillips curve: An identification robust econometric analysis. *Journal of Economic Dynamics and Control* 30(9-10), 1707–1727.
- Dufour, J.-M., L. Khalaf, and M. Kichian (2007). Structural multi-equation macroeconomic models: A system-based estimation and evaluation approach. Discussion paper, Bank of Canada.
- Dufour, J.-M. and M. Taamouti (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica* 73, 1351–1365.
- Dufour, J.-M. and M. Taamouti (2007). Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. *Journal of Econometrics* 139, 133–153.
- Fanelli, L. (2008). Testing the New Keynesian Phillips Curve Through Vector Autoregressive Models: Results from the Euro Area\*. *Oxford Bulletin of Economics and Statistics* 70(1), 53–66.
- 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.
- Gorodnichenko, Y. and S. Ng (2007). Estimation of dsge models when the data are persistent. Technical report. Presented at NBER Summer Institute.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.
- Hansen, L.P., J. Heaton and A. Yaron (1996). Finite Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics* 14, 262–280.
- Hansen, L. P. and T. J. Sargent (1982). Instrumental variables procedures for estimating linear rational expectations models. *Journal of Monetary Economics* 9, 263–296.
- Jondeau, E. and H. Le Bihan (2008). Examining bias in estimators of linear rational expectations models under misspecification. *Journal of Econometrics* 143(2), 375–395.

- Justiniano, A. and G. E. Primiceri (2006). The time varying volatility of macroeconomic fluctuations. NBER Working Papers 12022, National Bureau of Economic Research, Inc.
- Kapetanios, G. and M. Marcellino (2006). Factor-gmm estimation with large sets of possibly weak instruments. Working Papers 577, Queen Mary, University of London, Department of Economics.
- Kleibergen, F. (2002). Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression. *Econometrica* **70**, 1781–1803.
- Kleibergen, F. (2005). Testing Parameters in GMM without assuming that they are identified. *Econometrica* **73**, 1103–1124.
- Kleibergen, F. (2007). Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics* **139**, 181–216.
- Kleibergen, F. (2008). Size Correct Subset Statistics in the Linear IV Regression Model. Technical Report, Brown University.
- Kleibergen, F. and R. Paap (2006). Generalized Reduced Rank Tests using the Singular Value Decomposition. *Journal of Econometrics* **133**, 97–126.
- Kleibergen, F. and S. Mavroeidis (2008). Inference on subsets of parameters in GMM without assuming identification. Working Paper, Brown University.
- Krause, M., T. A. Lubik, and D. Lopez-Salido (2008). Inflation Dynamics with Search Frictions: A Structural Econometric Analysis. Working Paper 08-01, Richmond Fed.
- Kuester, K., G. Mueller, and S. Stoelting (2007). Is the New Keynesian Phillips Curve Flat? Working paper 809.
- Lewbel, A. (1991). The Rank of Demand Systems: Theory and Nonparametric Estimation. *Econometrica* **59**, 711–730.
- 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 Philips Curve and Labor Markets.*, Carnegie-Rochester Conference Series on Public Policy. Amsterdam: North-Holland.
- Ma, A. (2002). GMM estimation of the New Keynesian Phillips Curve. *Economics Letters* *76*, 411–417.
- Martins, L. F. and V. J. Gabriel (2006). Robust estimates of the new keynesian phillips curve. Department of Economics Discussion Papers 0206, Department of Economics, University of Surrey.
- Mavroeidis, S. (2005). Identification issues in forward-looking models estimated by GMM with an application to the Phillips Curve. *Journal of Money Credit and Banking* *37*(3), 421–449.
- Mavroeidis, S. (2006). Testing the New Keynesian Phillips Curve without assuming identification. Brown Economics Working Paper No. 2006-13, Brown University. Available at SSRN: <http://ssrn.com/abstract=905261>.

- Mavroeidis, S., G. Chevillon, and M. Massmann (2008). Inference in models with adaptive learning, with an application to the new Keynesian Phillips curve. mimeo, Brown University.
- Mavroeidis, S. and L. M. Magnusson (2008). Identifying Euler equation models via stability restrictions. mimeo, Brown University.
- McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the united states: What has changed since the early 1980's? *The American Economic Review* 90(5), 1464–1476.
- Moreira, M.J., (2003). A Conditional Likelihood Ratio Test for Structural Models. *Econometrica* 71, 1027–1048.
- Nason, J. M. and G. W. Smith (2005). Identifying the new keynesian phillips curve. Working Paper 2005-01, Federal Reserve Bank of Atlanta.
- Newey, W.K. and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing. In R. Engle and D. McFadden (Ed.), *Handbook of Econometrics, Volume 4*, Chapter 36, pp. 2113–2148. Elsevier Science B.V.
- Newey, W. K. and K. D. West (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708.
- Newey, W. and F. Windmeijer (2005). Gmm with many weak moment conditions. CeMMAP working papers CWP18/05, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- Pesaran, M. H. (1987). *The limits to Rational Expectations*. Oxford: Blackwell Publishers.
- Phillips, P.C.B. (1983). Exact Small Sample Theory in the Simultaneous Equations Model. In Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics, Vol.1*. North-Holland Publishing Co., Amsterdam.
- Robin, J.-M. and R.J. Smith (2000). Tests of Rank. *Econometric Theory* 16, 151–175.
- Robins, J.M. (2004). Optimal Structural Nested Models for Optimal Sequential Decisions. In Lin, D.Y. and P. Heagerty (Ed.), *Proceedings of the Second Seattle Symposium on Biostatistics*. New York, Springer.
- Rothenberg, T.J. (1984). Approximating the Distributions of Econometric Estimators and Test Statistics. In Z. Griliches and M.D. Intriligator (Ed.), *Handbook of Econometrics, Volume 2*, Chapter 15, pp. 881–935. Elsevier Science B.V.
- Rudd, J. and K. Whelan (2005). New tests of the new-keynesian phillips curve. *Journal of Monetary Economics* 52(6), 1167–1181.
- Rudd, J. and K. Whelan (2006). Can rational expectations sticky-price models explain inflation dynamics? *American Economic Review* 96(1), 303–320.
- Rudd, J. and K. Whelan (2007). Modelling Inflation Dynamics: A Critical Survey of Recent Research. *Journal of Money Credit and Banking* 39, 155–170.
- 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 (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *AER* 97(3), 586–606.

- Stock, J. and M. Watson (1999). Forecasting inflation. *Journal of Monetary Economics* 44(2), 293–335.
- Stock, J. H. and M. W. Watson (2008). Phillips Curve Inflation Forecasts. NBER Working Papers 14322, National Bureau of Economic Research.
- Stock, J.H. and J.H. Wright (2000). GMM with Weak Identification. *Econometrica* 68, 1055–1096.
- Stock, J. H., J. H. Wright, and M. Yogo (2002). A Survey of Weak instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics* 20, 518–530.
- Stock, J.H. and M. Yogo (2005). Testing for Weak Instruments in Linear IV Regression. In D.W.K. Andrews and J.H. Stock (Ed.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 80–108. Cambridge: Cambridge University Press. Prepared for Festschrift in honor of Thomas Rothenberg.
- West, K.D. (1997). Another heteroscedasticity- and autocorrelation-consistent covariance matrix estimator. *Journal of Econometrics* 76, 171–191.
- White, H. (1984). *Asymptotic Theory for Econometricians*. Academic Press.
- Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.
- Zivot, E., R. Startz, and C. R. Nelson (1998). Valid Confidence Intervals and Inference in the Presence of Weak Instruments. *International Economic Review* 39, 1119–1144.