LEANDRO M. MAGNUSSON
SOPHOCLES MAVROEIDIS

Identification-Robust Minimum Distance Estimation of the New Keynesian Phillips Curve

Limited-information identification-robust methods on the indexation and price rigidity parameters of the New Keynesian Phillips Curve yield very wide confidence intervals. Full-information methods impose more restrictions on the reduced-form dynamics and thus make more efficient use of the information in the data. However, such methods are also subject to weak instrument problems. We propose identification-robust minimum distance methods for exploiting these additional restrictions and show that they yield considerably smaller confidence intervals for the coefficients of the model compared to their limited-information generalized method of moments counterparts. In contrast to previous studies, we find evidence of partial but not full indexation, and obtain sharper inference on the degree of price stickiness. However, this parameter remains weakly identified.

JEL codes: C22, E31
Keywords: weak identification, GMM, present value models.

The new Keynesian Phillips Curve (NKPC) is a forward-looking model of sticky prices, according to which inflation is driven by the expected discounted stream of future marginal costs. In its purely forward-looking specification, the model is at odds with U.S. data over the postwar period, and it is typically replaced by a hybrid version that includes also lagged inflation

$$\pi_t = \lambda x_t + \gamma_f E_t(\pi_{t+1}) + \gamma_b \pi_{t-1} + \varepsilon_t, \tag{1}$$

where π_t denotes inflation, x_t is a measure of marginal costs, ε_t is an exogenous cost-push or markup shock (which could be zero), and E_t denotes expectations conditional on information available at time t (see Woodford 2003 for details).

Leandro M. Magnusson *is Assistant Professor, Tulane University* (*E-mail:* lmagnuss@tulane.edu). Sophocles Mavroeidis *is Assistant Professor, Brown University* (*E-mail:* Sophocles_Mavroeidis@Brown.edu).

Received February 17, 2009; and accepted in revised form November 20, 2009.

Journal of Money, Credit and Banking, Vol. 42, No. 2–3 (March–April 2010)

Equation (1) has been estimated in the literature using so-called limited- or full-information methods. Limited-information or "single-equation" methods do not require the specification of a model for the forcing variable x_t, and their results are therefore more general than those derived using full-information methods. A prominent limited-information approach for the estimation of the NKPC, popularized in a seminal paper by Galí and Gertler (1999), is based on replacing expectations of future inflation in equation (1) by their realization. The resulting equation is estimated by the generalized method of moments (GMM), using predetermined variables as instruments. It is by now well established that estimation of the NKPC may be subject to weak instrument problems (see, e.g., Kleibergen and Mavroeidis 2009 and the references therein). Recently, a number of studies applied identification-robust GMM-based tests to the NKPC, and reported very wide confidence intervals (see Dufour et al. 2006, 2008, Kleibergen and Mavroeidis 2009).

In contrast, full-information methods specify a complete model for the dynamics of inflation and the forcing variable. Various full-information approaches have been used for the estimation of the NKPC, which differ in terms of whether the model of the forcing variable is structural or reduced form, as well as in terms of econometric methodology, that is, minimum distance (MD) or likelihood-based inference. One such approach used by Sbordone (2002, 2005) is MD estimation based on a reduced-form vector autoregressive model (VAR) for π_t and x_t.[1] It is clear that this approach imposes more restrictions on the underlying data-generating process than limited-information GMM, so we expect it to make more efficient use of the information in the data. Thus, in view of the aforementioned negative results on the identification of the NKPC, reviewed in Kleibergen and Mavroeidis (2009), our main objective in this paper is to investigate the extent to which imposing more restrictions on the reduced-form dynamics reduces the uncertainty about the NKPC parameters. We do so using the identification-robust methods for MD that were recently developed in a different setting by Magnusson (Forthcoming). These are the MD versions of the identification-robust GMM tests proposed by Stock and Wright (2000) and Kleibergen (2005). Even though the focus of the present study is on the Phillips curve, our paper also provides a methodological contribution toward the development of identification-robust methods for full-information inference in dynamic stochastic general equilibrium (DSGE) models.[2]

Our analysis starts with a discussion of the identification-robust MD methods and a comparison with their limited-information GMM counterparts in the context of the NKPC. We clarify the sense in which the MD approach to the estimation of the NKPC brings in more information, and we exploit this feature to obtain identification-robust MD methods that are more efficient than the their GMM counterparts. Specifically,

1. MD was also used by Christiano, Eichenbaum, and Evans (2005) based on a structural model of the dynamics.

2. A related study in this direction is the paper by Dufour et al. (2008). This study differs from ours in that it does not impose assumptions on the reduced-form dynamics directly but rather it exploits the correlation of the structural errors in a system of structural equations.

we show that the difference in the two approaches essentially lies in the way they use past data to proxy the expectations of future inflation. This is akin to the issue of iterative versus direct multistep forecasting, which has been studied by the forecasting literature (see, e.g., Schorfheide 2005, Marcellino, Stock, and Watson 2006). Iterative forecasting corresponds to what the MD approach does, and it is more efficient, although less robust, than the direct multistep forecasting, which corresponds to GMM. We report power curves that show considerable power advantages of MD over GMM. These results are consistent with our subsequent empirical findings.

Our empirical analysis is based on a microfounded version of the NKPC that was studied by Sbordone (2005), according to which the so-called "semistructural" parameters λ, γ_f, and γ_b in equation (1) are functions of some deeper structural parameters that measure the degree of price stickiness and the degree of indexation of prices to past inflation. Following Sbordone (2005), we model the reduced-form dynamics of (π_t, x_t) by a VAR, and we use the labor share as a proxy for marginal costs, in accordance with Galí and Gertler (1999). We estimate the model using quarterly data from the U.S. economy over the period 1984 to 2008.

Our main finding is that exploiting the restrictions on the reduced-form dynamics reduces the uncertainty about the structural parameters of the NKPC considerably: the MD identification-robust confidence sets on the price rigidity and indexation parameters are about half the size of their GMM counterparts. Specifically, we find evidence of partial but not full indexation, so we can reject the pure forward-looking version of the NKPC, as well as the model with full indexation, which was used in Christiano, Eichenbaum, and Evans (2005). In terms of the price rigidity parameter, our confidence intervals are smaller than their GMM counterparts, but they are still fairly wide. This adds further weight to the view that the average duration of prices is hard to estimate accurately from aggregate macroeconomic data. Therefore, one important contribution of this paper is to show that the MD approach is also subject to the weak identification problem for the NKPC.

Use of identification-robust MD methods enables us to evaluate some of the criticism of the NKPC reported in previous studies. For example, Kurmann (2005) criticizes Sbordone's (2005) findings by showing that accounting for the uncertainty in the estimation of the VAR coefficients results in large uncertainty about the fit of the NKPC to the U.S. data. His results informally suggest that there is considerable uncertainty associated with the NKPC parameters, as well, since alternative calibrations produce similar fit. The present paper shows, using formal statistical tests, that this is indeed the case.

A related paper by Fuhrer and Olivei (2004) proposes an alternative way of selecting instruments for GMM, which they refer to as an "optimal instruments approach." This approach imposes the constraints implied by the structural model on the reduced-form model that is used to project future inflation on lagged data, so it exploits essentially the same restrictions as the MD approach we study here. Fuhrer and Olivei find, too, that their approach is more efficient than conventional GMM. However, similar to the MD approach, Fuhrer and Olivei's approach is not immune to weak identification problems, either, since the optimal instruments may still be arbitrarily weak in a given

application, and this needs to be taken into account for inference. For example, usual confidence intervals based on standard error bands around point estimates are not robust, irrespective of the estimation method on which they are based, and therefore, they are not reliable (see Dufour 1997).[3]

The structure of the paper is as follows. Section 1 presents the version of the NKPC model that we estimate and derives the identification-robust MD test statistics. Section 2 discusses the difference of the MD tests from their GMM counterparts. Section 3 reports the empirical results, and Section 4 offers some concluding remarks. Proofs are given in the Appendix.

1. IDENTIFICATION-ROBUST MD TESTS FOR THE NKPC

We will consider here the structural parameterization of the NKPC derived from Calvo (1983) pricing and indexation, as in Sbordone (2002, 2005) and Christiano, Eichenbaum, and Evans (2005). According to this specification, the parameters λ, γ_f, and γ_b are functions of the degree of price stickiness θ, which is the probability that a firm will be unable to change its price in a given period; the indexation parameter ϱ, which measures the degree of indexation to past inflation; and the discount factor β. As explained in Woodford (2003), in the special case $\beta = 1$, the coefficients γ_f and γ_b in equation (1) add to one. In this case, the model is identical to that of Fuhrer and Moore (1995), and also has identical implications to another popular version of the NKPC due to Galí and Gertler (1999), according to which there is no indexation, but a fraction of agents use a backward-looking rule of thumb to set their prices. Thus, in our empirical analysis we set β equal to one.[4] The model to be estimated is given by the equation

$$\pi_t = \frac{(1-\theta)^2}{\theta\,(1+\varrho)} x_t + \frac{1}{1+\varrho} E_t\,(\pi_{t+1}) + \frac{\varrho}{1+\varrho}\pi_{t-1} + \varepsilon_t. \tag{2}$$

To evaluate the central predictions of this model, we follow the approach of Sbordone (2005), which is based on the distance between the dynamics of inflation implied by the structural model and the dynamics implied by a reduced-form VAR forecasting model. The latter can be written in companion form as

$$\mathbf{z}_t = A\,(\varphi)\,\mathbf{z}_{t-1} + \epsilon_t. \tag{3}$$

Let e_π, e_x denote the unit vectors such that $e_\pi'\mathbf{z}_t = \pi_t$ and $e_x'\mathbf{z}_t = x_t$. Letting k denote the number of variables in the VAR and p the number of lags (order of the VAR), the

3. Any robust confidence interval must have the property that it is unbounded with positive probability, while Wald-based confidence intervals are almost surely bounded.

4. We note that the results are insensitive to alternative values for β near one. Moreover, Kleibergen and Mavroeidis (2009) report evidence that $\gamma_f + \gamma_b$ is very well identified and not significantly different from one.

reduced-form coefficient matrix $A(\varphi)$ is of dimension $kp \times kp$, and it contains $k^2 p$ unknown (reduced-form) parameters, denoted by φ.

Next, we need to link the reduced-form parameters φ to the structural parameters $\vartheta = (\theta, \varrho)'$ in the NKPC (2). We use the standard identifying assumption in the literature that $E_{t-1}\varepsilon_t = 0$ (see, e.g., Galí and Gertler 1999, Sbordone 2002).[5] Taking expectations with respect to information at $t-1$ on both sides of equation (1) yields

$$E_{t-1}\pi_t = \frac{(1-\theta)^2}{\theta(1+\varrho)} E_{t-1}(x_t) + \frac{1}{1+\varrho} E_{t-1}\pi_{t+1} + \frac{\varrho}{1+\varrho}\pi_{t-1}.$$

Substituting for $E_{t-1}\pi_{t+1} = e_\pi' A(\varphi)^2 \mathbf{z}_{t-1}$, $E_{t-1}\pi_t = e_\pi' A(\varphi)\mathbf{z}_{t-1}$, $E_{t-1}x_t = e_x' A(\varphi)\mathbf{z}_{t-1}$, and $\pi_{t-1} = e_\pi' \mathbf{z}_{t-1}$ yields

$$\left\{ e_\pi' \left[I - \frac{1}{1+\varrho}\mathbf{A}(\varphi) \right] - \frac{(1-\theta)^2}{\theta(1+\varrho)} \mathbf{e}_x' \right\} \mathbf{A}(\varphi) - \frac{\varrho}{1+\varrho}\mathbf{e}_\pi' = 0, \tag{4}$$

where I denotes the identity matrix of dimension kp. The kp restrictions (4) will be used to do inference on the structural parameters ϑ based on some estimator of the reduced-form parameters, $\hat{\varphi}$.

The restrictions (4) can be expressed using the distance function

$$g(\varphi, \vartheta) = \mathbf{A}(\varphi)' \left\{ \left[\mathbf{I} - \frac{1}{1+\varrho}\mathbf{A}(\varphi)' \right] \mathbf{e}_\pi - \frac{(1-\theta)^2}{\theta(1+\varrho)} \mathbf{e}_x \right\} - \frac{\varrho}{1+\varrho}\mathbf{e}_\pi. \tag{5}$$

We will also make use of the following Jacobian matrices:

$$G_\varphi(\varphi, \vartheta) \equiv \frac{\partial g(\varphi, \vartheta)}{\partial \varphi'}, \quad \text{and} \quad G_\vartheta(\varphi, \vartheta) \equiv \frac{\partial g(\varphi, \vartheta)}{\partial \vartheta'}. \tag{6}$$

Let $\hat{\varphi}$ denote a consistent and asymptotically normal estimator of the reduced-form parameters, with asymptotic variance matrix V_φ, and let $\hat{V}_\varphi$ be a consistent estimator of V_φ. By the Delta method, the asymptotic variance of $g(\hat{\varphi}, \vartheta)$ is $G_\varphi(\varphi, \vartheta)' V_\varphi G_\varphi(\varphi, \vartheta)$. Efficient MD estimation is based on the criterion function

$$Q(\vartheta) = g(\hat{\varphi}, \vartheta)' \hat{V}_{gg}\left(\bar{\vartheta}\right)^{-1} g(\hat{\varphi}, \vartheta), \tag{7}$$

where $\hat{V}_{gg}(\vartheta) = G_\varphi(\hat{\varphi}, \vartheta)' \hat{V}_\varphi G_\varphi(\hat{\varphi}, \vartheta)$. The value of the parameter $\bar{\vartheta}$ that appears in the weight matrix $\hat{V}_{gg}(\bar{\vartheta})^{-1}$ may be some preliminary (inefficient) estimator of ϑ, for example, an MD estimator using the identity weight matrix, in which case the procedure corresponds to classical two-step MD estimation (see Newey and McFadden 1994). However, $\bar{\vartheta}$ can also be equal to ϑ, in which case the criterion function (7)

5. This can be relaxed to allow for serial correlation in the cost push shock, for example, $E_{t-1}\varepsilon_t = \rho_\varepsilon \varepsilon_{t-1}$, at the cost of introducing additional unknown parameters.

corresponds to the so-called continuously updated (CU) GMM criterion function of Hansen, Heaton, and Yaron (1996).

Standard inference is based on the assumption that the Jacobian matrix $G_\vartheta(\varphi, \vartheta)$ (see equation (6)) is of full-rank. Under this assumption, two-step and CU-MD estimators are asymptotically equivalent, and since the former is computationally simpler, it is the one commonly used. In that case, the minimizer of the MD criterion function, $\hat{\vartheta}$, is asymptotically normal, and the usual Wald statistics for hypotheses on ϑ are asymptotically chi-square distributed (see Newey and McFadden 1994). However, when $G_\vartheta(\varphi, \vartheta)$ is nearly of reduced rank, these conventional asymptotic approximations break down, and inference based on Wald or conventional score tests can be seriously misleading.

To get some intuition for why this happens, consider first the conventional score statistic for the null hypothesis H_0: $\vartheta = \vartheta_0$. This statistic is a quadratic form of the score vector under H_0, $G_\vartheta(\hat{\varphi}, \vartheta_0)' \hat{V}_{gg}(\vartheta_0)^{-1} \sqrt{T}\, g(\hat{\varphi}, \vartheta_0)$, with respect to the inverse of its variance, $G_\vartheta(\hat{\varphi}, \vartheta_0)' \hat{V}_{gg}(\vartheta_0)^{-1} G_\vartheta(\hat{\varphi}, \vartheta_0)$, which can be written as

$$\mathrm{LM}(\vartheta_0) = T\, g(\hat{\varphi}, \vartheta_0)' \hat{V}_{gg}(\vartheta_0)^{-\frac{1}{2}} P_{\hat{V}_{gg}(\vartheta_0)^{-\frac{1}{2}} G_\vartheta(\hat{\varphi}, \vartheta_0)} \hat{V}_{gg}(\vartheta_0)^{-\frac{1}{2}} g(\hat{\varphi}, \vartheta_0), \tag{8}$$

where $P_X = X(X'X)^{-1}X'$ for any full-rank matrix X, and $V^{-1/2}$ denotes the symmetric square root inverse of an invertible matrix V. Since $\hat{V}_{gg}(\vartheta_0)$ is consistent for $V_{gg}(\vartheta_0)$, the asymptotic distribution of the score statistic depends mainly on the behavior of the terms $\sqrt{T}\, g(\hat{\varphi}, \vartheta_0)$ and $G_\vartheta(\hat{\varphi}, \vartheta_0)$. By the Delta method, the former is approximately normal if $\hat{\varphi}$ is asymptotically normal, so let us denote its limiting expression by a vector ξ_g that is normally distributed. The problematic term is $G_\vartheta(\hat{\varphi}, \vartheta_0)$. By a mean-value expansion of $G_\vartheta(\hat{\varphi}, \vartheta_0)$ around the true value φ, it can be shown that $G_\vartheta(\hat{\varphi}, \vartheta_0)$ is approximately equal to $G_\vartheta(\varphi, \vartheta_0) + T^{-1/2}\xi_G$, where ξ_G is a normally distributed random matrix. When the matrix $G_\vartheta(\varphi, \vartheta_0)$ is of full rank, it is the dominant term in the above expression, so $G_\vartheta(\hat{\varphi}, \vartheta_0)$ is approximately constant, and $G_\vartheta(\hat{\varphi}, \vartheta_0)' \hat{V}_{gg}(\vartheta_0)^{-1} G_\vartheta(\hat{\varphi}, \vartheta_0)$ converges to a nonrandom full-rank matrix. As a result, the score test is chi-square distributed.

However, the quality of this approximation clearly depends on how big the matrix $G_\vartheta(\varphi, \vartheta_0)$ is relative to the higher order random term $T^{-1/2}\xi_G$. To get a better approximation to the distribution of the statistic when $G_\vartheta(\varphi, \vartheta)$ is potentially small, we can adopt the device of Staiger and Stock (1997) and let $G_\vartheta(\varphi, \vartheta_0) = T^{-1/2}G$ for some fixed matrix G. With this assumption, the score statistic, in large samples, becomes approximately equal to

$$\xi_g' V_{gg}^{-1} (\xi_G + G) \left[(\xi_G + G)' V_{gg}^{-1} (\xi_G + G) \right]^{-1} (\xi_G + G) V_{gg}^{-1} \xi_g. \tag{9}$$

The distribution of the above statistic is different from a chi-square distribution. The key reason for this is that the random variable ξ_G is correlated with ξ_g. So, use of chi-square critical values will produce unreliable inference. Moreover, the parameter G that affects the shape of the distribution of the score statistic cannot be estimated consistently, so we cannot correct the critical values using some estimate of G.

The Wald statistic is a quadratic form of the MD estimator $\hat{\vartheta}$ with respect to the inverse of its variance $\hat{V}_{\vartheta\vartheta}$ where $\hat{V}_{\vartheta\vartheta} = T^{-1}[G_{\vartheta}(\hat{\varphi}, \hat{\vartheta})'\hat{V}_{gg}(\bar{\vartheta})^{-1}G_{\vartheta}(\hat{\varphi}, \hat{\vartheta})]^{-1}$. This statistic is not chi-square distributed in large samples, either, when $G_{\vartheta}(\varphi, \vartheta)$ is close to a reduced-rank matrix because $\hat{\vartheta}$ is not approximately normal and $\hat{V}_{\vartheta\vartheta}$ does not converge to a nonstochastic matrix. See Stock et al. (2002) for further details.

To avoid spurious inference, we consider here test statistics whose distribution under the null hypothesis is independent of the value of the Jacobian matrix $G_{\vartheta}(\varphi, \vartheta)$. Such statistics yield inference that is fully robust to weak identification or weak instruments problems. The difference of these statistics to their GMM counterparts will be highlighted in the next section.

The first test statistic we consider is the MD version of the statistic of Anderson and Rubin (1949), which is given by the CU-MD objective function, equation (7) with $\bar{\vartheta} = \vartheta$, scaled by the sample size

$$\text{MD-AR}(\vartheta_0) = T\, g\,(\hat{\varphi}, \vartheta_0)'\, \hat{V}_{gg}\,(\vartheta_0)^{-1}\, g\,(\hat{\varphi}, \vartheta_0)\,. \tag{10}$$

Like the original Anderson-Rubin statistic, AR_{MD} can be interpreted as a Wald test of the validity of the restrictions implied by the model (see equation (4)) at a given hypothesized value of the parameters (θ_0, ϱ_0). This statistic is robust to any value of the Jacobian matrix $G_{\vartheta}(\varphi, \vartheta)$ since it does not involve $G_{\vartheta}(\varphi, \vartheta)$ at all. Details of the argument are given in the proof of Proposition 1 in the Appendix.

The second statistic is a score statistic that is based on the derivative of the CU-MD objective function. This statistic is the MD version of the K-statistic proposed by Kleibergen (2002, 2005). It can be shown that

$$\frac{1}{2}\frac{\partial Q(\vartheta)}{\partial \vartheta'} = g\,(\hat{\varphi}, \vartheta)'\, \hat{V}_{gg}\,(\vartheta)^{-1}\, \hat{D}\,(\vartheta)\,, \tag{11}$$

where

$$\hat{D}\,(\vartheta) = \left[\frac{\partial g\,(\hat{\varphi}, \theta, \varrho)}{\partial \theta} - \hat{V}_{\theta g}(\vartheta)\hat{V}_{gg}\,(\vartheta)^{-1}\, g\,(\hat{\varphi}, \vartheta)\,,\right.$$
$$\left.\frac{\partial g\,(\hat{\varphi}, \theta, \varrho)}{\partial \varrho} - \hat{V}_{\varrho g}(\vartheta)\hat{V}_{gg}\,(\vartheta)^{-1}\, g\,(\hat{\varphi}, \vartheta)\right]. \tag{12}$$

$\hat{V}_{\theta g}(\vartheta) = \frac{\partial G_{\varphi}(\hat{\varphi}, \vartheta)}{\partial \theta}\hat{V}_{\varphi}G_{\varphi}(\hat{\varphi}, \vartheta)$ and $\hat{V}_{\varrho g}(\vartheta) = \frac{\partial G_{\varphi}(\hat{\varphi}, \vartheta)}{\partial \varrho}\hat{V}_{\varphi}G_{\varphi}(\hat{\varphi}, \vartheta)$. The MD-K statistic is given by a quadratic form of the derivative of the CU-MD objective function (11) with respect to an estimator of its variance, that is,

$$\text{MD-K}(\vartheta_0) = T g(\hat{\varphi}, \vartheta_0)'\hat{V}_{gg}(\vartheta_0)^{-\frac{1}{2}} P_{\hat{V}_{gg}(\vartheta_0)^{-\frac{1}{2}}\hat{D}(\vartheta_0)}\hat{V}_{gg}(\vartheta_0)^{-\frac{1}{2}} g(\hat{\varphi}, \vartheta_0). \tag{13}$$

By comparison to the conventional score statistic given in equation (8), we observe that the MD-K statistic uses the matrix $\hat{D}(\vartheta)$ instead of $G_{\varphi}(\hat{\varphi}, \vartheta)$. This difference is actually what gives the statistic its robustness property. The matrix $\hat{D}(\vartheta)$ is an

alternative estimator of the Jacobian $G_\vartheta(\varphi, \vartheta)$, which, unlike $G_\varphi(\hat{\varphi}, \vartheta)$, is asymptotically independent of $g(\hat{\varphi}, \vartheta)$. This means that we can derive the distribution of the resulting score statistic *conditional* on $\hat{D}(\vartheta)$, that is, treating $\hat{D}(\vartheta)$ as fixed. The asymptotic distribution of MD-K is also given by the expression (9), but with the random matrix ξ_G now being independent of ξ_g. It can be shown that, conditional on ξ_G, this distribution is chi-square, and since it is independent of ξ_G, this is also the unconditional distribution. See the proof of Proposition 1 in the Appendix for details. Therefore, the key insight, originally discovered by Kleibergen (2002), for getting an identification-robust score test is to use an estimator of the Jacobian that is asymptotically independent from the distance function.

Note that, like in the case of GMM, because the MD-K statistic depends on the derivative of the objective function, tests based on it may suffer a spurious decline of power at inflection points of the objective function, where the identifying restrictions are typically violated. Thus, we will follow the approach of Kleibergen (2005) and combine the MD-K statistic with a statistic that tests the validity of the overidentifying restrictions under H_0

$$\text{MD-J}\,(\vartheta_0) = \text{MD-AR}\,(\vartheta_0) - \text{MD-K}\,(\vartheta_0)\,.$$

The joint test will be referred to as MD-KJ test, and it will reject the null hypothesis at the $\alpha\%$ significance level whenever the hypothesis is rejected either by an $\alpha_1\%$ level MD-K test or an $\alpha_2\%$ MD-J test, where $\alpha_1 + \alpha_2 = \alpha$. In our empirical analysis below, we use $\alpha_1 = 0.8\alpha$ and $\alpha_2 = 0.2\alpha$.

The following result gives the asymptotic distribution of these statistics under the null hypothesis H_0: $\vartheta = \vartheta_0$.

PROPOSITION 1. *Assume that* $\sqrt{T}(\hat{\varphi} - \varphi) \xrightarrow{d} N(0, V_\varphi)$, $\hat{V}_\varphi \xrightarrow{p} V_\varphi$ *and the matrix* $G_\varphi(\varphi, \vartheta_0)V_\varphi G_\varphi(\varphi, \vartheta_0)'$ *is of full-rank* $k_g = kp$. *Then,*

$$\begin{aligned} \text{MD-AR}(\vartheta_0) &\xrightarrow{d} \chi^2(k_g) \\ \text{MD-K}(\vartheta_0) &\xrightarrow{d} \chi^2\,(2) \\ \text{MD-J}\,(\vartheta_0) &\xrightarrow{d} \chi^2(k_g - 2), \end{aligned} \tag{14}$$

where "$\xrightarrow{d}$*" and "*$\xrightarrow{p}$*" indicate convergence in distribution and in probability, respectively, and* $\chi^2(\kappa)$*indicates a chi square distribution with* κ *degrees of freedom.*

The proof of the proposition is given in the Appendix. Note that the assumptions we use are strictly weaker than the assumptions used to establish that classical MD t statistics are asymptotically normal. The latter requires in addition the identification condition that the Jacobian matrix $G_\vartheta(\varphi, \vartheta)$ should be of full rank (see Newey and McFadden 1994). The assumptions in Proposition 1 do rule out some interesting cases, such as unit roots in the VAR, but this limitation applies also to the nonrobust methods (see Moon and Schorfheide 2002). The development of methods that are

robust both to weak instruments as well as near unit roots in the VAR coefficients is a topic for future research.

Identification-robust $(1-\alpha)$ level confidence regions for the parameters (θ, ϱ) can be obtained by inverting the MD-AR and MD-KJ tests, that is, by collecting all the values of (θ_0, ϱ_0) that are not rejected by the tests at the α level of significance.

2. COMPARISON WITH GMM

For clarity, we discuss a simple special case of the NKPC with only one parameter

$$\pi_t = E_t \pi_{t+1} + \lambda x_t + \varepsilon_t. \tag{15}$$

The reduced-form model for (π_t, x_t) is given by

$$\begin{pmatrix} \pi_t \\ x_t \end{pmatrix} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} x_{t-1} + \begin{pmatrix} \eta_t \\ v_t \end{pmatrix}. \tag{16}$$

In this setting, the only relevant instrument is x_{t-1}, so the model is just-identified.

The distance function (5) is given by

$$g(\hat{\varphi}, \lambda) = \hat{\varphi}_1 (1 - \hat{\varphi}_2) - \lambda \hat{\varphi}_2,$$

where $\hat{\varphi}$ is the OLS estimator of φ. For GMM, we use the moment condition $E[f_t(\lambda)] = 0$, where $f_t(\lambda) = (\pi_t - \pi_{t+1} - \lambda x_t) x_{t-1}$, with corresponding sample moments

$$f_T(\lambda) = \frac{1}{T} \sum_{t=1}^{T} f_t(\lambda).$$

The GMM-based Anderson Rubin statistic for testing the null hypothesis $H_0: \lambda = \lambda_0$, is

$$\text{GMM-AR}(\lambda_0) = T \; f_T(\lambda_0)' \, \hat{V}_{ff}(\lambda_0)^{-1} \; f_T(\lambda_0),$$

where $\hat{V}_{ff}(\lambda_0)$ is an estimator of the asymptotic variance of $\sqrt{T} f_T(\lambda_0)$. Because $f_t(\lambda)$ is serially correlated, $\hat{V}_{ff}(\lambda_0)$ must be a heteroskedasticity- and autocorrelation-consistent (HAC) estimator. This is different from MD, where it suffices to use only a heteroskedasticity-consistent estimator of the variance of the reduced-form parameters, in view of the assumption that the proposed reduced-form VAR model represents the dynamics of (π_t, x_t) adequately. Though this difference between MD and GMM is of no consequence asymptotically, it could be important in finite samples, as HAC estimators are known to cause size distortions in finite samples (see, e.g., Sun, Phillips, and Jin 2008).

The main difference that we wish to emphasize is in the way in which the above two approaches make use of the identifying restriction $E_{t-1}\varepsilon_t = 0$. Specifically, the difference arises due to the presence of the forward-looking term $E_{t-1}\pi_{t+1}$: GMM proxies $E_{t-1}\pi_{t+1}$ by projecting π_{t+1} directly on the instrument x_{t-1}, while MD uses the iterative expectation based on the reduced-form model (16), namely, $E_{t-1}\pi_{t+1} = \varphi_1\varphi_2 x_{t-1}$. This makes it clear that GMM is more general, since the direct projection of π_{t+1} on the instruments will remain valid (although inefficient) when the reduced-form model is misspecified, while the MD iterative forecast will be biased when the reduced-form model is misspecified. The distinction is akin to the issue of direct versus iterative multistep forecasts that have been studied in the forecasting literature (see Marcellino, Stock, and Watson 2006). Iterative forecasts are more efficient, but direct forecasts are more robust to misspecification of the forecasting model.

From the above discussion, we expect that when the additional structure imposed by the MD approach is correct, the resulting estimates of $E_{t-1}\pi_{t+1}$ will be more accurate, and this will lead to more efficient inference on the structural parameter λ. To get a sense of the extent of efficiency gains one might expect, we report the power curves for the MD and GMM tests of the hypothesis H_0: $\lambda = \lambda_0$ in equation (15). The power curves are computed by simulation based on the reduced-form model given by equation (16), where $\eta_t = \varepsilon_t + (\lambda + \varphi_1)v_t$, and ε_t, v_t are drawn from a zero-mean bivariate Normal distribution with unit variances and correlation $\rho_{\varepsilon v} = 0.2$.

The strength of identification of λ can be measured by a quantity known as the concentration parameter μ^2 (see Stock, Wright, and Yogo 2002). The concentration parameter was originally defined in the linear instrumental variables (IV) regression model as a unitless measure of the amount of information in the data about the structural parameters, and it plays a role similar to the sample size in asymptotic analysis (see Stock, Wright, and Yogo 2002). For example, the rate of convergence of the IV estimator to the true value is μ. In the linear IV model with a single endogenous regressor, the concentration parameter can be defined in terms of the population R^2 of the so-called first-stage regression of the endogenous regressor on the instruments: $\mu^2 = T R^2/(1 - R^2)$, where T is the sample size. The intuition is that when instruments explain little of the variation in the endogenous regressor, the effective sample size is small, even if the actual sample size is large.

In the present example, the first-stage regression is the first-order autoregression $x_t = \varphi_2 x_{t-1} + v_t$, and its population R^2 is φ_2^2. Therefore, the concentration parameter is $\mu^2 = T\varphi_2^2/(1 - \varphi_2^2)$, so the strength of identification depends only on φ_2.

We use a sample size of 200 observations and we consider two different values for the concentration parameter: $\mu^2 = 4$ and $\mu^2 = 100$, corresponding to weak and strong identification, respectively. The empirical moments for GMM are $f_t(\lambda) = Z_t'\varepsilon_t$, with $Z_t = [x_{t-1}, \pi_{t-1}]$ and $Z_t = [x_{t-1}, \pi_{t-1}, x_{t-2}, \pi_{t-2}]$. This corresponds to VAR(1) and VAR(2) reduced-form models for MD. For the GMM test, we use the Newey and West (1987) estimator for the variance, while the MD tests are based on the OLS

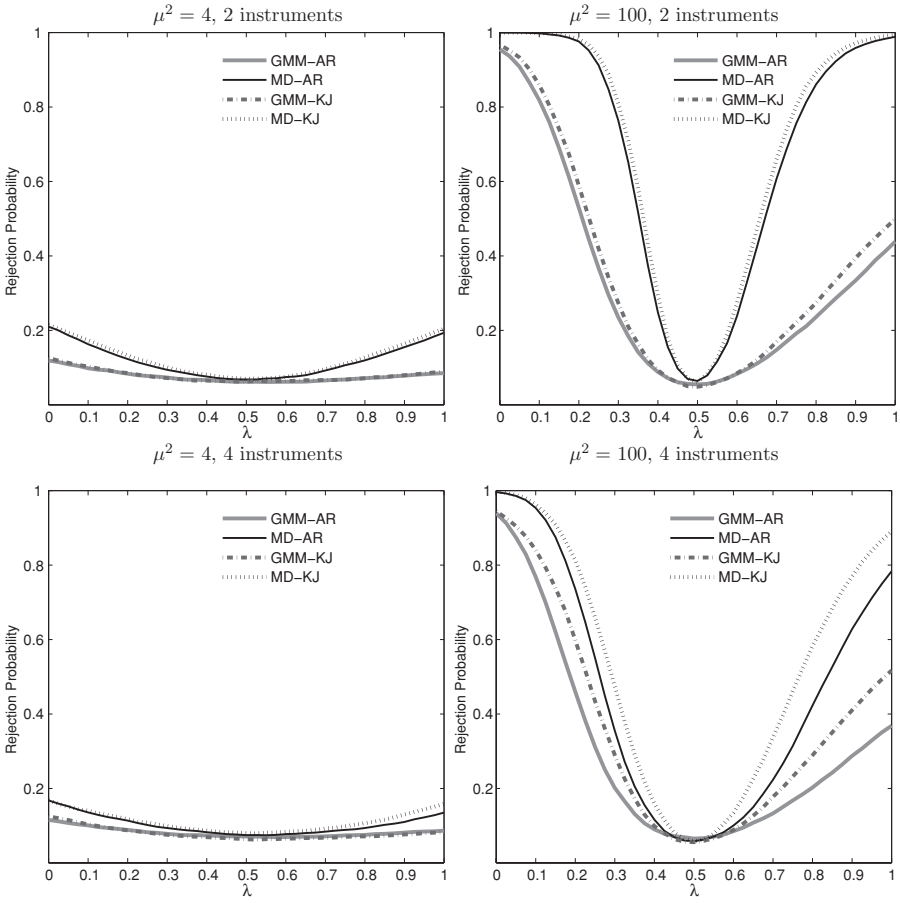

FIG. 1. Rejection Probability Using 5% Significant Level for Testing H_0: $\lambda = 0.5$ Against H_1: $\lambda \neq 0.5$, and $\rho_{\varepsilon v} = 0.2$.

NOTE: The sample size is 200 and the number of Monte Carlo replications is 10,000.

estimator for the reduced-form VAR coefficients and the White (1980) estimator of their variance. The results are reported in Figure 1.[6]

We see that in both cases of weak and strong instruments, the MD-AR and MD-KJ tests are substantially more powerful than their GMM counterparts. When the degree of overidentification is one, the MD-AR test is almost as powerful as the MD-KJ test as expected, but the difference increases as the degree of overidentification gets higher. Also, the MD tests based on a VAR(1) reduced form are more powerful than those based on a VAR(2), since the former exploits the additional restriction that the second-order lags are irrelevant, which is consistent with the DGP (16). So the power of the tests is increasing in the number of correctly specified restrictions.

6. Since λ is nonnegative, we choose $\lambda_0 = 1/2$ in order to produce equally tailed two-sided power curves. This does not affect the results, nor is it meant to be representative of estimates reported in the literature.

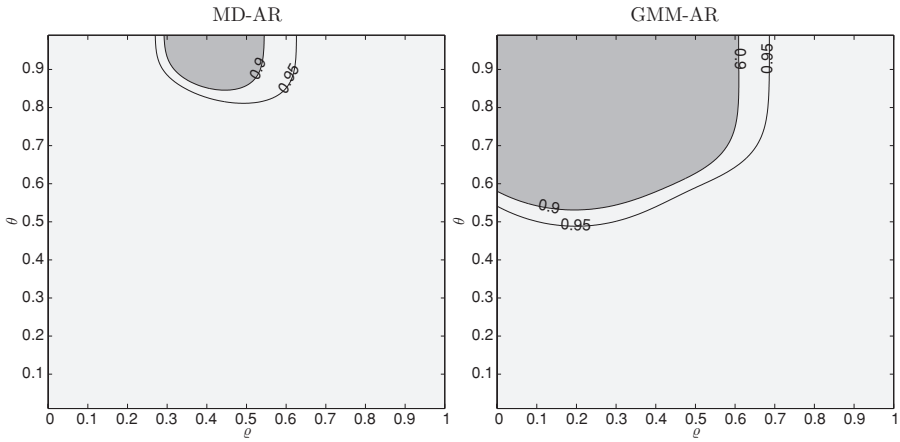

FIG. 2. MD-AR and GMM-AR Confidence Sets for the Parameters (θ, ϱ) of the NKPC.

NOTE: The dark and light gray areas are, respectively, the 90% and 95% confidence sets. The sample is 1984q1–2008q3.

3. EMPIRICAL RESULTS

We use quarterly data on inflation and the labor share over the period 1984 to 2008. Inflation is calculated from the quarterly GDP deflator, while the labor share is obtained from the Bureau of Labor Statistics and transformed according to procedure used in Sbordone (2002) that was also used by Galí and Gertler (1999).

We focus on the post-1983 sample so as to avoid issues of instability in the dynamics of inflation and the labor share induced by, among other things, changes in monetary policy in the late 1970s and early 1980s. This is in line with the recent literature (see, e.g., Krause, Lopez-Salido, and Lubik 2008) and is also further motivated by the stability tests reported in Kleibergen and Mavroeidis (2009), which indicate instabilities before 1984.

Confidence sets for the price stickiness and indexation parameters (θ, ϱ) are computed by grid search within the parameter space $\theta, \varrho \in [0, 1]$. Figures 2 and 3 report 95% and 90% level confidence sets based on inverting the AR tests and the KJ combination tests defined in Section 1 and contrast them with their GMM counterparts. The MD tests are based on a VAR(3) for inflation and the labor share, and GMM accordingly uses three lags of both of these variables as instruments, following Kleibergen and Mavroeidis (2009). The efficient weight matrices for MD and GMM are based on the White (1980) and Newey and West (1987) estimators, respectively, as discussed in Section 2.

The following conclusions emerge from these pictures. On the methodological side, we see that the confidence sets based on MD-AR test is less than half the size of their GMM counterpart, and a similar difference appears for the KJ sets. Thus, there are substantial power gains that arise from imposing the

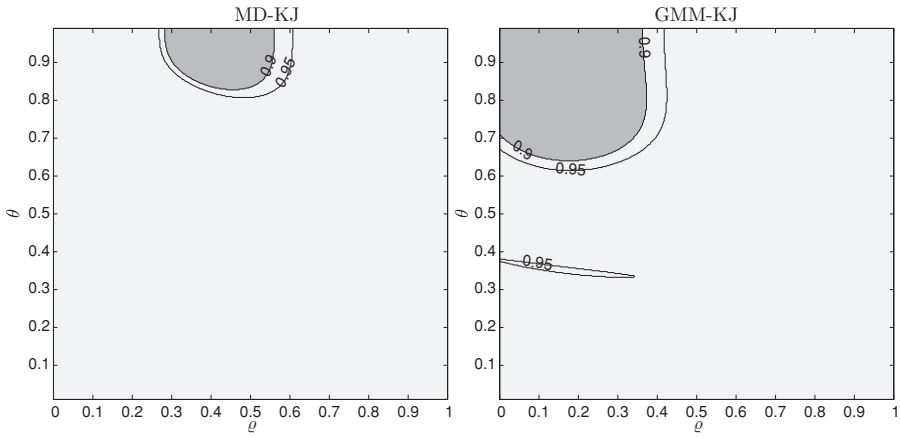

FIG. 3. MD-KJ and GMM-KJ Confidence Sets for the Parameters (θ, ϱ) of the NKPC.

NOTE: The dark and light gray areas are, respectively, the 90% and 95% confidence sets. The sample is 1984q1–2008q3.

full-information restrictions, which are consistent with the simulations reported in Section 2.

On the empirical side, we can infer from both the MD-AR and MD-KJ 90% sets that the indexation parameter lies roughly between 0.3 and 0.6. This indicates that indexation is significant, thus rejecting the pure forward-looking version of the NKPC, but that it is also less than 100%, as in the model of Christiano, Eichenbaum, and Evans (2005). With regard to the parameter θ that measures the degree of price stickiness, we also find that it is more precisely estimated by MD than by limited-information GMM, with the confidence intervals spanning the region 4/5 to 1. However, in terms of the average duration of prices, that is, the average time over which prices remain fixed, which is given by $1/(1-\theta)$, we see that the confidence sets remain unbounded from above, as in the GMM case. One implication of the fact that $\theta = 1$ is in the confidence set is that the slope of the NKPC is not significantly different from zero. This is also found in the limited-information studies cited above and could be interpreted either as evidence that the labor share is not the relevant variable that drives inflation or as evidence that the average duration of prices cannot be well identified by aggregate macroeconomic data.

4. CONCLUSION

This paper was motivated by the findings in the literature that the NKPC is not well identified. This finding was based on the use of limited-information methods, which make minimal assumptions about the dynamics of the labor share. We asked whether imposing the assumption that the dynamics of inflation and the labor share can be

represented by a finite-order VAR model improves the identification of the NKPC parameters. We proposed two different methods for answering this question. These are based on the MD approach used by Sbordone (2002, 2005) for the NKPC, but they do not require any assumption about identification, and therefore results derived from them are reliable even when identification fails. These methods are adaptations of the tests developed for GMM by Stock and Wright (2000) and Kleibergen (2005), which were recently developed in the context of limited dependent variable models by Magnusson (Forthcoming).

We found that the full-information approach reduces the uncertainty associated with one of the two key structural parameters of the NKPC. Specifically, we find evidence of partial but not full indexation of prices to past inflation. Somewhat sharper inference relative to GMM is obtained also for the parameter governing the average duration of prices, but the confidence intervals on this parameter remain unbounded from above, indicating that it remains weakly identified.

On the methodological side, the main message of this paper is that MD estimation is also subject to the problem of weak identification for the NKPC. On the other hand, the results of this paper suggest there are substantial efficiency gains in estimating forward-looking models by full- versus limited-information methods when the possibility of weak identification is taken into account. The methodology used here can be extended to estimate multiple-equation models, such as the DSGE model of Christiano, Eichenbaum, and Evans (2005). This is an important topic for future research, as DSGE models are currently estimated using procedures that are not robust to possible identification failure.

APPENDIX

PROOF OF PROPOSITION 1. The asymptotic distribution of MD-AR (ϑ_0) follows by applying the Delta method. For the distribution of the MD-K and MD-J statistics, it suffices to show that the matrix $\hat{D}(\vartheta)$ defined in equation (12) is asymptotically independent of $g(\hat{\varphi}, \vartheta)$. First, note that the $kp \times kp$ matrix $A(\varphi)$ in the reduced-form VAR (16) can be written as $A(\varphi) = B\Phi + C$ where $B = (I_k, 0, \ldots, 0)'$ is $kp \times k$ matrix, $\Phi = (\Phi_1, \ldots, \Phi_p)$ is $k \times kp$ matrix of the VAR(p) coefficients, and $C = (0, 0; I_{kp-1}, 0)$ is $kp \times kp$ and $\varphi = \text{vec}(\Phi)$. So $A(\varphi)$ is linear in φ, and the distance function (5) is differentiable and its Jacobian w.r.t. φ is given by

$$G_\varphi(\varphi, \vartheta) = I_{kp} \otimes \left(e_\pi - \frac{(1-\theta)^2}{\theta(1+\varrho)} e_x - \frac{1}{1+\varrho} A(\varphi) e_\pi \right)' B - \frac{1}{1+\varrho} [A(\varphi)' \otimes e_\pi' B].$$

Clearly, $\frac{\partial G_\varphi(\hat{\varphi}, \vartheta)}{\partial \theta}$ and $\frac{\partial G_\varphi(\hat{\varphi}, \vartheta)}{\partial \varrho}$ exist and are linear in $\hat{\varphi}$. Hence, by a first-order Taylor expansion around φ we obtain: $\hat{\xi}_g \equiv \sqrt{T} g(\hat{\varphi}, \vartheta) = G_\varphi(\varphi, \vartheta)\xi_\varphi +$

$o_p(1)$, $\hat{\xi}_\theta \equiv \sqrt{T}[\frac{\partial g(\hat{\varphi},\vartheta)}{\partial \theta} - \frac{\partial g(\varphi,\vartheta)}{\partial \theta}] = \frac{\partial G_\varphi(\varphi,\vartheta)}{\partial \theta}\xi_\varphi + o_p(1)$, and $\hat{\xi}_\varrho \equiv \sqrt{T}[\frac{\partial g(\hat{\varphi},\vartheta)}{\partial \varrho} - \frac{\partial g(\varphi,\vartheta)}{\partial \varrho}] = \frac{\partial G_\varphi(\varphi,\vartheta)}{\partial \varrho}\xi_\varphi + o_p(1)$, where $\xi_\varphi \sim N(0, V_\varphi)$. Thus, $(\hat{\xi}_g \vdots \hat{\xi}_\theta \vdots \hat{\xi}_\varrho) \xrightarrow{d} (\xi_g \vdots \xi_\theta \vdots \xi_\varrho)$, where ξ_g, ξ_θ, and ξ_ϱ are kp-dimensional jointly normally distributed random vectors with covariances $E(\xi_\theta \xi_g') \equiv V_{\theta g}(\vartheta) = \frac{\partial G_\varphi(\varphi,\vartheta)}{\partial \theta} V_\varphi G_\varphi(\varphi,\vartheta)'$ and $E(\xi_\varrho \xi_g') \equiv V_{\varrho g}(\vartheta) = \frac{\partial G_\varphi(\varphi,\vartheta)}{\partial \varrho} V_\varphi G_\varphi(\varphi,\vartheta)'$. $\hat{V}_{\theta g}(\vartheta)$ and $\hat{V}_{\varrho g}(\vartheta)$ defined in the text are consistent for $V_{\varrho g}(\vartheta)$ and $V_{\varrho g}(\vartheta)$, respectively, by Slutsky's theorem, since $V_{\varrho g}(\vartheta)$ and $V_{\varrho g}(\vartheta)$ are continuous in φ.

Let the $kp \times 2$ matrix $\hat{D}(\vartheta)$ defined in the text be written as $\hat{D}(\vartheta) = [\hat{D}_\theta(\vartheta), \hat{D}_\varrho(\vartheta)]$. Under strong identification, $\hat{D}(\vartheta) \xrightarrow{p} G_\vartheta(\varphi,\vartheta)$ has full rank. So, $\hat{D}(\vartheta)$ is trivially independent of $g(\hat{\varphi},\vartheta)$. Under weak identification, we can use the nesting $G_\vartheta(\varphi,\vartheta) = T^{-1/2}G$, where $G = [G_\theta \vdots G_\varrho]$ is a nonrandom matrix. This implies that $\sqrt{T}\hat{D}_\theta(\vartheta) = \xi_{\theta.g} + o_p(1)$, where $\xi_{\theta.g} = \xi_\theta - V_{\theta g}(\vartheta)V_{gg}(\vartheta)^{-1}\xi_g + G_\theta$, and $\xi_{\theta.g}$ is independent of ξ_g, and similarly for $\hat{D}_\varrho(\vartheta)$. Thus, $\sqrt{T}\hat{D}(\vartheta)$ is asymptotically independent of $\sqrt{T}\,g(\hat{\varphi},\vartheta)$. Hence, in both cases, $[\hat{D}(\vartheta)'\hat{V}_{gg}(\vartheta)^{-1}\hat{D}(\vartheta)]^{-1/2}\hat{D}(\vartheta)'\hat{V}_{gg}(\vartheta)^{-1}\sqrt{T}\,g(\hat{\varphi},\vartheta)$ is approximately normal with identity variance matrix. The asymptotic distribution of MD-K and MD-J follows by the continuous mapping theorem. □

LITERATURE CITED

Anderson, Theodore W., and Herman Rubin. (1949) "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Annals Mathematical Statistics,* 20, 46–63.

Calvo, Guillermo A. (1983) "Staggered Prices in a Utility Maximizing Framework." *Journal of Monetary Economics,* 12, 383–98.

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans. (2005) "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy,* 113, 1–45.

Dufour, Jean-Marie. (1997) "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models." *Econometrica,* 65, 1365–87.

Dufour, Jean-Marie, Lynda Khalaf, and Maral Kichian. (2006) "Inflation Dynamics and the New Keynesian Phillips Curve: An Identification Robust Econometric Analysis." *Journal of Economic Dynamics and Control,* 30, 1707–27.

Dufour, Jean-Marie, Lynda Khalaf, and Maral Kichian. (2007) "Structural Multi-Equation Macroeconomic Models: A System-Based Estimation and Evaluation Approach." Discussion paper, Bank of Canada.

Dufour, Jean-Marie, Lynda Khalaf, and Maral Kichian. (2008) "On the Precision of the Calvo Parameter in Structural NKPC Models." Technical report.

Fuhrer, Jeffrey C., and George R. Moore. (1995) "Inflation Persistence." *Quarterly Journal of Economics,* 110, 127–59.

Fuhrer, Jeffrey C., and Giovanni P. Olivei. (2004) "Estimating Forward-Looking Euler Equations with GMM Estimators: An Optimal Instruments Approach." Working Paper No. 04-2, Federal Reserve Bank of Boston.

Galí, Jordi, and Mark Gertler. (1999) "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics,* 44, 195–222.

Hansen, Lars P., John Heaton, and Amir Yaron. (1996) "Finite Sample Properties of Some Alternative GMM Estimators." *Journal of Business and Economic Statistics,* 14, 262–80.

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

Kleibergen, Frank. (2005) "Testing Parameters in GMM without Assuming that They Are Identified." *Econometrica,* 73, 1103–23.

Kleibergen, Frank and Sophocles Mavroeidis. (2009) "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve." *Journal of Business and Economic Statistics,* 27, 293–339.

Krause, Michael U., David López-Salido, and Thomas A. Lubik. (2008) "Inflation Dynamics with Search Frictions: A Structural Econometric Analysis." *Journal of Monetary Economics,* 55, 892–916.

Kurmann, André. (2005) "Quantifying the Uncertainty About the Fit of a New Keynesian Pricing Model." *Journal of Monetary Economics,* 52, 1119–34.

Magnusson, Leandro M. (Forthcoming) "Inference in Limited Dependent Variable Models Robust to Weak Identification." *Econometrics Journal*.

Marcellino, Massimiliano, James H. Stock, and Mark W. Watson. (2006) "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series." *Journal of Econometrics,* 135, 499–526.

Moon, Hyungsik R., and Frank Schorfheide. (2002) "Minimum Distance Estimation of Nonstationary Time Series Models." *Econometric Theory,* 18, 1385–1407.

Moreira, Marcelo J. (2003) "A Conditional Likelihood Ratio Test for Structural Models." *Econometrica,* 71, 1027–48.

Newey, Whitney K., and Daniel McFadden (1994) Large Sample Estimation and Hypothesis Testing. In *The Handbook of Econometrics*, edited by Robert F. Engle and Daniel McFadden, Volume 4, pp. 2111–245. Amsterdam: North-Holland.

Newey, Whitney K., and Kenneth D. West. (1987) "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica,* 55, 703–8.

Sbordone, Argia M. (2002) "Prices and Unit Labor Costs: A New Test of Price Stickiness." *Journal of Monetary Economics,* 49, 265–92.

Sbordone, Argia M. (2005) "Do Expected Future Marginal Costs Drive Inflation Dynamics?" *Journal of Monetary Economics,* 52, 1183–97.

Schorfheide, Frank. (2005) "VAR Forecasting under Misspecification." *Journal of Econometrics,* 128, 99–136.

Staiger, Douglas, and James H. Stock. (1997) "Instrumental Variables Regression with Weak Instruments." *Econometrica,* 65, 557–86.

Stock, James H., and Jonathan H. Wright. (2000) "GMM with Weak Identification." *Econometrica,* 68, 1055–96.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. (2002) "GMM, Weak Instruments, and Weak Identification." *Journal of Business and Economic Statistics,* 20, 518–30.

Sun, Yixiao, Peter C. B. Phillips, and Sainan Jin. (2008) "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing." *Econometrica,* 76, 175–94.

White, Halbert. (1980) "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica,* 48, 817–38.

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