Comment

John C. Chao
Department of Economics, University of Maryland, College Park, MD 20742 (*Chao@Econ.umd.edu*)

Norman R. Swanson
Department of Economics, Rutgers University, New Brunswick, NJ 08901 (*nswanson@econ.rutgers.edu*)

In their compelling and noteworthy new piece, Professors Kleibergen and Mavroeidis have given us a lucid exposition of some of the test procedures that have been introduced in recent years to deal with the problem of weak instruments. They have also applied these methods to an empirical analysis of the New Keyesian Phillips Curve (NKPC) model of inflation dynamics. Their study of the NKPC model is particularly timely in light of the fact that recently there has been a surge of interest in this model in macroeconomics. We would, thus, like to congratulate them on a most stimulating and thought-provoking article.

We will focus our discussion on two of the many interesting issues discussed in their paper. Namely: (i) the notion that the NKPC model is weakly identified if marginal costs fail to explain inflation in the structural equation, and (ii) the literature on many instrument and many weak instrument asymptotics.

(i) Kleibergen and Mavroeidis (KM, henceforth) consider estimating the following structural equation:

$$\pi_t = \lambda x_t + \gamma_f \pi_{t+1} + \gamma_b \pi_{t-1} + e_t, \tag{1}$$

where π_t denotes inflation, x_t is some measure of marginal costs, and, under the formulation of the model, x_t and π_{t+1} are taken to be endogenous variables correlated with the error, e_t. In analyzing the identification of this model, KM focus on the special case of a purely forward-looking model with $\gamma_b = 0$, and also specify an AR(2) process for the generating mechanism of x_t, that is, the system:

$$\pi_t = \lambda x_t + \gamma_f \pi_{t+1} + e_t, \tag{2}$$

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \upsilon_t. \tag{3}$$

They show that, under this setup, the mean function of the first-stage equation is given by:

$$\begin{pmatrix} E_{t-1}(\pi_{t+1}) \\ E_{t-1}(x_t) \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_0((\rho_1 + \rho_2\gamma_f)\rho_1 + \rho_2) & \alpha_0(\rho_1 + \rho_2\gamma_f)\rho_2 \\ \rho_1 & \rho_2 \end{pmatrix}}_{\Pi} \times \underbrace{\begin{pmatrix} x_{t-1} \\ x_{t-2} \end{pmatrix}}_{Z_t}, \tag{4}$$

where $\alpha_0 = \lambda/[1-\gamma_f(\rho_1+\gamma_f\rho_2)]$ and $\alpha_1 = \lambda\gamma_f\rho_2/[1-\gamma_f(\rho_1+\gamma_f\rho_2)]$. From (4), or Equation (6) in the authors' article, it is easy to see that under this setup the rank condition for identification is satisfied if and only if $\lambda \neq 0$ and $\rho_2 \neq 0$, as KM point out in their article. It seems at least in part on the basis of this analysis that KM have argued that the empirical NKPC model is weakly identified, since estimates of λ given in this and other articles have found it to be insignificantly different from zero.

We wish to show that whether or not λ being close to zero is indicative of weak identification might depend on particular features of the specification. To see this, consider the following scenario. Suppose that $\gamma_f + \gamma_b = 1$, so that we can rewrite the structural Equation (1) as:

$$\Delta\pi_t = \lambda x_t + \gamma_f \pi^*_{t+1} + e_t,$$

where $\pi^*_{t+1} = \pi_{t+1} - \pi_{t-1}$. Note that we focus on this restricted case with $\gamma_f + \gamma_b = 1$, since this restriction is imposed for much of KM's empirical analysis given in Section 5. In addition, consider the case where $1/2 < \gamma_f \leq 1$, so that, in light of the result given in Rudd and Whelan (2006), we have that π_t is an $I(0)$ process provided that x_t is an $I(0)$ process. Furthermore, suppose that $\mathbf{z}_t = (x_t, \pi_t)'$ follows a stable VAR(1) process, so that:

$$\begin{pmatrix} x_t \\ \pi_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ \pi_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \tag{5}$$

or in more succinct notation

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t,$$

where by stability $\det\{\mathbf{I}_2 - \mathbf{A}L\} \neq 0$ for $|L| \leq 1$. Note that the specification (5) is compatible with the authors' use of lagged values of x_t and π_t as instruments; it is also essentially a stripped-down version of a specification used in Rudd and Whelan (2005). Given this setup, it follows from a result given in Rudd and Whelan (2006) (and cited on page 20 of this article) that:

$$\begin{aligned}
\pi_t &= \left(\frac{1-\gamma_f}{\gamma_f}\right)\pi_{t-1} + \frac{\lambda}{\gamma_f}\sum_{j=0}^{\infty} E_t(x_{t+j}) + \frac{u_t}{\gamma_f} \\
&= \left(\frac{1-\gamma_f}{\gamma_f}\right)\pi_{t-1} + \frac{\lambda}{\gamma_f}\mathbf{e}_1'(\mathbf{I}_2 - \mathbf{A})^{-1}z_t + \frac{u_t}{\gamma_f} \\
&= \left[\left(\frac{\lambda}{\gamma_f}\right)\frac{(1-a_{22})}{\det[\mathbf{I}_2 - \mathbf{A}]}\right]x_{t-1} \\
&\quad + \left[\left(\frac{1-\gamma_f}{\gamma_f}\right) + \left(\frac{\lambda}{\gamma_f}\right)\frac{a_{12}}{\det[\mathbf{I}_2 - \mathbf{A}]}\right]\pi_{t-1} + \frac{u_t}{\gamma_f},
\end{aligned}$$

where $\mathbf{e}_1 = (1, 0)'$ and where the second equality above stems from the fact that $E_t(x_{t+j}) = \mathbf{e}_1'\mathbf{A}^j\mathbf{z}_t$. Although it is difficult in this case to solve explicitly for a_{21} and a_{22} as a function of λ, γ_f, a_{11}, and a_{12}, it is easy to see that in the case where $\lambda = 0$,

Journal of Business & Economic Statistics
July 2009, Vol. 27, No. 3
DOI: 10.1198/jbes.2009.08264

we have:

$$a_{21} = 0 \qquad \text{and} \qquad a_{22} = \frac{1-\gamma_f}{\gamma_f}.$$

It then follows by direct calculation that the mean function for the first-stage equation in this case is:

$$\begin{pmatrix} E_{t-1}(x_t) \\ E_{t-1}(\pi^*_{t+1}) \end{pmatrix} = \begin{pmatrix} E_{t-1}(x_t) \\ E_{t-1}(\pi_{t+1} - \pi_{t-1}) \end{pmatrix} = \underbrace{\begin{pmatrix} a_{11} & a_{12} \\ 0 & (1-2\gamma_f)/\gamma_f^2 \end{pmatrix}}_{\Pi} \underbrace{\begin{pmatrix} x_{t-1} \\ \pi_{t-1} \end{pmatrix}}_{Z_t},$$

so that the rank condition is satisfied in this case even when $\lambda = 0$. Moreover, note that in this case the rank condition fails only if either $a_{11} = 0$ or $\gamma_f = 1/2$, but, based on the authors' empirical results, there does not seem to be any statistical evidence for the latter, since from Table 2 of the article we see that the 95% confidence interval for γ_f using the robust MQLR (Moreira's quasi-likelihood ratio) test is [0.556, 1.053], so that $\gamma_f = 1/2$ lies outside of this interval.

Our purpose for constructing the above example is not to argue whether the specification given in this example is more or less reasonable than the authors' specification. We do feel, however, that the question of whether the empirical NKPC model studied in this article is weakly identified or not might require further examination. We also feel that it is important to check for the presence of weak identification since although there are test procedures which are robust to weak identification, the same could not be said for point estimation. Hence, in applications such as this one, where one might be interested in the reliability and the interpretability of the point estimates, it is important to have some diagnostic procedures which allow one to assess how weakly identified the underlying model is. While we realize that it would be difficult to get an explicit formula for the concentration parameter in the more complicated cases that are studied in Section 5 of the article, one way to obtain some information on the degree of identification weakness might be for the authors to try to estimate the concentration parameter by modifying or adapting the estimator proposed in Hansen, Hausman, and Newey (2008) to the present situation.

(ii) An area where we would have liked to see more discussion in this article is the growing literature on many instrument and many weak instrument asymptotics. We think the many instrument literature could be especially relevant for the NKPC model since empirical analysis of this model could be formulated in a natural way in terms of a conditional moment restriction setup, leading to possibly many unconditional moment conditions. KM do touch upon the subject of many instruments in Section 6 of their article on *Directions for Future Research*, in which the authors remark that:

> 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 instrument problem," which biases GMM in the direction of least squares; see Newey and Windmeijer (2005).

While we certainly do not advocate "the indiscriminate use of many instruments" (to quote what the authors wrote later in the same paragraph), we do wish to point out that, although there are some estimators such as 2SLS and the two-step generalized method of moments (GMM) estimator that suffer from large biases when many instruments are used, there are also others that do not. In fact, one achievement of the many instrument literature has been to identify those estimators that are not susceptible to what the authors refer to as the "many instrument problem" in the sense that the biases of these estimators do not increase appreciably even when the degree of overidentification is large. Indeed, Hausman et al. (2008) have recently proposed modified versions of limited-information maximum likelihood (LIML) and of the Fuller estimator (called HLIM and HFUL, respectively) that are consistent even in the presence of many weak instruments and heteroscedasticity. Moreover, in the nonlinear context, Newey and Windmeijer (2008) have shown that members of the class of Generalized Empirical Likelihood (GEL) estimators are more robust to the presence of many weak moment conditions, in the sense of being consistent and also having much smaller bias in finite sample, than the more commonly-used two-step GMM estimator. These results have added to earlier results by Morimune (1983), Bekker (1994), Chao and Swanson (2005), Stock and Yogo (2005), and Han and Phillips (2006), which collectively have shown in a variety of settings that consistent estimation is possible not only under many instruments, but also for models that do not satisfy the conventional assumptions of strong identification. The wealth of theoretical and also numerical/simulation results that have accumulated in the many instrument literature show that the so-called "many instrument problem" is not one for which we presently do not have a solution.

As a more general point, we are very positive about what the many instrument literature could offer for econometric practice. Under the many instrument asymptotics, estimators that are robust to the presence of many IV's (such as the HLIM, HFUL, and GEL estimators) can be shown to have convenient Gaussian limit distributions, although the form of the covariance matrix would involve an extra adjustment term relative to that obtained under the case of conventional asymptotics. In addition, Hansen, Hausman, and Newey (2008), Newey and Windmeijer (2008), Chao et al. (2008), and Hausman et al. (2008) have provided estimators of the covariance matrices that are consistent under both standard and many weak instrument asymptotics. Hence, under this approach, estimator uncertainty can be assessed and confidence sets can be constructed with the same ease as they are done under standard procedures—the only difference being that a (many instrument) corrected standard error is used in lieu of the usual standard error. In contrast, confidence sets based on inverting a weak instrument robust test statistic, such as the Kleibergen Lagrange Multiplier (KLM), the Jackknife version of Kleibergen's Lagrange Multiplier (JKLM), or the MQLR statistic discussed in this article, tend to be computationally more difficult. Moreover, while it is true that confidence sets based on the many instrument (MI) approach will no longer have the correct coverage probability when instrument weakness is too severe, theoretical and simulation results given in Hansen, Hausman, and Newey (2008), Newey and Windmeijer (2008), Chao et al. (2008), and Hausman et al. (2008) nevertheless suggest that the range of situations for which they are valid and applicable is likely to be very broad. In particular, Hansen, Hausman, and Newey (2008) performed an interesting exercise where they not only examined the performance of MI-corrected test procedures across a wide range of Monte Carlo designs with varying degrees of instrument weakness and endogeneity (where instrument weakness is measured by the value of the concentration parameter μ^2 and endogeneity is measured by ρ, the correlation coefficient between the error of the structural equation and that of the first-stage equation), but they also presented summary statistics for estimated μ^2 and ρ from applied microeconomic studies that had been published in the AER, the JPE, and the QJE roughly over the period 1999–2004. Matching the simulation results with the level of instrument weakness and endogeneity that was encountered in this sample of actual empirical studies, one could see that the MI-corrected test procedures would have performed well for a vast majority of the studies surveyed.

Finally, we note that much of the research on MI asymptotics has been conducted in settings that would seem more suitable for applied microeconomics in the sense that an independence assumption is typically maintained for the error process and sometimes for the instruments as well. However, it is precisely because such asymptotics have not been as well studied in the time series context that we think an MI analysis of the NKPC model would be an especially interesting topic for future research.

ADDITIONAL REFERENCES

Bekker, P. A. (1994), “Alternative Approximations to the Distributions of Instrumental Variable Estimators,” *Econometrica*, 62, 657–681.

Chao, J. C., and Swanson, N. R. (2005), “Consistent Estimation With a Large Number of Weak Instruments,” *Econometrica*, 73, 1673–1692.

Chao, J. C., Swanson, N. R., Hausman, J. A., Newey, W. K., and Woutersen, T. (2008), “Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression With Many Instruments,” working paper, MIT.

Han, C., and Phillips, P. C. B. (2006), “GMM With Many Moment Conditions,” *Econometrica*, 74, 147–192.

Hansen, C., Hausman, J. A., and Newey, W. K. (2008), “Estimation With Many Instrumental Variables,” *Journal of Business & Economic Statistics*, 26, 398–422.

Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. C., and Swanson, N. R. (2008), “Instrumental Variable Estimation With Heteroskedasticity and Many Instruments,” working paper, MIT.

Morimune, K. (1983), “Approximate Distributions of *k*-Class Estimators When the Degree of Overidentification Is Large Compared With the Sample Size,” *Econometrica*, 51, 821–842.

Newey, W. K., and Windmeijer, F. (2008), “GMM With Many Weak Moment Conditions,” working paper, MIT.

Rudd, J., and Whelan, K. (2005), “Does Labor’s Share Drive Inflation?” *Journal of Money, Credit, and Banking*, 37, 297–312.

Stock, J. H., and Yogo, M. (2005), “Asymptotic Distributions of Instrumental Variables Statistics With Many Instruments,” in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, eds. D. W. K. Andrews and J. H. Stock, Cambridge, MA: Cambridge University Press, pp. 109–120.

Comment

Jean-Marie Dufour

Department of Economics, McGill University, Leacock Building 519, 855 Sherbrooke W., Montréal, Québec H3A 2T7, Canada, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ) (*jean-marie.dufour@mcgill.ca*)

1. INTRODUCTION

The ground covered by Kleibergen and Mavroeidis (2009) is impressive. First, the authors survey new methods for inference in models with possibly weak instruments, especially in view of dealing with parameter subset inference in a GMM context. The test statistics considered include a number of *concentrated test statistics*: an *S*-type statistic based on the one proposed by Stock and Wright (2000), a Kleibergen-type (KLM) statistic (Kleibergen 2005), an overidentification test (JKLM) derived from the two previous procedures, and a conditional likelihood-ratio-type (LR-type) statistic (which extends the method of Moreira 2003). Second, the methods are applied to study a currently popular macroeconomic relation, the new Keynesian Phillips curve (NKPC), which now plays an important role in decisions about monetary policy. This type of model is especially important in countries that practice “inflation targeting” (like New Zealand, Canada, Australia, U.K., etc.).

The contribution of the authors is quite welcome, because for many years it appeared that macroeconomists had walked out of econometrics and serious empirical work. Recent econometric activity around the NKPC is certainly a comforting development for econometricians.

I will discuss the article by Kleibergen and Mavroeidis (2009) in the light of my own work on the econometric problems associated with weak identification (Dufour 1997, 2003; Dufour and Jasiak 2001; Dufour and Taamouti 2005, 2007; Doko Tchatoka and Dufour 2008) as well as NKPCs (Dufour, Khalaf, and Kichian 2006, 2007a, 2007b, 2008). I intend to focus on some pitfalls associated with the econometric methods proposed by the authors and potential research directions. Specific issues that will be discussed include:

1. Concerning econometric theory:
   (a) Inference in the presence of weak identification.
   (b) Limited information and robustness to missing instruments.
   (c) Projection methods and subset inference.
2. The meaning of the empirical results presented on NKPCs.

2. WEAK IDENTIFICATION AND STATISTICAL INFERENCE

In my view, recent econometric work on weak identification provides three main lessons:

Journal of Business & Economic Statistics
July 2009, Vol. 27, No. 3
DOI: 10.1198/jbes.2009.08283

1. Asymptotic approximations can easily be misleading. It is especially important in this area to produce a finite sample theory at least in a number of reference cases.
2. In structural models with identification difficulties, several of the intuitions that people draw from studying the linear regression model and using standard asymptotic approximations can easily be misleading. In particular, standard errors do not constitute a valid way of assessing parameter uncertainty and do not yield valid confidence intervals (Dufour 1997). Furthermore, individual parameters in statistical models are not generally meaningful, although parameter vectors are. Restrictions on the values of individual coefficients may be empirically empty, while restrictions on the whole parameter vector are empirically meaningful.
3. To build confidence sets (and to a lesser extent, tests), it is important to look for pivotal functions. Pivots are not generally available for individual parameters, but they can be obtained for appropriately selected parameter vectors. Given a pivot for a parameter vector, we can construct valid tests and confidence sets for the parameter vector. Inference on individual coefficients may then be derived through projection methods.

It is now widely accepted that inference in structural models should take into account the fact that identification may be weak. In so-called "linear IV regressions," this means taking care of the possibility of "weak instruments." In particular, this has led to the development of "identification robust" methods, which are based on first deriving some pivotal functions (at least asymptotically).

The point of departure of this work has been the finite sample procedure proposed long ago by Anderson and Rubin (1949, AR). However, it was soon noted that the AR procedure may involve sizeable "power losses" when the number of instruments used is large, and various methods aimed at improving this feature have been proposed (Kleibergen 2002; Moreira 2003). However, these "improvements" come at a cost. First, the justification of the methods is only asymptotic, which of course leaves open the possibility of arbitrary large size distortions even with fairly stringent distributional assumptions (convergence results are not uniform). Second, they are not robust to "missing instruments" and, more generally, to the formulation of a model for the explanatory endogenous variables. This latter problem has received little attention in the literature, so it is worthwhile to explain it in greater detail.

3. LIMITED INFORMATION AND ROBUSTNESS TO MISSING INSTRUMENTS

A central feature of most situations where IV methods are required comes from the fact that instruments may be used to solve an endogeneity or an errors-in-variables problem. It is very rare that one can or should use all the possible valid instruments.

Consider the standard model:

$$y = Y\beta + X_1\gamma + u, \tag{1}$$

$$Y = X_1\Pi_1 + X_2\Pi_2 + V, \tag{2}$$

where y and Y are $T \times 1$ and $T \times G$ matrices of endogenous variables, X_i is a $T \times k_i$ matrix of exogenous variables (instruments), $i = 1, 2$, β and γ are $G \times 1$ and $k_1 \times 1$ vectors of unknown coefficients, Π_1 and Π_2 are $k_1 \times G$ and $k_2 \times G$ matrices of unknown coefficients, u is a vector of structural disturbances, V is a $T \times G$ matrix of reduced-form disturbances, and $X = [X_1, X_2]$ is a full-column rank $T \times k$ matrix ($k = k_1 + k_2$). We wish to test

$$H_0(\beta_0) : \beta = \beta_0. \tag{3}$$

As mentioned above, a solution to the problem of testing in the presence of weak instruments has been available for more that 50 years (Anderson and Rubin 1949). On observing that

$$y - Y\beta_0 = X_1\theta_1 + X_2\theta_2 + \varepsilon, \tag{4}$$

where $\theta_1 = \gamma + \Pi_1(\beta - \beta_0)$, $\theta_2 = \Pi_2(\beta - \beta_0)$, and $\varepsilon = u + V(\beta - \beta_0)$, $H_0(\beta_0)$ can be tested by testing

$$H_0' : \theta_2 = 0. \tag{5}$$

If u is independent of X and $u \sim \mathrm{N}[0, \sigma_u^2 I_T]$, the AR statistic is the usual F-statistic for H_0':

$$\mathrm{AR}(\beta_0) = \frac{(y - Y\beta_0)'[M(X_1) - M(X)](y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(X)(y - Y\beta_0)/(T - k)} \sim F(k_2, T - k), \tag{6}$$

which yields the confidence set $C_\beta(\alpha) = \{\beta_0 : \mathrm{AR}(\beta_0) \le F_\alpha(k_2, T - k)\}$ for β.

A drawback of the AR method is that it loses power when too many instruments (X_2) are used. Potentially more powerful methods can be obtained by exploiting the special form (2) of the model for Y, which entails (among other things) the assumption that the mean of Y only depends on X_1 and X_2:

$$\mathsf{E}(Y) = X_1\Pi_1 + X_2\Pi_2. \tag{7}$$

This is what, in the end, methods like those proposed by Kleibergen (2002) or Moreira (2003) do.

Now suppose model (2) is in fact incomplete and a third matrix of instruments does indeed appear in the reduced form for Y:

$$Y = X_1\Pi_1 + X_2\Pi_2 + X_3\Pi_3 + V, \tag{8}$$

where X_3 is a $T \times k_3$ matrix of explanatory variables (not necessarily strictly exogenous). Equation (4) then becomes:

$$y - Y\beta_0 = X_1\Delta_1 + X_2\Delta_2 + X_3\Delta_3 + \varepsilon, \tag{9}$$

where $\Delta_1 = \gamma + \Pi_1(\beta - \beta_0)$, $\Delta_2 = \Pi_2(\beta - \beta_0)$, $\Delta_3 = \Pi_3(\beta - \beta_0)$, and $\varepsilon = u + V(\beta - \beta_0)$. Since $\Delta_2 = 0$ and $\Delta_3 = 0$ under H_0, it is easy to see that the null distribution of $\mathrm{AR}(\beta_0)$ remains $F(k_2, T - k)$. The AR procedure is *robust to missing instruments* (or *instrument exclusion*). It is also interesting to observe that the vectors $V_1, \ldots, V_T$ may not follow a Gaussian distribution and may be heteroscedastic. A similar result is obtained if

$$Y = g(X_1, X_2, X_3, V, \Pi), \tag{10}$$

where $g(\cdot)$ is an arbitrary (possibly nonlinear) function.

Alternative methods of inference aimed at being robust to weak identification (Wang and Zivot 1998; Kleibergen 2002; Moreira 2003) do not enjoy this type of robustness. The reason is that most of these methods exploit the specification

$$Y = X_1\Pi_1 + X_2\Pi_2 + V \tag{11}$$

for the reduced form.

In Dufour and Taamouti (2007), we present the results of a simulation study based on a model of the following form:

$$\begin{aligned} y &= Y_1\beta_1 + Y_2\beta_2 + u, \\ (Y_1, Y_2) &= X_2\Pi_2 + X_3\delta + (V_1, V_2), \end{aligned} \tag{12}$$

$$(u_t, V_{1t}, V_{2t})' \overset{\text{iid}}{\sim} \mathrm{N}(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.3 \\ 0.8 & 0.3 & 1 \end{pmatrix}, \tag{13}$$

where δ represents the importance of the excluded instrument; a sample of these results is reproduced in Table 1. These show clearly that methods that depend heavily on the specification (2) can suffer from large size distortions. This suggests that the problem of missing instruments may be as important in practice as the problem of weak instruments.

Methods that yield "power gains" by relying on additional restrictions on the reduced form of the model are closer to full information methods. Adding restrictions typically allows one to obtain more power (precision). However, if the restrictions used are not really part of the null hypothesis of interest, the resulting tests will be plagued by size distortions. This is the old trade-off between the inefficiency of limited information methods and the fragility of full information methods. In general, the latter cannot be viewed as substitutes for the former. What do we do when results conflict?

An important challenge consists in finding methods that are more powerful than AR-type procedures and are robust to missing instruments. This is feasible, for example, by using instrument reduction and transformation methods in conjunction with split-sample techniques; see Dufour and Taamouti (2003a, 2003b) and Dufour (2003).

Concerning the GMM procedures used by Kleibergen and Mavroeidis (2009), there is no proof or discussion as to whether these enjoy robustness to missing instruments. For example, problems similar to "missing instruments" may be introduced when potentially informative moment equations are not considered or dropped from the equations used for the GMM inference. Indeed, in the GMM setup considered by the authors, assumption (2) is replaced by high-level assumptions on the asymptotic distribution of the derivatives $q_t(\theta) = \partial f_t(\theta)/\partial\theta'$ of the moment equations (see Kleibergen and Movroeidis 2008). The latter appear to involve restrictions on the "reduced form" (model solution), though the exact nature of these restrictions is unclear. It seems plausible that the S-type procedure is less affected by such problems than the other statistics (since it is the procedure closest to the original AR method), but this remains to be seen. Anyway, I suggest it would be important to study this type of difficulty in the context of the models and methods considered by Kleibergen and Movroeidis (2009).

Table 1. Instrument exclusion and the size of tests robust to weak instruments. Random missing instruments. Nominal size = 0.05. Results are given in percentages

k_2	AR	ARS	K	LM	LR	LR1	LR2	AR	ARS	K	LM	LR	LR1	LR2
				(a) $\delta = 0$ and $\rho = 0.01$							(b) $\delta = 0$ and $\rho = 1$			
2	5.0	5.2	5.2	4.8	5.1	5.1	5.2	5.5	5.9	5.9	5.0	5.8	5.8	5.9
3	3.8	4.6	5.6	3.5	3.6	4.5	4.5	5.0	6.2	5.6	2.0	1.7	5.8	5.8
4	5.4	5.7	5.7	4.9	4.1	5.4	5.6	4.8	5.6	5.5	1.3	1.1	5.6	5.5
5	6.6	7.7	5.9	5.6	3.9	7.4	7.7	4.3	5.0	4.6	0.4	0.4	4.9	5.1
10	4.3	5.6	6.0	4.1	1.7	6.0	6.2	4.2	5.6	4.6	0.0	0.0	4.2	4.3
20	5.5	9.0	8.4	3.0	0.5	9.1	9.2	4.9	7.7	4.8	0.0	0.0	5.3	5.5
40	4.8	12.4	16.5	0.9	0.0	14.6	14.9	4.1	11.0	5.8	0.0	0.0	6.3	6.2
				(c) $\delta = 1$ and $\rho = 0.01$							(d) $\delta = 1$ and $\rho = 1$			
2	4.9	5.5	5.5	4.9	5.3	5.3	5.5	4.4	4.8	4.8	4.2	4.8	4.8	4.8
3	5.0	5.5	7.4	4.6	5.3	5.7	5.7	4.4	4.9	5.1	1.8	2.5	5.0	5.0
4	5.0	5.7	11.5	4.5	5.7	5.8	5.9	5.2	6.3	4.7	0.6	0.8	4.6	4.7
5	5.4	6.3	15.7	4.7	5.9	6.6	6.7	5.1	6.2	5.2	0.4	0.8	5.7	6.0
10	4.9	7.2	34.5	3.8	7.7	8.0	7.8	4.8	6.7	6.4	0.1	0.1	6.6	6.7
20	4.7	7.2	56.9	2.9	9.3	10.7	7.8	4.8	7.7	6.6	0.0	0.0	6.7	7.0
40	4.2	11.8	77.3	1.0	29.8	33.5	12.9	5.3	12.5	11.9	0.0	0.0	14.4	15.6
				(e) $\delta = 10$ and $\rho = 0.01$							(f) $\delta = 10$ and $\rho = 1$			
2	4.4	4.7	4.7	4.2	4.5	4.5	4.7	5.0	5.4	5.4	4.9	5.2	5.2	5.4
3	4.3	4.4	9.6	4.0	4.4	4.6	4.8	4.8	5.6	5.0	1.8	4.6	6.1	6.3
4	3.3	3.9	15.9	3.1	3.8	3.9	4.0	5.0	6.0	6.6	0.8	5.2	6.1	6.4
5	5.3	5.7	28.9	4.6	5.6	5.8	5.9	4.4	4.9	6.1	0.4	4.4	5.2	5.5
10	5.2	7.0	74.7	4.2	7.5	8.0	7.6	5.0	6.7	15.0	0.1	6.0	7.8	7.4
20	5.1	7.9	94.6	2.6	11.7	12.5	8.9	4.5	7.1	39.8	0.0	8.9	10.7	7.7
40	5.0	10.8	97.9	0.7	33.5	36.2	12.8	5.2	12.4	73.6	0.0	30.5	34.7	14.1

4. PROJECTION METHODS AND SUBSET INFERENCE

Inference on individual coefficients can be performed by using a projection approach. If

$$\mathsf{P}[\beta \in C_\beta(\alpha)] \geq 1 - \alpha, \tag{14}$$

then, for any function $g(\beta)$,

$$\mathsf{P}\big[g(\beta) \in g[C_\beta(\alpha)]\big] \geq 1 - \alpha. \tag{15}$$

If $g(\beta)$ is a component of β [or a linear transformation $g(\beta) = w'\beta$], the projection-based confidence set can be obtained very easily (Dufour and Jasiak 2001; Dufour and Taamouti 2005, 2007). This is a generic method with a finite sample justification. Furthermore, no restriction on the form of $g(\cdot)$ is required.

Kleibergen (2007) and Kleibergen and Mavroeidis (2008) claim it is possible to produce more efficient methods for subset inference by considering test statistics where the "nuisance parameters" have been replaced by point estimates (under the null hypothesis). This is certainly an interesting contribution. But there are three main limitations:

1. The argument is asymptotic.
2. In the GMM case, the "regularity conditions" bear not only on the moment variables $f_t(\theta)$ but also on the derivatives of these $q_t(\theta) = \frac{\partial f_t(\theta)}{\partial \theta}$, which involve implicit restrictions on the solution (reduced form) of the model.
3. As a result of the previous point, validity in cases where instruments are "missing" remains unproved (and doubtful).

The projection approach is applicable as soon as a test of the null hypothesis $\theta = \theta_0$ is feasible for all θ_0, which requires weaker assumptions than those used by the authors to ensure the validity of the concentrated identification robust GMM tests they propose. Of course, an interesting related issue would consist in studying to what extent these assumptions could be relaxed while preserving the validity of the concentrated test procedures.

5. WORK ON NEW KEYNESIAN PHILLIPS CURVES

In our own work on NKPCs (Dufour, Khalaf, and Kichian 2006), we focus on AR-type methods for producing inference on the parameters. Because of the arguments above, we think such methods are more robust and reliable. We have no reason to change our mind on that issue. If there is a strong disagreement between such methods and other identification robust methods (which may not be robust to the specification of the reduced form), we think conclusions from AR-type methods should prevail.

The parameters of NKPCs depend on deeper structural parameters, on which it is interesting to draw inference. This is done in our work using AR-type methods. It would be interesting to show that the methods proposed by the authors can be applied for that purpose and the results that can be obtained thereby.

We agree with the authors that many NKPC specifications are plagued with identification problems. But results may change dramatically when the definitions of variables, instruments, or small elements of the specification are modified. Identification robust methods in this context can prove to be very useful. Their work and ours (as well as others) provide an interesting illustration of that.

ACKNOWLEDGMENTS

This article was presented at the 2008 annual meeting of the American Statistical Association (Denver, Colorado). This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad Adenauer Fellowship (Alexander von Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

ADDITIONAL REFERENCES

Doko Tchatoka, F., and Dufour, J.-M. (2008), "Instrument Endogeneity and Identification-Robust Tests: Some Analytical Results," *Journal of Statistical Planning and Inference*, 138, 2649–2661.

Dufour, J.-M., and Taamouti, M. (2003a), "On Methods for Selecting Instruments," discussion paper, Université de Montréal, C.R.D.E.

——— (2003b), "Point-Optimal Instruments and Generalized Anderson-Rubin Procedures for Nonlinear Models," discussion paper, Université de Montréal, C.R.D.E.

Dufour, J.-M., Khalaf, L., and Kichian, M. (2007a), "Structural Multi-Equation Macroeconomic Models: A System-Based Estimation and Evaluation Approach," discussion paper, McGill University, Dept. of Economics, Montréal, Canada and Bank of Canada, Ottawa, Canada.

——— (2007b), "Which NKPC? Identification-Robust Estimates of Price Stickiness and Real Wage Rigidity," discussion paper, McGill University, Dept. of Economics, Montréal, Canada and Bank of Canada, Ottawa, Canada.

——— (2008),"How Much Do Real Wage Rigidities Matter for Canadian Inflation?" discussion paper, McGill University, Dept. of Economics, Montréal, Canada and Bank of Canada, Ottawa, Canada.

Kleibergen, F. (2007), "Subset Statistics in the Linear IV Regression Model," discussion paper, Brown University, Dept. of Economics, Providence, RI.

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

Wang, J., and Zivot, E. (1998), "Inference on Structural Parameters in Instrumental Variables Regression With Weak Instruments," *Econometrica*, 66 (6), 1389–1404.

Comment

Anna MIKUSHEVA
Department of Economics, Massachussetts Institute of Technology, Cambridge, MA 02141 (amikushe@mit.edu)

We are witnessing a growing awareness among applied researchers about the possibility of weak identification and its consequences, such as large biases, incorrect coverage of confidence sets, and the wrong size of tests.

Weak identification literature began more than a decade ago by documenting cases of weak instruments in important economic applications and by creating tests for weak identification. Two-step procedures, or so-called pretesting procedures, were popular for an extended period of time. In such a procedure one initially tests for weak instruments and then chooses a testing procedure or a confidence set depending on the result of the pretest; in particular, the 2SLS results are usually used if the pretest indicates strong instruments. However, pretesting procedures are not always reliable in controlling size and are considered an inferior solution to the weak identification problem (see, e.g., Andrews and Stock 2005). A qualitative change in the current literature is an attempt to create and use weak identification robust procedures as opposed to pretesting procedures. The article by Kleibergen and Mavroeidis is a big step forward in this direction.

In this comment I first discuss the theoretical advantage provided by a series of papers by Kleibergen and Mavroeidis that I really like. Then I move on to the applicability of the suggested methodology to the Phillips curve, which, to my understanding, still poses some questions.

1. PROCEDURES ROBUST TO WEAK INSTRUMENTS

The ideal weak identification robust procedure would possess two properties: (1) it should control for size no matter how strong or weak identification is; (2) it should not lose efficiency when compared to standard procedures if the identification is strong. Such a procedure may not exist if robustness incurs costs in terms of a power loss. Until now only one case (linear IV regression with a single endogenous regressor) seemed to be fully solved, and for this case the robustness to weak identification is achieved without loss of power in a strong instrument asymptotic. In this case we have several testing procedures [examples are Anderson and Rubin's 1949 test, the Lagrange multiplier (LM) test of Kleibergen 2002 and the conditional likelihood ratio (LCLR) test by Moreira 2003] satisfying the two conditions stated above. Andrews, Moreira, and Stock (2006) studied the optimality properties of such tests under an assumption of homoscedasticity. There is also software (`condivreg` procedure in STATA) that is easy to use and allows for production of robust tests and confidence sets in no time (Mikusheva and Poi 2006).

The article we are discussing today is a big step forward in creating robust procedures in a more general situation. The authors consider a generalized method of moments (GMM) setting, which includes linear IV as a special case, and allow for a multidimensional parameter space (in the IV case, it corresponds to more than one endogenous regressor).

While there are several known tests that control for size under weak instruments, such as a GMM analog of the Anderson–Rubin test (Stock and Wright 2000) and Kleibergen's KLM, JKLM, and MQLR, they produce conservative confidence sets if they are applied to a subset of parameters and if instruments are strong. That is, a trade-off exists between robustness to weak identification and efficiency under strong identification. Such a trade-off creates incentives to use pretesting. The article by Kleibergen and Mavroeidis solves this problem by correcting critical values for subset testing. The procedures suggested by the authors would asymptotically be the same as the usual GMM procedures if identification were strong, but would have correct coverage if in some dimensions identification were to fail. The main message of the discussed article is that one should not use pretesting, but rather a fully robust procedure.

2. APPLICATION TO PHILLIPS CURVE

Before discussing the Phillips curve application I want to draw special attention to the question of what robustness to weak identification means. The problem of weak identification can be explained as a problem of nonuniform or discontinuous asymptotics.

In the discussion below I distinguish between a parameter of interest, θ, and a nuisance (potentially infinitely dimensional) parameter, $\mathfrak{F}$. In the example of the Phillips curve the parameter of interest consists of coefficients $(\gamma_b, \gamma_f, \lambda)$, while the nuisance parameter includes everything else that can affect the finite sample distribution of the data; for example, it contains the parameters of a data-generating process for x_t and the distribution of error terms.

A GMM model parameter of interest is identified if the nuisance parameter is such that there exists a unique value of the parameter of interest that solves the theoretical moment condition. According to classic GMM asymptotic theory, if a parameter of interest is identified, then, as the sample size increases, the GMM estimate for θ converges to a normal distribution. In other words, the usual GMM t-statistic-based inferences are point-wise asymptotically correct. The statement does not hold if the parameter is not identified. We can conclude that the asymptotic theory is discontinuous at the point of nonidentification. It means that the closer the nuisance parameter is to the value for which the parameter of interest is not identified, the larger the sample required for the asymptotics to provide an accurate approximation. Put differently, the usual GMM inferences are not uniformly asymptotically correct. In other words, if the parameter is very close to a nonidentification point, the normal distribution provides a very poor description of the finite sample distribution of GMM t-statistics.

Journal of Business & Economic Statistics
July 2009, Vol. 27, No. 3
DOI: 10.1198/jbes.2009.08263

Kleibergen and Mavroeidis make the inferences robust to weak instruments by suggesting a procedure that is point-wise asymptotically correct and in which asymptotics is uniform along this one specific direction, closeness to a nonidentification point. The authors do not attempt to make inferences uniformly asymptotically correct over the whole space of possible values of $\mathfrak{F}$. The main reason is that asymptotic uniformity often bears some costs in terms of power.

The main assumption (Assumption 1) states that the nuisance parameter is such that the convergence below holds:

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

It differs from the standard GMM assumptions by substituting an identification assumption with an assumption that the sum of the derivative q_t is an asymptotically Gaussian process. The authors claim that Assumption 1 is very general. Assumption 1 is a point-wise asymptotic assumption, while the inferences about θ using a realization of $(\psi_f(\theta), \psi_\theta(\theta))$ are done uniformly with respect to the rank of the covariance matrix.

Whenever we want to apply the suggested procedures we have to ask ourselves whether Assumption 1 seems reasonable in the setting discussed. In the Phillips curve application I can see at least two complications: the persistence of variables and the many instruments problem.

Assumption 1 suggests that the sums of moment conditions and their first derivatives should be approximately normal. The inflation π_t (lags of which are used as instruments) is a highly persistent time series; the unit root hypothesis for it usually cannot be rejected. One can observe the normalized sum $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{f}_t(\theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t e_t$ has $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \pi_{t-1} e_t$ as one of the components. The normality of this sum for a highly persistent series could be problematic. For example, consider a case when $\lambda = 0, \gamma_b + \gamma_f = 1$, then π_t is a unit root process and $\frac{1}{T} \sum_{t=1}^{T} \pi_{t-1} e_t \to^d \omega \int_0^1 w(t)\, dw(t)$. Notice two things: a nonstandard normalization should be used to get a limiting distribution, and the limit is not normal. A similar problem arises when the process does not have a unit root but is modeled as local to unit root (see, e.g., Bobkovski 1983 and Phillips 1987). We can also argue that the partial sums of $\bar{q}_t$ may also be nonnormal for nonstationary components. In general, a use of persistent instruments may lead to Assumption 1's failure to hold.

One of the solutions suggested in the article is to use only stationary series as instruments; for example, lags of differenced inflation $\Delta\pi$ rather than lags of inflation itself. This is, however, only a partial solution. If all instruments are stationary, then it seems plausible that the sum $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{f}_t(\theta_0)$ is approximately normally distributed. This suggests that an S-statistic for the *whole* parameter vector θ will be approximately χ-square distributed and can produce a test with a good size property. However, since γ_b and γ_f are coefficients on inflation, the first derivative of the moment condition, q_t, involves persistent summands such as a lag of inflation. We have returned to the problem discussed in the previous paragraph; namely, that the validity of Assumption 1 can be questioned. This makes the critical value correction suggested by Kleibergen and Mavroeidis unapplicable to the Phillips curve. While Kleibergen and Mavroeidis's procedure is robust to weak identification, the assumption it is based on is nonuniform to the persistence of regressors and instruments. This may lead to unreliable inferences.

The second concern I have is that requiring normality of sums of $\bar{q}_t$ in addition to the sums of $\bar{f}_t$ increases the dimensionality of the applied central limit theorem and makes the many instrument problem more severe. In the Phillips curve case the authors consider 6 instruments, which is a moderate number for a sample size of slightly fewer than 200. While the dimensionality of $\bar{f}_t$ is 6, the dimensionality of $\bar{q}_t$ is 18, which makes the total dimensionality of Assumption 1 equal to 24. It is very hard to believe that a 24-dimensional central limit theorem would provide a good approximation when one has only 200 time periods, even though in a case of stationarity one can easily believe in the 6-dimensional Central Limit Theorem for $\bar{f}_t$ only.

To summarize, Assumption 1 may be somewhat restrictive and questionable in the Phillips curve application. It is partially due to the time series nature of the problem and small sample sizes. An asymptotic efficiency-oriented Kleibergen and Mavroeidis procedure seems to be more applicable to large cross-section datasets.

ADDITIONAL REFERENCES

Bobkovski, M. J. (1983), "Hypothesis Testing in Nonstationary Time Series," unpublished Ph.D. thesis, University of Wisconsin, Madison, Dept. of Statistics.

Mikusheva, A., and Poi, B. (2006), "Tests and Confidence Sets With Correct Size When Instruments Are Potentially Weak," *Stata Journal*, 6 (3), 335–347.

Phillips, P. C. B. (1987), "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74 (3), 535–547.

Comment

Jonathan H. Wright

Department of Economics, Johns Hopkins University, Baltimore, MD 21218 (*wrightj@jhu.edu*)

Forward-looking macroeconomic models, such as the new Keynesian Phillips curve (NKPC) or forecast-based Taylor rules, are routinely estimated by replacing expectations with future realized values and then estimating by instrumental variables. Under rational expectations, any variable in the information set at the time that the expectation is being formed must be a valid instrument in the sense of being orthogonal to the error term. But these instruments may still be weak, and this motivates approaches to inference that are robust in the sense that

Journal of Business & Economic Statistics
July 2009, Vol. 27, No. 3
DOI: 10.1198/jbes.2009.08274

tests will continue to be correctly sized even if the instruments are weak. Most of the literature on the NKPC has ignored issues of weak instruments. The few papers that have used identification robust methods tend to find very wide confidence sets.

Perhaps this is a correct statement of our lack of ability to identify the structural parameters of a NKPC, but a great deal of progress has been made in the last decade in making identification robust methods more efficient. The earliest identification robust approach to inference, proposed by Anderson and Rubin (1949), tests any hypothesized parameter value by computing the implied structural errors and projecting these onto the space of instruments. The more recent identification robust tests have instead projected the putative structural errors onto a subspace of dimension equal to the number of parameters. That is likely to be more efficient if there are surplus instruments.

In this article, alternative identification robust approaches to inference are compared and the MQLR method is found to be best in the sense of controlling coverage while minimizing the size of the confidence sets. That is a very useful result, but the identification of the NKPC is still weak and this motivates the quest for more identifying information. The article uses a few lags of inflation and marginal cost as instruments. As the authors observe, simply bringing in a much larger number of instruments is unlikely to help. But, in a time series context, an instrument is a forecast. The choice of instruments is a choice about how we want to make forecasts. Fortunately, we know quite a bit about forecasting inflation and measures of economic slack.

First, we know from authors like Croushore (2006) that data vintage can matter. Second, we know that pooling the information in large datasets helps with forecasting inflation, whether using factor methods, model averaging, or other techniques suitable in a data-rich environment (see, for example, Stock and Watson 1999). Third, judgmental forecasts such as the Survey of Professional Forecasters (SPF) or the Federal Reserve's Greenbook provide excellent forecasts of inflation that seem to beat all time series methods (Ang, Bekaert, and Wei 2007 and Faust and Wright 2009), perhaps because there is something about the low frequency behavior of inflation that judgment captures much better than time series models. Finally, for forecasting economic growth, the very simplest autoregressive time series forecasts are hard to beat (Faust and Wright 2009).

To illustrate how these principles of forecasting might help with the problem of the NKPC, I considered inference in the NKPC imposing that the coefficients γ_f and γ_b sum to one; that is, using the moment condition

$$E\big[Z_t(\pi_t - \lambda x_t - \gamma_f(\pi_{t+1} - \pi_{t-1}))\big] = 0$$

in the notation of the article. The authors kindly provided me with their data and I compared the Anderson–Rubin, or S-set, for λ and γ_f in this model, computing in three ways:

(i) Using the sample period 1969Q1–2007Q4 with two lags of inflation changes and three lags of marginal cost as instruments, as in the article.

(ii) Repeating exactly the same data, but with real-time inflation data (defining inflation as the series observed in the middle of the quarter after the quarter to which they refer, obtained from the Philadelphia Fed real-time dataset).

(iii) Using real-time inflation data and an alternative set of instruments: one lag of the inflation change, the Greenbook forecast of the change in inflation (and the SPF within the last five years), and one lag of marginal cost.

The three alternative confidence sets are shown in Figure 1. Comparing (i) and (ii), it seems that using real-time inflation data makes a difference, but they both give confidence sets of about the same size. However, using the alternative instruments in (iii) gives a dramatic reduction in the size of the confidence set. The instruments in (iii) seem closer to what we would want to use if our goal is to forecast inflation and marginal cost. This exercise is simply intended as an illustration of the gains that may be available from using a choice of instruments that is motivated by thinking of it as a forecasting problem. The idea of using judgmental forecasts to provide instruments for future inflation is close to—but not exactly the same as—a method of inference for the NKPC considered by several papers including Roberts (1995), Dufour, Khalaf, and Kichian (2006), Nason and Smith (2008), and Smith (2009), which is to measure expected future inflation by these judgmental forecasts. I am, however, not aware of existing papers that put realized future inflation on the right-hand side of the NKPC and then use judgmental forecasts as instrumental variables.

If the researcher prefers not to rely on judgmental forecasts, then another option is to predict inflation using methods appropriate to a large dataset (factor methods, model averaging, etc.) and to use the resulting forecasts as instruments for future inflation in the NKPC. In either case, the idea is to use some preliminary dimensionality reduction step—which is implicit in judgmental forecasts and explicit in factor methods or model averaging—to expand the information set without running into the "many instruments" problem.

On a different topic, several authors have found considerable empirical support for a model in which inflation has trend and transitory components, including Kozicki and Tinsley (2001), Gürkaynak, Sack, and Swanson (2005), Cogley, Primiceri, and Sargent (2007), and Stock and Watson (2007). Intuitively, permanent and transitory shocks to inflation expectations should have very different effects on the behavior of households and firms. A firm may be willing to absorb a temporary increase in expected inflation in reduced profit margins, but cannot absorb a permanent increase in inflation expectations. Accordingly, the relevant forward-looking inflation measure is not just next period's inflation.

Indeed, the microfoundations of the NKPC considered in this article are based on an approximation around a steady state with constant inflation (actually zero inflation). But Cogley and Sbordone (2008) derive an NKPC with drifting trend inflation. As one might expect, it is not just next period's inflation that appears on the right-hand side—longer-term inflation expectations matter, too. In future work, it would be interesting to apply weak identification methods to Cogley and Sbordone's forward-looking Phillips curve. It may be hard because it has far more parameters than the standard NKPC. Identification robust methods are all based on inverting the acceptance regions of suitable test statistics and, as such confidence sets, are hard to represent in a high-dimensional parameter space. Moreover, as Kleibergen and Mavroeidis explain clearly, while there are methods for doing identification robust inference on subsets of

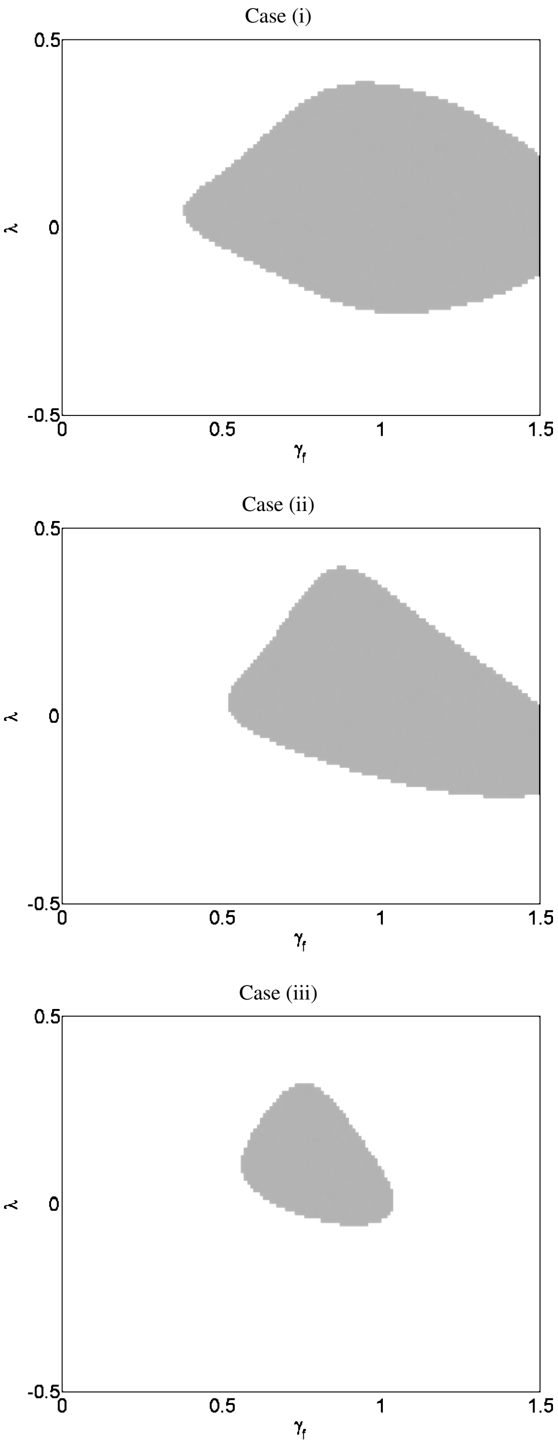

Figure 1. Three alternative confidence sets for the parameters of the NKPC model: Case (i): Kleibergen–Mavroeidis instruments; revised data. Case (ii): Kleibergen–Mavroeidis instruments; real-time inflation data. Case (iii): alternative instruments; real-time inflation data. *Notes*: Cases (i) and (ii) use two lags of inflation changes and three lags of marginal cost as instruments. In Case (ii), the inflation data are as observed in the first quarter after the quarter to which the data refer from the real-time dataset of the Philadelphia Fed. In Case (iii), the instruments are instead one lag of inflation changes, one lag of marginal cost, and the Greenbook forecast of the change of inflation (SPF used for the last five years).

parameters, these procedures are all asymptotically conservative. A cheap and cheerful alternative would be to fit the usual NKPC but replace inflation everywhere by the gap between inflation and trend inflation that might, for example, be proxied by the long-term SPF inflation forecast.

The possibility of a nonstationary component in inflation has a more technical—but nonetheless potentially important—econometric implication. In Assumption 1, the authors assume that the sample moment conditions evaluated at the true parameter value are asymptotically normal. That is a central assumption to all the work on weak identification in the generalized method of moments (GMM). But, if inflation is nonstationary—and is used in levels—then Assumption 1 cannot hold. Instead some sample moment conditions will converge to a functional of a Brownian motion.

One final comment is that the approach to inference used in this article is to abandon conventional inference and go straight to methods that are robust to weak identification without first checking that there is a failure of identification in the first place. Indeed, the estimation of the NKPC shows all the symptoms of weak identification and I am convinced that identification robust methods are the only ones that are appropriate in this context. There is a logic to using only identification robust methods. But there are costs to using methods that are identification robust. First, inference on a subset of parameters can only be done by asymptotically conservative methods. Second, identification robust confidence sets are hard to represent, especially when the dimension of the parameter space is big. Third, the robust approach to inference gives up on point estimation in the standard sense of the term (although Hodges–Lehman estimators—which are not consistent under weak identification—are still available). As a result, I believe that it is helpful to also report tests for whether there is a weak identification problem, along the lines of the test proposed by Stock and Yogo (2005). In the (perhaps rare) cases where these tests indicate that parameters are well identified, researchers can report conventional point estimates and standard errors.

But in the case at hand, the enormous difference between confidence sets using Wald and identification robust methods and the sensitivity of the conventional NKPC parameter estimates to normalization proves the case for weak identification beyond reasonable doubt. Methods that are robust to weak identification, yet are as efficient as possible, as proposed by the authors, should become the standard approach to inference on the NKPC and in estimation of other related forward-looking macroeconomic models.

ACKNOWLEDGMENTS

I am grateful to Jim Nason and Gregor Smith for their helpful comments. All errors are my sole responsibility.

ADDITIONAL REFERENCES

Ang, A., Bekaert, G., and Wei, M. (2007), "Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?" *Journal of Monetary Economics*, 54, 1163–1212.

Cogley, T., and Sbordone, A. M. (2008), "Trend Inflation, Indexation and Inflation Persistence in the New Keynesian Phillips Curve," *American Economic Review*, 98, 2101–2126.

Cogley, T., Primiceri, G. E., and Sargent, T. J. (2007), "Inflation-Gap Persistence in the U.S.," working paper, New York University and Northwestern University.

Croushore, D. (2006), "Forecasting With Real-Time Macroeconomic Data," in *Handbook of Economic Forecasting*, eds. G. Elliott, C. W. J. Granger, and A. Timmermann, Amsterdam: North-Holland.

Faust, J., and Wright, J. H. (2009), "Comparing Greenbook and Reduced Form Forecasts Using a Large Realtime Dataset," *Journal of Business & Economic Statistics*, to appear.

Gürkaynak, R. S., Sack, B., and Swanson, E. T. (2005), "The Excess Sensitivity of Long Term Interest Rates: Evidence and Implications for Macroeconomic Models," *American Economic Review*, 95, 425–436.

Kozicki, S., and Tinsley, P. A. (2001), "Shifting Endpoints in the Term Structure of Interest Rates," *Journal of Monetary Economics*, 47, 613–652.

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

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

Smith, G. W. (2009), "Pooling Forecasts in Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, to appear.

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

Comment

Motohiro Yogo

The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 and the National Bureau of Economic Research, Cambridge, MA 02138 (*yogo@wharton.upenn.edu*)

1. PROBLEM OF WEAK IDENTIFICATION

Under the restriction that the time discount factor is one, the new Keynesian Phillips curve is given by

$$\pi_t = \lambda s_t + \gamma_f E_t[\pi_{t+1}] + (1 - \gamma_f)\pi_{t-1}, \tag{1}$$

where π_t is inflation and s_t is the output gap at time t. The underlying structural parameters are $\theta \in [0, 1]$, the frequency of price adjustment, and $\omega \in [0, 1]$, the fraction of backward-looking price setters. They are related to the parameters in the Phillips curve through

$$\lambda = \frac{(1-\omega)(1-\theta)^2}{\theta + \omega},$$

$$\gamma_f = \frac{\theta}{\theta + \omega}.$$

Galí and Gertler (1999, p. 212) recognized a "small sample normalization problem in GMM." Because of this problem, they estimate the Phillips curve based on two alternative moment restrictions,

$$E_t\big[(\theta + \omega)(\pi_t - \lambda s_t - \gamma_f \pi_{t+1} - (1 - \gamma_f)\pi_{t-1})z_t\big] = 0, \tag{2}$$

$$E_t\big[(\pi_t - \lambda s_t - \gamma_f \pi_{t+1} - (1 - \gamma_f)\pi_{t-1})z_t\big] = 0. \tag{3}$$

Table 1 reports their estimates.

Based on moment restriction (2), the 95% asymptotic confidence intervals are $\lambda \in [0.02, 0.04]$ and $\gamma_f \in [0.74, 0.80]$. Based on moment restriction (3), the 95% asymptotic confidence intervals are $\lambda \in [0.00, 0.01]$ and $\gamma_f \in [0.58, 0.65]$. Quite disturbingly, the confidence intervals produced by the two alternative normalizations are disjoint.

The fact that the asymptotic confidence interval can be disjoint, depending on the normalization of the moment restriction, has been emphasized by Hahn and Hausman (2002) and Hamilton, Waggoner, and Zha (2007). Another economic application for which this same problem arises is the estimation of the elasticity of intertemporal substitution (Neely, Roy, and Whiteman 2001). In that application, however, precise estimates can be obtained using methods that are robust to weak identification (Stock and Wright 2000; Yogo 2004).

2. ESTIMATES ROBUST TO WEAK IDENTIFICATION

Kleibergen and Mavroeidis (2009) estimate the Phillips curve using the methodology developed by Kleibergen (2005), which leads to valid confidence sets even in the presence of weak identification. Table 2 reports their estimates.

The 95% confidence intervals are $\lambda \in [0.00, 0.17]$ and $\gamma_f \in [0.56, 1.00]$. These confidence intervals are much wider than those reported by Galí and Gertler (1999), which indicates that weak identification is a serious issue in estimation of the Phillips curve. To get a sense of the imprecision of these estimates, the implied 95% confidence intervals for the underlying structural parameters are $\omega \in [0.00, 0.80]$ and $\theta \in [0.50, 1.00]$. The confidence interval for the frequency of price adjustment implies that the duration is between six months and forever! Because of weak identification, structural parameters cannot be identified based on macrodata alone.

3. COMPARISON TO MICROESTIMATES

In order to understand the Phillips curve, it is important to examine its microfoundations through firms' pricing behavior. In a recent study, Nakamura and Steinsson (2008) estimate the frequency of price adjustment for various goods and services using microdata from the Bureau of Labor Statistics. Table 3 reports some relevant numbers from their study.

The median duration between price adjustment is 11 months for all sectors, which is remarkably close to the point estimate of 13 months obtained by Kleibergen and Mavroeidis (2009) for the 1960–2007 sample. There is substantial cross-sectional variation in the duration between price adjustment across sectors, ranging from 0.5 months for vehicle fuel to 27.3 months for apparel.

Journal of Business & Economic Statistics
July 2009, Vol. 27, No. 3
DOI: 10.1198/jbes.2009.08275

Table 1. Estimates from Galí and Gertler (1999)

	λ	γ_f	ω	θ	Duration
		Panel A: Estimate based on moment restriction (2)			
Point estimate	0.03	0.77	0.24	0.80	15
95% CI	[0.02, 0.04]	[0.74, 0.80]	[0.19, 0.30]	[0.77, 0.84]	[13, 18]
		Panel B: Estimate based on moment restriction (3)			
Point estimate	0.01	0.62	0.52	0.84	19
95% CI	[0.00, 0.01]	[0.58, 0.65]	[0.44, 0.61]	[0.79, 0.89]	[14, 28]

NOTE: This table is taken from Galí and Gertler (1999, table 2). Estimation is by GMM with the Newey–West covariance matrix. The instruments include four lags of inflation, labor income share, long-short yield spread, output gap, wage inflation, and commodity price inflation. The brackets contain asymptotic 95% confidence intervals. The last column reports the implied duration between price adjustment in months. The data are quarterly for the sample period 1960:1–1997:4.

Table 2. Estimates from Kleibergen and Mavroeidis (2009)

	λ	γ_f	ω	θ	Duration
		Panel A: 1960–2007 sample			
Point estimate	0.04	0.77	0.23	0.77	13
95% CI	[0.00, 0.17]	[0.56, 1.00]	[0.00, 0.80]	[0.50, 1.00]	$[6, \infty]$
Restricted 95% CI	[0.02, 0.10]	[0.56, 1.00]	[0.00, 0.58]	0.73	11
		Panel B: 1960–1983 sample			
Point estimate	0.21	0.81	0.14	0.58	7
95% CI	[0.00, 0.62]	[0.46, 1.00]	[0.00, 1.00]	[0.27, 1.00]	$[4, \infty]$
Restricted 95% CI	[0.01, 0.10]	[0.46, 1.00]	[0.00, 0.87]	0.73	11
		Panel C: 1984–2007 sample			
Point estimate	0.01	0.79	0.23	0.87	24
95% CI	[0.00, 0.23]	[0.57, 1.00]	[0.00, 0.77]	[0.46, 1.00]	$[6, \infty]$
Restricted 95% CI	[0.03, 0.10]	[0.57, 1.00]	[0.00, 0.56]	0.73	11

NOTE: This table is taken from Kleibergen and Mavroeidis (2009, tables 2 and 3). Estimation is by continuous updating GMM with the Newey–Weight covariance matrix. The instruments include three lags of first differenced inflation and output gap. The brackets contain 95% confidence intervals based on the test in Kleibergen (2005). The restricted confidence interval, which imposes $\theta = 0.73$, is based on the author's calculations. The last column reports the implied duration between price adjustment in months. The data are quarterly for the sample period 1960:1–2007:3.

Table 3. Frequency of price adjustment

Measure	Frequency	Duration
All sectors	8.7	11.0
Processed food	10.5	9.0
Unprocessed food	25.0	3.5
Household furnishing	6.0	16.1
Apparel	3.6	27.3
Transportation goods	31.3	2.7
Recreation goods	6.0	16.3
Other goods	15.0	6.1
Utilities	38.1	2.1
Vehicle fuel	87.6	0.5
Travel	41.7	1.9
Services (excluding travel)	6.1	15.8

NOTE: This table is taken from Nakamura and Steinsson (2008, table 2), for median estimates that exclude sales and substitutions. The frequency of price adjustment is reported as percent per month, and the implied duration is reported in months. The sample period is 1998–2005.

One way to improve the precision of the estimates in the Phillips curve is to impose the frequency of price adjustment measured in microdata. A duration of 11 months implies a quarterly frequency of price adjustment of $\theta = 0.73$. The last row of each panel in Table 2 reports the 95% confidence intervals for λ and γ_f under the restriction that $\theta = 0.73$. For the 1960–2007 sample, the 95% confidence interval for λ shrinks to [0.02, 0.10], compared to [0.00, 0.17] without the restriction. However, the restriction does not improve the confidence interval for γ_f. Intuitively, it is difficult to identify whether firms are responding to $E_t[\pi_{t+1}]$ or π_{t-1} due to the persistence in inflation.

In conclusion, Kleibergen and Mavroeidis (2009) have made a convincing case that weak identification is a serious issue in estimation of the Phillips curve. Future studies should strive to achieve better identification through the use of rich microdata.

ACKNOWLEDGMENTS

I thank Frank Kleibergen, Sophocles Mavroeidis, Jonathan Wright, and participants of the 2008 Joint Statistical Meetings for helpful comments and discussions.

ADDITIONAL REFERENCES

Hahn, J., and Hausman, J. (2002), "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163–189.

Hamilton, J. D., Waggoner, D. F., and Zha, T. (2007), "Normalization in Econometrics," *Econometric Reviews*, 26, 221–252.

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

Nakamura, E., and Steinsson, J. (2008), "Five Facts About Prices: A Reevaluation of Menu Cost Models," *Quarterly Journal of Economics*, 123, 1415–1464.

Neely, C. J., Roy, A., and Whiteman, C. H. (2001), "Risk Aversion versus Intertemporal Substitution: A Case Study of Identification Failure in the Intertemporal Consumption Capital Asset Pricing Model," *Journal of Business & Economic Statistics*, 19, 395–403.

Yogo, M. (2004), "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," *Review of Economics and Statistics*, 86, 797–810.

Comment

Eric ZIVOT
Department of Economics, University of Washington, Seattle, WA 98195-3330 (*ezivot@u.washington.edu*)

Saraswata CHAUDHURI
Department of Economics, University of North Carolina, Chapel Hill, NC 27599 (*saraswata_chaudhuri@unc.edu*)

1. INTRODUCTION

The paper by Kleibergen and Mavroeidis (2009), hereafter KM, is an excellent survey of the current state of the art in the weak instrument robust econometrics for testing subsets of parameters in the generalized method of moments (GMM), and provides an important and relevant application of the econometric theory to the analysis of the new Keynesian Phillips curve. We are extremely grateful to have the opportunity to comment on this very nice paper. Our comments will focus on the weak instrument robust tests for subsets of parameters, and in particular on the projection-based test that KM refer to as the Robins (2004) test.

We show that KM's implementation of the Robins test is inefficient, and provide an efficient implementation that performs nearly as well as the MQLR test recommended by KM. Our comment proceeds as follows. Section 2 reviews the tests used for inference on subsets of parameters in GMM and discusses in detail the implementation of the Robins test, which we call the new method of projection. Section 3 reports the results of a small simulation study to demonstrate that the new method of projection performs nearly as well as the tests recommended by KM. Section 4 contains our concluding remarks.

2. INFERENCE ON SUBSETS OF PARAMETERS IN GMM

In this section we describe inference on subsets of parameters in the GMM framework. We follow the notation and assumptions of KM regarding the GMM framework. Interest centers on a p-dimensional vector of parameters θ identified by a set of $k \geq p$ moment conditions

$$E[f_t(\theta)] = 0.$$

Let $\theta = (\alpha', \beta')'$, where α is $p_\alpha \times 1$ and β is $p_\beta \times 1$. The parameters of interest are β, and α are considered nuisance parameters. The weak identification robust methods of inference on θ are based on the (efficient) continuous updating (CU) GMM objective function

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

where $f_T(\theta) = T^{-1} \sum_{t=1}^{T} f_t(\theta)$ and $\hat{V}_{ff}(\theta)$ is a consistent estimator of the $k \times k$ dimensional covariance matrix $V_{ff}(\theta)$ of the vector of sample moments. Let $q_t(\theta) = \mathrm{vec}(\frac{\partial f_t(\theta)}{\partial \theta'})$ and define $\bar{f}_t(\theta) = f_t(\theta) - E[f_t(\theta)]$ and $\bar{q}_t(\theta) = q_t(\theta) - E[q_t(\theta)]$. Assumption 1 of KM states that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{bmatrix} \bar{f}_t(\theta) \\ \bar{q}_t(\theta) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \psi_f(\theta) \\ \psi_\theta(\theta) \end{bmatrix} \sim \mathrm{N}(0, V(\theta)),$$

$$V(\theta) = \begin{bmatrix} V_{ff}(\theta) & V_{f\theta}(\theta) \\ V_{\theta f}(\theta) & V_{\theta\theta}(\theta) \end{bmatrix}.$$

The gradient of (1) with respect to θ is given by

$$\nabla_\theta Q(\theta) = \frac{\partial Q(\theta)}{\partial \theta'} = 2 f_T(\theta)' \hat{V}_{ff}(\theta)^{-1} \hat{D}_T(\theta),$$

where $\hat{D}_T(\theta) = \sum_{t=1}^{T} D_t(\theta)$ and $D_t(\theta) = \mathrm{devec}_k[q_t(\theta) - \hat{V}_{\theta f}(\theta) \hat{V}_{ff}(\theta)^{-1} f_t(\theta)]$. For the definition of the devec operator see Chaudhuri (2008).

2.1 Tests for the Full Parameter Vector

Valid tests of the hypothesis $H_0 : \theta = \theta_0$ were developed in Stock and Wright (2000) and Kleibergen (2005). Stock and Wright's S-statistic is a generalization of the Anderson–Rubin statistic (see Anderson and Rubin 1949) and is given by $S(\theta) = Q(\theta)$. Kleibergen's K-statistic is a score-type statistic based on $Q(\theta)$ and may be expressed as

$$K(\theta) = \frac{1}{4} (\nabla_\theta Q(\theta)) [\hat{D}_T(\theta)' \hat{V}_{ff}(\theta)^{-1} \hat{D}_T(\theta)]^{-1} (\nabla_\theta Q(\theta))'. \tag{2}$$

Journal of Business & Economic Statistics
July 2009, Vol. 27, No. 3
DOI: 10.1198/jbes.2009.08305

Under the null $H_0:\theta=\theta_0$, $S(\theta_0)\overset{A}{\sim}\chi^2_k$ and $K(\theta_0)\overset{A}{\sim}\chi^2_p$.

2.2 Tests for Subsets of Parameters

For testing hypotheses on subsets of parameters of the form $H_0:\beta=\beta_0$, subset versions of the S- and K-statistics were also considered by Stock and Wright (2000) and Kleibergen (2005). These statistics are based on the plug-in principle and utilize the constrained CU-GMM estimate $\tilde{\alpha}(\beta_0)=\arg\min_\alpha Q(\alpha,\beta_0)$. Letting $\tilde{\theta}_0=(\tilde{\alpha}(\beta_0)',\beta_0')'$, the subset S- and K-statistics are given by $S(\tilde{\theta}_0)$ and $K(\tilde{\theta}_0)$, respectively. Under the null $H_0:\beta=\beta_0$ and under the assumption that α is well identified, Stock and Wright (2000) and Kleibergen (2005) showed that $S(\tilde{\theta}_0)\overset{A}{\sim}\chi^2_{k-p_\alpha}$ and $K(\tilde{\theta}_0)\overset{A}{\sim}\chi^2_{p_\beta}$. This result is based on the fact that when α is well identified, $\tilde{\alpha}(\beta_0)$ is $\sqrt{n}$ consistent for α under $H_0:\beta=\beta_0$. When α is not well identified, $\tilde{\alpha}(\beta_0)$ is no longer $\sqrt{n}$ consistent for α and hence the S- and K-statistics are not asymptotically chi-square distributed. However, Theorem 1 of KM shows that irrespective of the identification of α, the S- and K-statistics are always bounded from above by the $\chi^2_{k-p_\alpha}$ and $\chi^2_{p_\beta}$ distributions, respectively.

2.3 Usual Method of Projection

Dufour (1997), Dufour and Jasiak (2001), and Dufour and Taamouti (2005, 2007) showed that the usual projection approach could always be used to obtain valid inference for subsets of parameters provided there exists an asymptotically (boundedly) pivotal statistic for testing the joint hypothesis $H_0:\theta=\theta_0$. Let $R(\theta)$ denote such a statistic and assume that $R(\theta)\overset{A}{\sim}\chi^2_v$. Suitable choices for $R(\theta)$ are $S(\theta)$, for which $v=k$, and $K(\theta)$, for which $v=p$. The usual method of projection rejects $H_0:\beta=\beta_0$ at level (at most) ζ if

$$\inf_{\alpha\in\Theta_\alpha} R(\alpha,\beta_0)>\chi^2_v(1-\zeta),$$

where Θ_α denotes the parameter space for α, and $\chi^2_v(1-\zeta)$ denotes the $1-\zeta$ quantile of the chi-square distribution with v degrees of freedom. The asymptotic size of the projection test cannot exceed ζ irrespective of the identification of α or β or both. However, the power of the test can be very low if v is large compared to p_β.

2.4 New Method of Projection

Chaudhuri et al. (2007), Chaudhuri (2008), and Chaudhuri and Zivot (2008) proposed a new method of projection for making inferences on subsets of parameters in the presence of potentially unidentified nuisance parameters that are based on ideas presented in Robins (2004). The new method of projection requires (i) a uniform asymptotic $(1-\xi)\cdot 100\%$ confidence set, $C_\alpha(1-\xi,\beta_0)$, for α when the null hypothesis $H_0:\beta=\beta_0$ is true, and (ii) an asymptotically pivotal statistic $R(\theta)$. In most cases, as described in Table 1, $R(\theta)\overset{A}{\sim}\chi^2_v$ for some v depending upon the choice of $R(\theta)$.

Then the new method of projection rejects $H_0:\beta=\beta_0$ if

(1) $C_\alpha(1-\xi,\beta_0)=\emptyset$, or

(2) $\inf_{\alpha_0\in C_\alpha(1-\xi,\beta_0)} R(\alpha_0,\beta_0)>\chi^2_v(1-\zeta)$.

Table 1. Confidence sets, test statistics, and degrees of freedom for new-projection-type tests

$C_\alpha(1-\xi,\beta_0)$	$R(\alpha,\beta)$	v
$C^K_\alpha(1-\xi,\beta_0)$	$S(\alpha,\beta_0)$	k
$C^K_\alpha(1-\xi,\beta_0)$	$K(\alpha,\beta_0)$	p
$C^{K_\alpha}_\alpha(1-\xi,\beta_0)$	$K_{\beta.\alpha}(\alpha,\beta_0)$	p_β
$C^S_\alpha(1-\xi,\beta_0)$	$S(\alpha,\beta_0)$	k
$C^S_\alpha(1-\xi,\beta_0)$	$K(\alpha,\beta_0)$	p
$C^S_\alpha(1-\xi,\beta_0)$	$K_{\beta.\alpha}(\alpha,\beta_0)$	p_β

Under the null hypothesis $H_0:\beta=\beta_0$, $C_\alpha(1-\xi,\beta_0)$ asymptotically contains α with probability at least $1-\xi$, and hence it follows from Bonferroni's inequality that the asymptotic size of the new projection type test cannot exceed $\zeta+\xi$. The new method of projection can be expected to be generally less conservative than the usual method of projection because the infimum for the new method is only computed over $C_\alpha(1-\xi,\beta_0)$, whereas the infimum is computed over the whole space Θ_α for the usual method. Similar projection methods have also been employed by Dufour (1990), Berger and Boos (1994), and Silvapulle (1996).

To implement the new method of projection in the context of GMM, $C_\alpha(1-\xi,\beta_0)$ can be constructed by inverting the S- or K-tests as

$$C^S_\alpha(1-\xi,\beta_0)=\{\alpha:S(\alpha,\beta_0)\le\chi^2_k(1-\xi)\}\qquad\text{or}$$
$$C^K_\alpha(1-\xi,\beta_0)=\{\alpha:K(\alpha,\beta_0)\le\chi^2_p(1-\xi)\}.$$

An advantage of using $C^K_\alpha(1-\xi,\beta_0)$ is that it will never be empty, and the asymptotic properties of the test will only depend on $R(\theta)$ when α is well identified. However, it will also include saddlepoints α^* where $K(\alpha^*,\beta_0)=0$ and these points are associated with spurious declines in power of the K-statistic. In contrast, the set $C^S_\alpha(1-\xi,\beta_0)$ can be empty and this will occur for values β_0 at which the overidentifying restrictions are rejected (at level ξ). As we show in the next section, this can lead to improved power properties of the new method of projection.

While the new method of projection can be implemented using any asymptotically pivotal statistic $R(\theta)$, Robins (2004) showed that there are certain advantages of using an efficient score-type statistic for $R(\theta)$. The efficient score for β (given α), in the terminology of van der Vaart (1998), is the part of the score (gradient of the objective function with respect to) for β that is orthogonal to the score for α. The efficient score statistic for β is a quadratic form in the efficient score for β with respect to an estimator of its asymptotic variance. In the context of GMM, Chaudhuri (2008) and Chaudhuri and Zivot (2008) decomposed the K-statistic (2) into two orthogonal statistics: a K-statistic for α (given β known) and an efficient (score) K-statistic for β

$$K(\theta)=K_\alpha(\theta)+K_{\beta.\alpha}(\theta),$$

where

$$K_\alpha(\theta) = \frac{1}{4}(\nabla_\alpha Q(\theta))\big(\hat{D}_{T\alpha}(\theta)'\hat{V}_{ff}(\theta)^{-1/2}\hat{D}_{T\alpha}(\theta)\big)^{-1} \times (\nabla_\alpha Q(\theta))',$$

$$K_{\beta.\alpha}(\theta) = \frac{1}{4}(\nabla_{\beta.\alpha} Q(\theta))\big(\hat{D}_{T\beta}(\theta)'\hat{V}_{ff}(\theta)^{-1/2}N_{\hat{V}_{ff}(\theta)^{-1/2\prime}\hat{D}_{T\alpha}(\theta)} \times \hat{V}_{ff}(\theta)^{-1/2}\hat{D}_{T\beta}(\theta)\big)^{-1}(\nabla_{\beta.\alpha} Q(\theta))',$$

and $\nabla_{\beta.\alpha}Q(\theta)$ is the estimated efficient score for β defined as

$$\nabla_{\beta.\alpha}Q(\theta) = f_T(\theta)'\hat{V}_{ff}(\theta)^{-1/2} \times N_{\hat{V}_{ff}(\theta)^{-1/2\prime}\hat{D}_{T\alpha}(\theta)}\hat{V}_{ff}(\theta)^{-1/2}\hat{D}_{T\beta}(\theta).$$

The above expressions use the partition $\hat{D}_T(\theta) = [\hat{D}_{T\alpha}(\theta), \hat{D}_{T\beta}(\theta)]$ and $\hat{V}_{\theta f} = [\hat{V}_{\alpha f}(\theta)', \hat{V}_{\beta f}(\theta)']'$.

It can be shown that under $H_0: \theta = \theta_0$, $K_\alpha(\theta_0) \overset{A}{\sim} \chi^2_{p_\alpha}$ and $K_{\beta.\alpha}(\theta_0) \overset{A}{\sim} \chi^2_{p_\beta}$. Furthermore, if θ_0 belongs to the $\sqrt{n}$-neighborhood of θ, then $K_{\beta.\alpha}(\theta_0) = K_{\beta.\alpha}(\theta) + o_p(1)$. This latter property of $K_{\beta.\alpha}(\theta)$ makes it ideally suited for use in the new method of projection. Indeed, Chaudhuri (2008) proved that if $C_\alpha(1-\xi,\beta_0)$ is nonempty with probability approaching one and if α is well identified then the new-method-of-projection-type test that rejects $H_0: \beta = \beta_0$ when $\inf_{\alpha_0 \in C_\alpha(1-\xi,\beta_0)} K_{\beta.\alpha}(\theta_0) > \chi^2_{p_\beta}(1-\zeta)$ is asymptotically equivalent to the size (at most) ζ K-test for β against local alternatives. This means that the new method of projection with $R(\theta) = K_{\beta.\alpha}(\theta)$ is size controlled when α is not identified and can be made asymptotically equivalent to Kleibergen's K-test when α is well identified.

Table 1 summarizes the possible ways of implementing the new-method-of-projection-type tests for testing $H_0: \beta = \beta_0$. KM illustrated the use of the new method of projection with $C_\alpha(1-\xi,\beta_0) = C^K_\alpha(1-\xi,\beta_0)$ and $R(\theta) = S(\alpha,\beta_0)$ and concluded that the Robins test, proposed in Chaudhuri (2008) and Chaudhuri et al. (2007), does not outperform the usual method of projection based on $R(\theta) = S(\alpha,\beta_0)$. However, this is not what Chaudhuri (2008) and Chaudhuri et al. (2007) refer to as the Robins test. In the context of GMM, Chaudhuri (2008) and Chaudhuri and Zivot (2008) recommend using $C_\alpha(1-\xi,\beta) = C^S_\alpha(1-\xi,\beta)$ and $R(\theta) = K_{\beta.\alpha}(\theta)$. The power of this method is largely driven by the choice of the statistic $R(\theta)$. In addition, the choice $R(\theta) = K_{\beta.\alpha}(\theta)$ (i.e., the efficient K-statistic) can make this test asymptotically equivalent the K-test when α is well identified. In the next section we show, using the same simulation experiment as KM, that this latter implementation of the new method of projection performs comparably to the tests recommended by KM.

3. SIMULATIONS

To illustrate the finite sample properties of the new method of projection based on $C^S_\alpha(1-\xi,\beta_0)$ and $K_{\beta.\alpha}(\alpha,\beta_0)$ we utilize the same simulation experiment described in Section 4 of KM. We are grateful to Frank Kleibergen and Sophocles Mavroeidis for sharing their Matlab code with us.

The data-generating process is

$$\pi_t = \lambda x_t + \gamma_f E_t[\pi_{t+1}] + u_t,$$
$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + v_t,$$
$$\pi_{t+1} = (\alpha_0\rho_1 + \alpha_1)x_t + \alpha_0\rho_2 x_{t-1} + \eta_{t+1},$$

where $\eta_t = u_t + \alpha_0 v_t$. There error terms η_t and v_t are jointly normal with unit variances and correlation $\rho_{\eta v} = 0.2$. The parameter of interest is γ_f and λ is the nuisance parameter. Identification of the structural parameters λ and γ_f is controlled by the concentration parameter μ^2, which is a complicated nonlinear function of the model parameters.

KM's Figure 3 illustrates the power curves for testing $H_0: \gamma_f = 1/2$ against $H_1: \gamma_f \neq 1/2$ at the 5% level for the subset S, usual method of projection based on S, and the new method of projection based on $C^K_\lambda(1-\xi,\gamma_f = 1/2)$ and $S(\lambda,\gamma_f = 1/2)$ with $\xi = 0.02$ and $\zeta = 0.03$. The figure shows that the power curves of the usual method of projection and an inefficient application of the new method are indistinguishable and are dominated by the subset S-statistic.

Figure 1 in this note shows the power curves of the new method of projection based on $C^S_\lambda(1-\xi,\gamma_f = 1/2)$ and $K_{\lambda.\gamma_f}(\lambda,\gamma_f = 1/2)$ with $\xi = 0.005$ and $\zeta = 0.045, 0.05$, along with the recommended tests of KM. The graphs show that the new method of projection actually performs as well as the MQLR and KJ tests recommended by KM. For the strong identification case, use of $C^S_\lambda(1-\xi,\gamma_f = 1/2)$ avoids the spurious decline in power observed for the KLM statistic.

4. CONCLUSION

KM show that the subset versions of the S, K, and MQLR statistics are valid tests even when the nuisance parameters are unidentified. This is an important theoretical and practical result. Their simulation results calibrated to a stylized new Keynesian Phillips curve show that projection-type tests are too conservative and are dominated by the subset S, K, and MQLR statistics. We show that a version of the Robins test, which we call the new method of projection, based on an efficient score-type statistic performs nearly as well as the MQLR statistic and provides an alternative approach to weak instrument robust inference for subsets of parameters in models estimated by GMM.

A practical drawback of the weak instrument robust tests is that they are based on the CU-GMM objective function. The CU-objective function can be ill-behaved, even for linear models, and finding the global minimum can be difficult. Moreover, most commercial software implementations of GMM do not support CU-GMM. Until commonly used software implementations of GMM catch up with the important theoretical developments surveyed by KM, it is not likely that weak instrument robust methods will be widely used in practice.

ACKNOWLEDGMENTS

Professor Zivot gratefully acknowledges the support of the Gary Waterman Distinguished Scholar Fund and the University

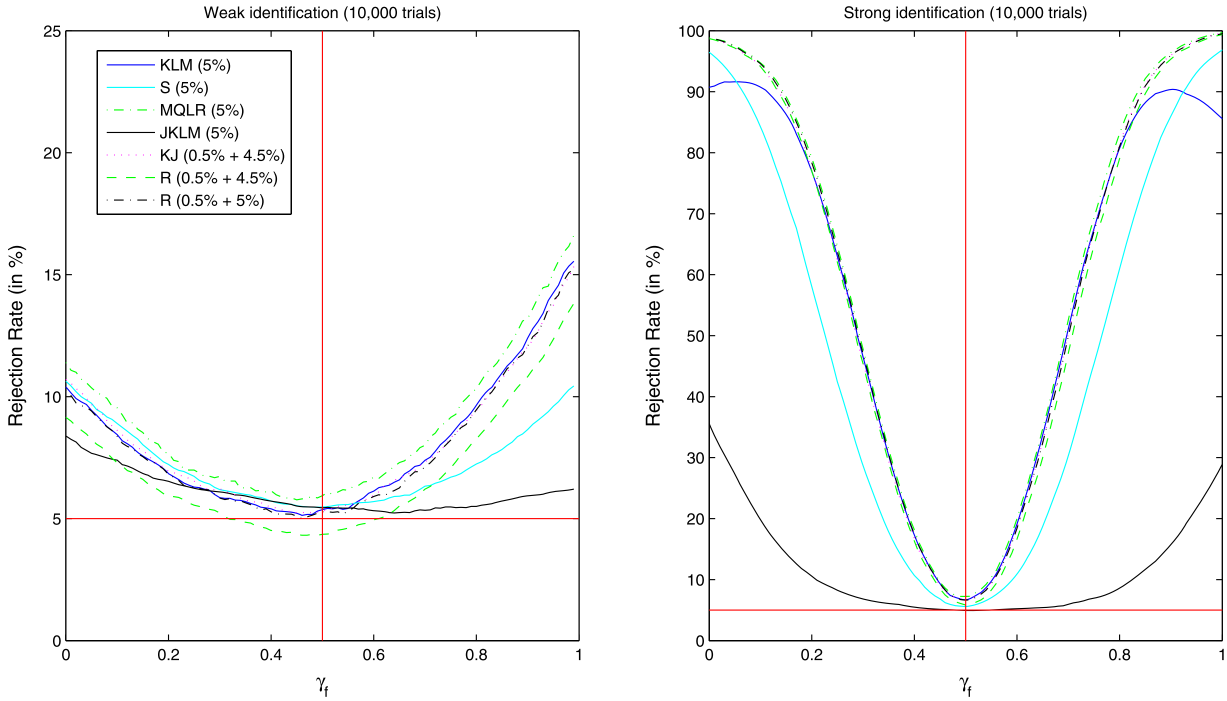

Figure 1. Power curves of 5% level tests for $H_0: \gamma_f = 0.5$ against $H_1: \gamma_f \neq 0.5$. The sample size is 1,000 and the number of Monte Carlo simulations is 10,000.

of Washington Center for Statistics in the Social Sciences Seed Grant.

ADDITIONAL REFERENCES

Berger, R. L., and Boos, D. D. (1994), "P Values Maximized Over a Confidence Set for the Nuisance Parameter," *Journal of the American Statistical Association*, 89, 1012–1016.

Chaudhuri, S. (2008), "Projection-Type Score Tests for Subsets of Parameters," Ph.D. thesis, University of Washington, Dept. of Economics.

Chaudhuri, S., and Zivot, E. (2008), "A New Method of Projection-Based Inference in GMM With Weakly Indentified Nuisance Parameters," unpublished manuscript, University of North Carolina, Dept. of Economics.

Chaudhuri, S., Richardson, T. S., Robins, J., and Zivot, E. (2007), "A New Projection-Type Split-Sample Score Test in Linear Instrumental Variables Regression," unpublished manuscript, University of Washington, Dept. of Economics.

Dufour, J. M. (1990), "Exact Tests and Confidence Sets in Linear Regressions With Autocorrelated Errors," *Econometrica*, 58, 475–494.

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

Silvapulle, M. J. (1996), "A Test in the Presence of Nuisance Parameters," *Journal of the American Statistical Association*, 91, 1690–1693.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge, U.K.: Cambridge University Press.

Rejoinder

Frank Kleibergen and Sophocles Mavroeidis
Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912
(*Frank_Kleibergen@brown.edu*; *Sophocles_Mavroeidis@brown.edu*)

1. INTRODUCTION

We would like to thank Fabio Canova, Saraswata Chaudhuri, John Chao, Jean-Marie Dufour, Anna Mikusheva, Norman Swanson, Jonathan Wright, Moto Yogo, and Eric Zivot for their stimulating discussions of our article. We especially like the diversity of the different discussions, which caused them to have hardly any overlap while all of them provide insightful comments from the discussant's own research perspective. Because of the small overlap of the different discussions, we comment on them separately and do so in alphabetical order.

2. CANOVA

In his discussion, Canova brings up the issue of the structural versus semistructural specifications of the model. Indeed, we received similar comments when we presented the article at the Joint Statistical Meetings (JSM) conference in Denver, so we revised the article somewhat to give results for a particular structural specification proposed by Galí and Gertler (1999).

Journal of Business & Economic Statistics
July 2009, Vol. 27, No. 3
DOI: 10.1198/jbes.2009.08350

Canova points out that not all of the parameters are identifiable in this structural specification, and this can only be achieved by a system approach. Since we did not discuss this, Canova's discussion complements our article in an important way.

3. CHAO AND SWANSON

In their discussion of our article Chao and Swanson make two points. First, they argue that the new Keynesian Phillips curve (NKPC) might not be weakly identified. Then, partly on the basis of their first argument, they advocate the use of many instrument robust procedures for inference on the parameters of the NKPC. These two arguments are related because the many instrument robust methods require that the parameters are well identified. We discuss these two arguments in turn.

3.1 Is the NKPC Weakly Identified?

In our article, we explain that the coefficients of the NKPC are not identified when $\lambda = 0$. For analytical tractability, our identification analysis was based on the pure forward-looking version of the NKPC but the result also applies to the hybrid version [$\gamma_b \neq 0$ in Equation (2) in the article]; see, for example, Mavroeidis (2006). Chao and Swanson show that when the restriction $\gamma_f + \gamma_b = 1$ is imposed so that the number of unknown parameters is reduced by one, the condition $\lambda \neq 0$ is no longer necessary for identification provided that $\gamma_f > 1/2$. This result was also shown in Mavroeidis (2002, section 4.2), who, in addition, showed that the restricted hybrid model is unidentified when $\lambda = 0$ and $\gamma_f \leq 1/2$. Furthermore, if x_t is not Granger-caused by π_t in the example of Chao and Swanson, as seems to be the case with U.S. data, it can be shown that the restricted hybrid model is unidentified when $\gamma_f \leq 1/2$ for any value of λ, see Mavroeidis (2002, section 4.2). However, since one of the subset 95% confidence intervals that we report has $\gamma_f > 1/2$, Chao and Swanson argue that whether the NKPC is weakly identified or not needs further examination. Incidentally, note that in the example of Chao and Swanson, where π_t Granger-causes x_t, it is not clear that the model will be identified for all $\gamma_f > 1/2$ when $\lambda \neq 0$.

We want to emphasize that our conclusion that the NKPC is weakly identified is based on the size of the identification robust confidence sets of the parameters. It does not stem from any a priori identification analysis or from the estimates of λ. Since the confidence sets for the semistructural parameters λ and γ_f reported in Figure 6 are bounded, our empirical results show that the model is not completely unidentified. The confidence sets for the structural parameters α and ω reported in Figure 7 are, however, completely uninformative about the degree of backward-looking behavior, ω. They are partly informative about α. Since $\alpha = 1$ is in the 90% confidence set, the 90% confidence set for the average duration over which prices remain fixed, which is measured by $(1 - \alpha)^{-1}$, is unbounded. We believe that these results are sufficient to conclude that the NKPC is weakly identified.

Chao and Swanson argue that it is useful to have a measure of identification to assess the reliability of the point estimates of the parameters. The identification-robust confidence sets, however, already provide such a measure of identification. In case of unbounded confidence sets, the model is unidentified and we show that this result cannot be attributed to a lack of power of these identification-robust procedures relative to nonrobust methods. Moreover, Equation (26) in the article shows that the S-statistic, which is the generalization of the Anderson–Rubin (AR) statistic to the generalized method of moments (GMM), resembles a test for the rank of the Jacobian at distant values of the parameter of interest. Since the parameters are identified when the Jacobian has a full rank value, this shows that bounded S-based confidence sets provide evidence that the model is not completely unidentified. Hence, statistical inference can be conducted directly and there is no need for pretesting, even if such pretesting were possible (which is currently not the case for models like the NKPC).

The a priori identification analysis we conducted in our article intended to highlight the dependence of the strength of identification on the value of the structural parameters in the NKPC. The strength of identification in the linear instrumental variables regression model is measured by the concentration parameter and does not depend on the structural parameters. We construct the concentration parameter for the NKPC and show that it depends in a complicated way on the structural parameters. This dependence makes it difficult to extend methods that have been proposed to gauge the strength of the instruments in the linear instrumental variable regression model (e.g., Stock and Yogo 2003 and Hansen, Hausman, and Newey 2008) to the NKPC.

Finally, we numerically evaluate the concentration parameter for the restricted hybrid NKPC model over the admissible range of the structural parameters $\alpha, \omega \in [0, 1]$. Recall that, under the restriction $\beta = 1$, Equations (33) and (34) in our article become: $\gamma_f = \alpha/(\alpha + \omega)$ and $\lambda = (1 - \omega)(1 - \alpha)^2/(\alpha + \omega)$. Figure 1 reports the contours of the concentration parameter

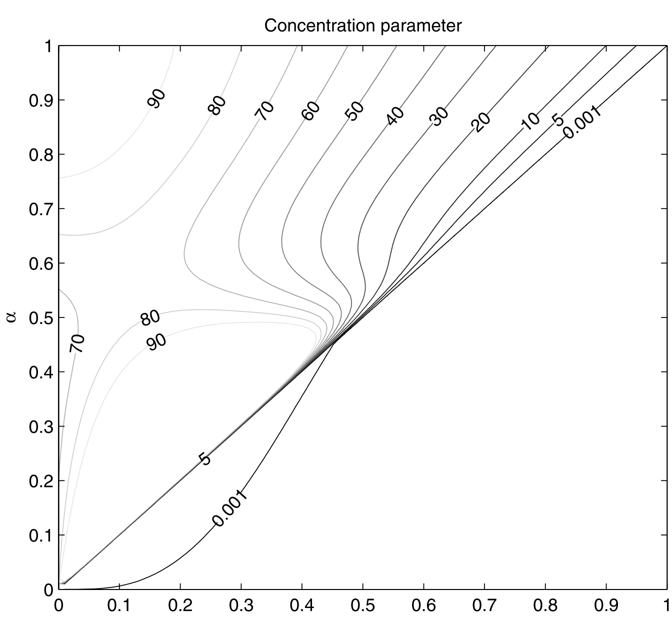

Figure 1. Approximate contours of the concentration parameter in structural NKPC with $\beta = 1$, conditional on $x_t \sim$ AR(2). Reduced form coefficients and error variances are estimated using U.S. data over our estimation sample. The sample size for the concentration parameter is assumed to be 100.

when x_t follows an autoregressive model of order 2 and all nuisance parameters are calibrated to U.S. data. Upon comparison with the 90% confidence sets reported in Figure 7 of our article, we see that values of the concentration parameter of 10^{-3} lie within the 90% confidence set. Hence, we cannot rule out that weak identification is of empirical relevance even in the example discussed by Chao and Swanson.

3.2 Many Instruments

Chao and Swanson observe that in the NKPC and most macro models the unconditional moments from which the parameters are estimated result from conditional moment restrictions. This typically implies a large number of candidate instrumental variables. Therefore, they argue that methods derived in the so-called many instrument literature could be relevant for the NKPC.

We agree that if one is willing to use more than a handful of instruments, it is important to use methods that are more robust to the number of instruments than the conventional two-step GMM procedure. The subset Kleibergen Lagrange multiplier (KLM) and MQLR statistics that we use in our empirical analysis are examples of such procedures. However, there are a few reasons that make us rather cautious in recommending the use of many moment conditions for estimating the NKPC and other related macro models.

First, it is well known that the many instrument robust procedures referred to by Chao and Swanson are not robust to weak identification. Our previous discussion therefore demonstrates that we do not have sufficient evidence to argue that the NKPC is sufficiently well identified for these many instrument robust methods to perform adequately.

Second, in forward-looking models the relevance of an instrument is directly linked to its predictive content for the future value of the endogenous regressor. In the case of the NKPC, instruments carry incremental information insofar as they help to predict future inflation beyond the first few lags of inflation and the labor share. However, there is a large literature documenting that it is rather hard to beat random walk forecasts of inflation; see Stock and Watson (2008), which leads to a relatively small number of relevant instruments.

Third, another important reason to use a limited number of instruments, which is not related to the issue of identification, is the estimation of the long-run variance of the moments. It is well known that nonparametric spectral density estimators of the long-run variance of dependent variables, such as the Newey–West estimator (see Newey and West 1987), cause significant size distortions to GMM tests; see, for example, Sun, Phillips, and Jin (2008). Neither the identification robust nor the many instrument robust procedures are immune to this problem. We have conducted several simulations of the size of GMM tests, including the ones we recommend, with different HAC estimators and found that the size can increase dramatically with the number of instruments.

For the aforementioned reasons and since the KLM and MQLR statistics are as robust to many instrument sequences as the many instrument robust statistics and are also robust to weak identification of the parameters, unlike the many instrument robust statistics, we believe they are preferable for inference on the NKPC.

4. DUFOUR

Dufour's discussion primarily focuses on the effect of so-called missing instruments on the limiting distributions of the KLM and MQLR statistics. We therefore first address the effect of missing instruments and then proceed with some further discussion of projection-based testing.

4.1 Missing Instruments

Dufour refers to the simulation study conducted by Dufour and Taamouti (2007), which shows the sensitivity of the limiting distributions of the KLM and MQLR statistics to missing instruments. The simulation results reported in Table 1 of Dufour's comment are striking, but the sensitivity to "missing instruments" mentioned by Dufour does not result from the omission of relevant instruments. It results from their peculiar specification of the data generating process, which has the missing instruments orthogonal *in sample* to the included ones. This means that the missing instruments are not independent of the included instruments, and they depend on each other in a way that is unlikely to occur in practice. We show that if the missing instruments are not orthogonal to the included instruments that the size distortions of the KLM test reduce to the problem of many included instruments as discussed in Bekker and Kleibergen (2003).

To proceed, we first state the data generating process used in the simulation study by Dufour and Taamouti (2007), which is identical to the one in Dufour's comment. We then show that the size distortions result from the orthogonality of the missing instruments to the included ones. We explain that this size distortion arises mainly from the estimation of the covariance of the errors, which is used in the computation of the KLM and MQLR statistics; see Kleibergen (2002). Since these covariance estimators are not needed for the AR statistic, its distribution is not affected by the missing instruments under the null. However, we show that this orthogonality assumption adversely affects the power of the AR statistic, which becomes biased, and has a similar effect on the first-stage F-statistic for the coefficients of the included instruments.

The simulation study of Dufour and Taamouti (2007) uses the model

$$\begin{aligned} y_1 &= Y_1\beta_1 + Y_2\beta_2 + u, \\ (Y_1 \vdots Y_2) &= X_2\Pi_2 + X_3\delta + (V_1 \vdots V_2), \end{aligned} \tag{1}$$

with y_1, Y_1, Y_2, $X_3 : T \times 1$, $X_2 : T \times k_2$, and

$$\begin{aligned} &\operatorname{vec}(u_t \vdots V_{1t} \vdots V_{2t}) \overset{\text{iid}}{\sim} \mathrm{N}(0, \Sigma), \\ &\Sigma = \begin{pmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.3 \\ 0.8 & 0.3 & 1 \end{pmatrix}. \end{aligned} \tag{2}$$

The $T \times k_2$ matrix X_2 contains the included instruments while the $T \times 1$ vector X_3 is a vector of omitted instruments. Dufour and Taamouti (2007) specify the vector of missing instruments such that it is orthogonal to X_3, $X_3'X_2 \equiv 0$. They simulate the elements of X_2 and X_3 as iid $\mathrm{N}(0, 1)$ random variables and orthogonalize X_3 with respect to X_2. Dufour and

Taamouti (2007) keep the instruments fixed in repeated samples, but this is inessential for the results. In his discussion, Dufour uses random instruments, so we do the same in our simulations here. The results are qualitatively the same with fixed X. Dufour and Taamouti (2007) set the parameter values at $\beta_1 = \frac{1}{2}$, $\beta_2 = 1$, and $\delta = \lambda(1\vdots 1)$. Their parameter matrix Π_2 is such that $\Pi_2 = \rho\Pi/\sqrt{T}$, where ρ is 0.01 or 1, with the elements on the main diagonal of the $k_2 \times 2$ matrix Π equal to one and the remaining off-diagonal elements are all equal to zero. This implies that the concentration matrices, which are equal to $\Pi_2' X_2 X_2 \Pi_2$, are roughly equal to $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0.001 & 0 \\ 0 & 0.001 \end{pmatrix}$, which shows that the included instruments are basically weak ($\rho = 1$) or irrelevant ($\rho = 0.01$). The sample size T is equal to 100 and the number of instruments k_2 varies between 2 and 40.

To show the implications of the orthogonality of X_2 and X_3, we consider the score vector with respect to (β_1, β_2) of the linear instrumental variables regression model at the true value of β_1 and β_2, which is proportional to (see Kleibergen 2002):

$$\begin{aligned} u'Z\tilde{\Pi}_2(\beta_1,\beta_2) &= u'P_{X_2}(Y_1\vdots Y_2) - u'M_{X_2}(Y_1\vdots Y_2)\frac{u'P_{X_2}u}{u'M_{X_2}u} \\ &= u'P_{X_2}(X_2\Pi_2 + X_3\delta + (V_1\vdots V_2)) \\ &\quad - u'M_{X_2}(X_2\Pi_2 + X_3\delta + (V_1\vdots V_2))\frac{u'P_{X_2}u}{u'M_{X_2}u} \\ &= u'P_{X_2}\Pi_2 + u'P_{X_2}(X_3\delta + (V_1\vdots V_2)) \\ &\quad - u'M_{X_2}(X_3\delta + (V_1\vdots V_2))\frac{u'P_{X_2}u}{u'M_{X_2}u}, \end{aligned} \tag{3}$$

where $\tilde{\Pi}_2(\beta_1,\beta_2) = (X_2'X_2)^{-1}X_2'((Y_1\vdots Y_2) - u\frac{\hat{\sigma}_{uV}}{\hat{\sigma}_{uu}})$, $u = y_1 - Y_1\beta_1 - Y_2\beta_2$, $\hat{\sigma}_{uV} = \frac{1}{T-k_2}u'M_{X_2}(Y_1\vdots Y_2)$, $\hat{\sigma}_{uu} = \frac{1}{T-k_2}u'M_{X_2}u$, and $P_{X_2} = X_2(X_2'X_2)^{-1}X_2'$, $M_{X_2} = I_T - P_{X_2}$. The KLM statistic is a quadratic form of the score vector in (3). Hence, any peculiar behavior of the KLM statistic is directly related to the score vector. If X_2 and X_3 are orthogonal, so $P_{X_2}X_3 = 0$, the score vector in (3) simplifies to

$$\begin{aligned} u'Z'\tilde{\Pi}_2(\beta_1,\beta_2) &= u'P_{X_2}\Pi_2 + u'P_{X_2}(V_1\vdots V_2) \\ &\quad - u'M_{X_2}(X_3\delta + (V_1\vdots V_2))\frac{u'P_{X_2}u}{u'M_{X_2}u}. \end{aligned} \tag{4}$$

It is important to note that except for the last element of the above expression, the missing instruments have dropped out of all the other elements.

When X_3 is missing, it becomes part of the residual in the first-stage regression, and the key idea behind the score vector is to orthogonalize the first-stage residual with respect to the structural error u. Now, when u and X_3 are independent in the population, there is always some nonzero sample correlation between u and X_3. The expression of the score vector in (4) shows that the orthogonality of X_2 and X_3 prevents the sample correlation between u and X_3 from affecting the first element of the score vector. The sample correlation between u and X_3 still affects the second part of the score vector. Hence, because of the orthogonality of X_2 and X_3, u and $[(Y_1\vdots Y_2) - u\frac{\hat{\sigma}_{uV}}{\hat{\sigma}_{uu}}]$ exhibit sample correlation although they are uncorrelated in the population. One can therefore obtain any kind of size distortion by appropriately specifying δ or λ.

Another consequence of the orthogonality of X_2 and X_3 is that the AR test is not unbiased because its power can be lower than the level of the test under the alternative, and when the included instruments are sufficiently weak, the power of the AR test is actually maximized under the null. To see this, note that the specification of the AR statistic that tests $H_0\colon \beta = (\beta_{1,0}, \beta_{2,0})'$ when the true value is $\beta = (\beta_1, \beta_2)'$ reads

$$\begin{aligned} &\text{AR}(\beta_{1,0},\beta_{2,0}) \\ &= \left(y_1 - (Y_1:Y_2)\begin{pmatrix}\beta_{1,0}\\ \beta_{2,0}\end{pmatrix}\right)' \\ &\quad \times P_{X_2}\left(y_1 - (Y_1:Y_2)\begin{pmatrix}\beta_{1,0}\\ \beta_{2,0}\end{pmatrix}\right)\Big/k_2 \\ &\quad \Big/\left(\left(y_1 - (Y_1:Y_2)\begin{pmatrix}\beta_{1,0}\\ \beta_{2,0}\end{pmatrix}\right)' \right.\\ &\quad \left.\times M_{X_2}\left(y_1 - (Y_1:Y_2)\begin{pmatrix}\beta_{1,0}\\ \beta_{2,0}\end{pmatrix}\right)\Big/(T-k_2)\right) \\ &= \left(u + [X_2\Pi_2 + (V_1:V_2)]\begin{pmatrix}\beta_1-\beta_{1,0}\\ \beta_2-\beta_{2,0}\end{pmatrix}\right)' \\ &\quad \times P_{X_2}\left(u + [X_2\Pi_2 + (V_1:V_2)]\begin{pmatrix}\beta_1-\beta_{1,0}\\ \beta_2-\beta_{2,0}\end{pmatrix}\right)\Big/k_2 \\ &\quad \Big/\left(\left(u + [X_3\delta + (V_1:V_2)]\begin{pmatrix}\beta_1-\beta_{1,0}\\ \beta_2-\beta_{2,0}\end{pmatrix}\right)' \right.\\ &\quad \times M_{X_2}\left(u + [X_3\delta + (V_1:V_2)]\begin{pmatrix}\beta_1-\beta_{1,0}\\ \beta_2-\beta_{2,0}\end{pmatrix}\right) \\ &\quad \left.\Big/(T-k_2)\right), \end{aligned} \tag{5}$$

where we used the fact that

$$\begin{aligned} &y_1 - (Y_1\vdots Y_2)\begin{pmatrix}\beta_{1,0}\\ \beta_{2,0}\end{pmatrix} \\ &\qquad = u + [X_2\Pi_2 + X_3\delta + (V_1\vdots V_2)]\begin{pmatrix}\beta_1-\beta_{1,0}\\ \beta_2-\beta_{2,0}\end{pmatrix} \end{aligned} \tag{6}$$

and that, because of the orthogonality of X_2 and X_3, X_3 drops out of the numerator. When δ is different from zero, the estimate of the variance in the denominator is inflated, and this unambiguously reduces power. In the limiting case with $\Pi_2 = 0$, the power of the AR test is equal to its level when $\delta = 0$, and it must therefore be less than the level for $\delta > 0$. This is shown in the simulations reported below based on Dufour's example where the included instruments are very weak. This also implies that for a given nonzero value of δ the rejection frequency of the AR statistic is maximized at the true value of β. Hence, the orthogonality of X_2 and X_3 leads to a biased AR test, which leads to the strange outcome that the true value is the most objectionable value.

The odd behavior of the AR statistic when X_2 and X_3 are orthogonal extends to similar nonstandard behavior of other statistics, such as the first-stage Wald statistic that tests whether Π_2 is equal to its true value. Because the covariance matrix estimator is the only element of this Wald statistic that depends on δ, it overstates the variance so the Wald test will be undersized.

The above shows that the orthogonality of the missing instrument X_3 compared to the included instrument X_2 explains the strange behavior of the different statistics, including the AR statistic, whose power function is maximized under the null when identification is weak. If X_2 and X_3 are not orthogonal, the effect of missing instruments on the limiting distributions of the KLM and MQLR statistics is identical to the effect of many instruments on the limiting distribution of the KLM statistic, which is discussed in Bekker and Kleibergen (2003). This is because missing relevant instruments essentially blow up the variance of the error term in the first-stage regression and therefore reduce the concentration parameter. When the number of included instruments is also large, this leads to the many weak instruments problem discussed in Bekker and Kleibergen (2003).

To show this, we consider the estimator of Π_2, $\tilde{\Pi}_2(\beta_1, \beta_2)$, that is used for the score vector in (3):

$$\begin{aligned}\tilde{\Pi}_2(\beta_1, \beta_2) &= (X_2'X_2)^{-1}X_2'\left((Y_1 \vdots Y_2) - u\frac{\hat{\sigma}_{uV}}{\hat{\sigma}_{uu}}\right) \\ &= (X_2'X_2)^{-1}X_2'\left((Y_1 \vdots Y_2) - u\frac{\sigma_{uV}}{\sigma_{uu}}\right) \\ &\quad + (X_2'X_2)^{-1}X_2'u\left(\frac{\sigma_{uV}}{\sigma_{uu}} - \frac{\hat{\sigma}_{uV}}{\hat{\sigma}_{uu}}\right),\end{aligned} \tag{7}$$

with $\sigma_{uV} = E(\lim_{T\to\infty} \frac{1}{T}u'((V_1 \vdots V_2))$ and $\sigma_{uu} = E(\lim_{T\to\infty} \frac{1}{T} \times u'u)$. The expression of $\tilde{\Pi}_2(\beta_1, \beta_2)$ (7) shows that when we use the true values of the covariance parameters σ_{uV} and σ_{uu}, there is no sensitivity to missing instruments. Thus the sensitivity to missing instruments results from the last part of the expression in (7). We therefore analyze the behavior of the covariance estimator $\hat{\sigma}_{uV}$:

$$\begin{aligned}\hat{\sigma}_{uV} &= \frac{1}{T-k_2}u'M_{X_2}(X_3\delta + (V_1 \vdots V_2)) \\ &= \frac{1}{T-k_2}u'M_{X_2}X_3\delta + \frac{1}{T-k_2}u'M_{X_2}(V_1 \vdots V_2) \\ &\approx \sigma_{uV} + \frac{1}{\sqrt{T-k_2}}\psi_{uV} + \frac{1}{\sqrt{T-k_2}}\psi_{uX_3}\delta.\end{aligned} \tag{8}$$

The convergence rates of ψ_{uV} and ψ_{uX_3} in the expression for $\hat{\sigma}_{uV}$ result from specifying M_{X_2} as $P_{X_{2,\perp}}$, with $X_{2,\perp} : T \times (T - k_2)$ and $X_2'X_{2,\perp} \equiv 0$, and we also used the fact that

$$\begin{aligned}\sqrt{T-k_2}\left(\frac{1}{T-k_2}u'M_{X_2}(V_1 \vdots V_2) - \sigma_{uV}\right) &\xrightarrow{d} \psi_{uV}, \\ \frac{1}{\sqrt{T-k_2}}u'M_{X_2}X_3 &\xrightarrow{d} \psi_{uX_3},\end{aligned} \tag{9}$$

where ψ_{uV} and ψ_{uX_3} are independently normally distributed with mean zero.

The expression of the covariance estimator in (8) shows that the sensitivity to missing instruments results from the higher-order elements, $\frac{1}{\sqrt{T-k_2}}\psi_{uV}$ and $\frac{1}{\sqrt{T-k_2}}\psi_{uX_3}\delta$. The convergence rate of these elements depends on the number of instruments. Bekker and Kleibergen (2003) show that the limiting distribution of the KLM statistic when the number of instruments is proportional to the sample size is bounded between a $\chi^2(m)$ and a $\frac{\chi^2(m)}{1-\alpha}$ distribution with m the number of structural parameters and $\alpha = \lim_{k_2,T\to\infty} \frac{k_2}{T}$. The $\chi^2(m)$ lower bound is attained when the structural parameters are well identified and the $\frac{\chi^2(m)}{1-\alpha}$ upper bound is attained when the structural parameters are completely unidentified.

To illustrate the above remarks, we revisit the simulation exercise of Dufour and Taamouti (2007). Table 1 shows the rejection frequencies of the KLM test using the same data generating process as in Dufour's comment. We only show the results for the KLM statistic since the results for the MQLR statistic are similar. Table 1 shows rejection frequencies for the KLM test under four different scenarios. The first scenario, listed as "Orth," matches the results reported in Dufour's Table 1 under the heading "K" and has the missing instrument orthogonal to the included ones. The second scenario, listed as "Known cov," uses the true values of the covariance parameters σ_{uV} and σ_{uu} in the computation of the KLM statistic. The third scenario, listed as "Nonorth," does not force the missing instrument to be orthogonal to the included ones. The fourth scenario, listed as "Upper," uses nonorthogonal missing instruments and critical values from the $\frac{\chi^2(2)}{1-\alpha}$ upper bound on the limiting distribution obtained by Bekker and Kleibergen (2003).

It is clear from the results that the large size distortions occur only in the case of orthogonalized missing instruments, and the size distortions get larger as the relevance of the missing instrument increases (the case $\lambda = 10$ means that the concentration parameter associated with the missing instrument is 20,000). Table 1 shows that there is no size distortion when we use the known values of the covariance parameters σ_{uV} and σ_{uu}, in accordance with our above discussion. When we use missing instruments that are not orthogonal to the included ones, Table 1 shows that the size distortions do not depend on λ (i.e., the strength of the missing instruments) and only depend on the number of included instruments k_2. This confirms our previous analysis, which showed that the effect of missing instruments on the limiting distribution results basically from the use of many instruments. This is further confirmed when we use critical values that result from the upper bound on the limiting distribution obtained by Bekker and Kleibergen (2003). Since the included instruments are very weak, the limiting distribution of the KLM statistic coincides with the upper bound, which is confirmed by the rejection frequencies that coincide with the size of the test.

Next, we turn to our remarks concerning the effect of orthogonal missing instruments on the power of the AR statistic. Figure 2 shows the power of the AR statistic in (5) for testing $H_0 : \beta_1 = \frac{1}{2}$, $\beta_2 = 1$ for different values of β_1 while β_2 is fixed at its true value (when we vary β_2, the results are qualitatively the same). The power curves in Figure 2 result from the data generating process used by Dufour and Taamouti (2007) with X_2 orthogonal to X_3, which is indicated by "Orth," and the

Table 1. Rejection frequencies of KLM statistic in case of missing instruments

k_2	Orth	Known cov	Nonorth	Upper	Orth	Known cov	Nonorth	Upper
	$\lambda = 0$ and $\rho = 0.01$				$\lambda = 0$ and $\rho = 1$			
2	5.18	4.87	5.18	4.85	5.18	4.87	5.18	4.85
3	5.52	4.94	5.52	5.03	5.53	5.3	5.53	5.14
4	5.54	4.88	5.54	4.93	5.46	5	5.46	4.8
5	5.93	5.13	5.93	5.23	5.78	4.88	5.78	5.04
10	7.3	5.17	7.3	5.52	6.4	5.06	6.4	4.76
20	9.45	4.97	9.45	5.4	7.82	4.83	7.82	4.11
40	17.03	4.87	17.03	5.58	13.55	5.08	13.55	3.93
	$\lambda = 1$ and $\rho = 0.01$				$\lambda = 1$ and $\rho = 1$			
2	5.18	4.87	5.18	4.85	5.18	4.87	5.18	4.85
3	7.7	4.94	5.58	5.23	5.78	5.3	5.7	5.26
4	11.07	4.88	5.82	5.25	6.26	5	5.95	5.25
5	14.8	5.13	6.31	5.43	7.12	4.88	6.37	5.52
10	33.32	5.17	7.35	5.55	14.94	5.06	7.33	5.36
20	59.28	4.97	9.62	5.61	37.89	4.83	9.43	5.35
40	78.04	4.87	16.83	5.45	68.23	5.08	16.63	5.22
	$\lambda = 10$ and $\rho = 0.01$				$\lambda = 10$ and $\rho = 1$			
2	5.18	4.87	5.18	4.85	5.18	4.87	5.18	4.85
3	11.26	4.94	5.52	5.15	9.34	5.3	5.6	5.23
4	19.34	4.88	5.84	5.24	15.36	5	6.07	5.47
5	28.82	5.13	6.27	5.44	22.79	4.88	6.26	5.42
10	73.36	5.17	7.42	5.56	63.09	5.06	7.41	5.33
20	95.16	4.97	9.65	5.57	91.22	4.83	9.64	5.33
40	97.82	4.87	16.92	5.51	96.68	5.08	16.63	5.4

NOTE: Orth: orthogonal instruments. Known cov: known covariance. Nonorth: nonorthogonal instruments. Upper: critical values result from the upper bound on the limiting distribution.

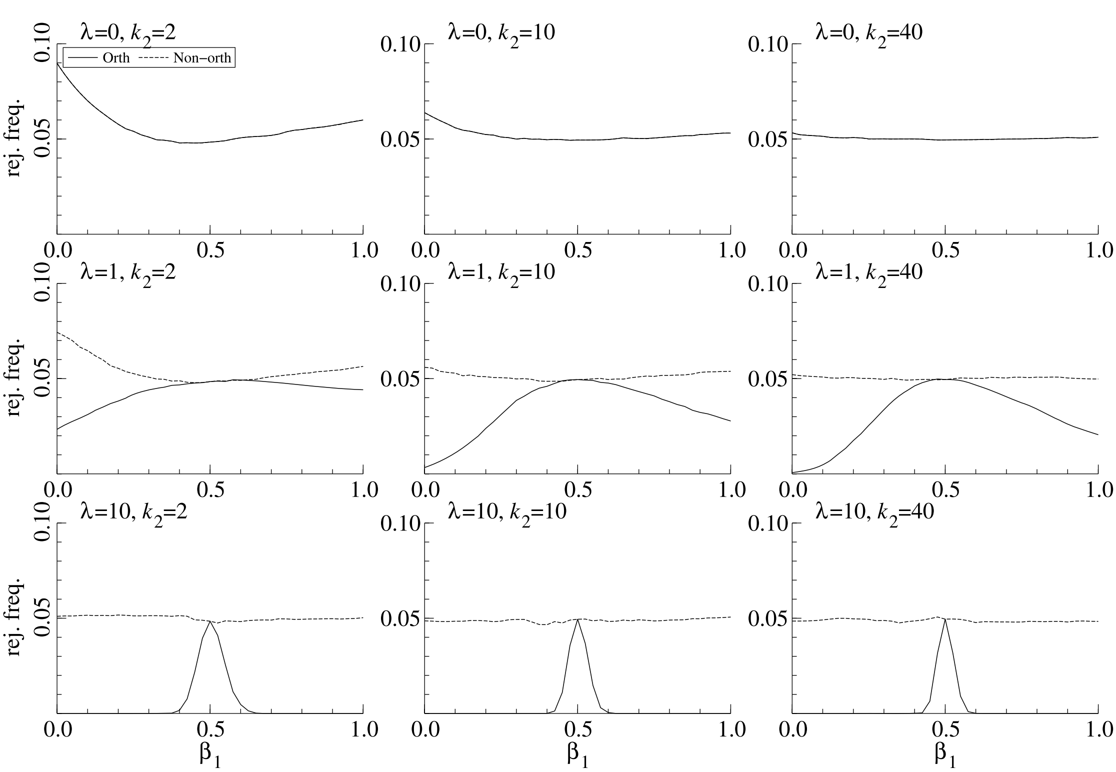

Figure 2. Power of the AR statistic for testing $H_0 : \beta_1 = 0.5$, $\beta_2 = 1$ for different values of β_1 and β_2 equal to 1. Orth: orthogonal instruments. Nonorth: nonorthogonal instruments.

same data generating process with X_2 not orthogonal to X_3, indicated by "Nonorth." The power curves in Figure 2 that result from the orthogonal and nonorthogonal specification are identical when $\lambda = 0$, which is the case of no missing instrument. When $\lambda = 1$ or 10, the power of the AR statistic is minimized at the true value in case of nonorthogonal instruments while it is maximized at the true value in case of orthogonal instruments. Figure 2 therefore shows that the odd behavior of the power curve of the AR statistic results from the orthogonality of X_2 and X_3 and not from the property that X_3 is a missing instrument.

Table 1, Figure 2, and our previous discussion show that the size distortions reported by Dufour and the peculiar behavior of the AR statistic with respect to its power result from the orthogonality of the missing instrument compared to the included ones. When the missing instruments are not orthogonal: (i) the effect of the missing instruments on the limiting distribution results from the sensitivity of the limiting distribution to many instruments and (ii) the AR test is unbiased. Instruments that are orthogonal to one another are not independently distributed. This is the reason for the peculiar behavior of the different statistics, and it makes the relevance of the simulation study of Dufour and Taamouti (2007) somewhat questionable for practical purposes. In fact, if one insists on the principle of robustness against such a type of missing variables, it is easy to construct examples in which no statistic (including the AR statistic) has correct size. Specifically, such an example can be constructed by considering the case of an exogenous variable X_3 that has been omitted from the structural equation, that is, $y = Y_1\beta_1 + Y_2\beta_2 + X_3\delta + u$. In applied work researchers rarely observe all the relevant measurable characteristics that may affect a particular outcome of interest, and these omitted variables form part of the error term. If the unobserved characteristic X_3 is uncorrelated with the included instruments, X_2, valid inference can be drawn for β_1 and β_2. However, if the omitted variable is orthogonal to X_2, then even the AR statistic will be severely size distorted. We think that this situation is as unrealistic as missing instruments in the first-stage regression that are orthogonal to the included ones.

4.2 Projection Methods

For the same statistic of interest, the subset versions that we use dominate their projection-based counterparts with respect to power and also show that the projection-based statistics are conservative. Dufour advocates the use of projection-based statistics and puts forward three arguments:

1. Use of the subset statistics relies on an asymptotic argument. We agree that one has to be careful with applying asymptotic arguments. The research on weak instrument robust statistics is largely motivated by the bad performance of such asymptotic arguments with respect to the traditionally used statistics. The reason why these asymptotic arguments do not perform very well is, however, the high-level assumption with respect to the rank of the Jacobian and not the inappropriateness of asymptotic arguments in general. For example, central limit theorems hold under mild conditions (see, e.g., White 1984) and provide the foundation for statistical inference in most econometric models. The limiting distributions of the weak instrument robust statistics only require such a central limit theorem to hold (see Assumption 1) and therefore apply quite generally. Use of such central limit theorems leads to more robust statistical inference than the one which results from assuming a specific distribution for the data and which would lead to so-called exact inference.
2. The regularity conditions for the derivative in GMM puts restrictions on the model. We agree that making assumptions about the derivative puts some restrictions on the model compared to only using the moments themselves. As usual, imposing more structure on the model trades off robustness for power. Arguably, there are cases in which the additional restrictions imposed by our assumption on the derivative are *a priori* questionable (e.g., the example studied in Mavroeidis, Chevillon, and Massmann 2008, which uses projection-based tests only). However, we think that for the application that we looked at, this assumption is mild. See also our response to Mikusheva's comment below.
3. The limiting distributions of subset statistics are sensitive to missing instruments. We have extensively addressed this point above.

5. MIKUSHEVA

Mikusheva raises two concerns about the relevance of Assumption 1 for the NKPC. Assumption 1 states that the moment conditions $\bar{f}_t(\theta)$ and their Jacobian $\bar{q}_t(\theta)$ are approximately normally distributed. Mikusheva argues that the normal approximation might be inaccurate because of (i) the high persistence in inflation and (ii) the large dimension of $\bar{q}_t(\theta)$ relative to the sample size. We discuss each of these arguments consecutively.

We agree with Mikusheva that persistence in inflation can invalidate Assumption 1. To address this issue, in our empirical analysis we used the lags of $\Delta\pi$ instead of π as instruments. Mikusheva rightly observes that this is sufficient to ensure the approximate normality of the moment conditions under the null $f_t(\theta_0)$, which will contain terms of the form $\Delta\pi_{t-i}e_t$, where e_t is MA(1), but not necessarily under the alternative, in general. This is because under the alternative, $f_t(\theta)$ also contains terms of the form $\Delta\pi_{t-i}(\gamma_f - \gamma_{f,0})\pi_{t+1}$ and $\Delta\pi_{t-i}(\gamma_b - \gamma_{b,0})\pi_{t-1}$, which will not be approximately normal when π_t is nearly integrated.

However, there are special cases in which Assumption 1 can be verified, and our empirical application with the restriction $\gamma_f + \gamma_b = 1$ is such a special case. This is because the restriction $\gamma_f + \gamma_b = 1$ makes the model linear in $\Delta_2\pi_{t+1}$ [see Equation (32) in the article], and the moment conditions become $Z_t(\Delta\pi_t - \gamma_f\Delta_2\pi_{t+1} - \lambda x_t)$. As a result, neither the moment conditions under the alternative, nor their Jacobian involve any levels of the potentially integrated series π_t, and our Assumption 1 can be verified. This argument does not apply more generally to the unrestricted specification. Indeed, this is precisely the reason why in a related article studying the NKPC with learning, Mavroeidis, Chevillon, and Massmann (2008) use only the S-statistic for the full parameter vector, for which the relevant part of Assumption 1 can be verified under the null hypothesis. Nonetheless, in view of similar results in the cointegration literature, where the estimators of the parameters of the cointegrating vector have normal limiting distributions, it is possible that we might have (mixed) asymptotic normality of the score vector while the derivatives of the moments do not converge to a normal distribution. Even though we did not need to investigate this further for our application, we believe that this issue is of considerable general interest and deserves further study.

Mikusheva's second point concerns the dimensionality of the moment and derivative vector that is involved in Assumption 1. We agree that the dimensionality of the moment and derivative vector is quite large. Although the limiting distributions of the identification robust KLM and MQLR subset statistics do not alter under a many instrument limit sequence, the dimensionality is an issue especially with respect to the involved HAC covariance matrix estimators. We have conducted several simulations of the size of GMM tests, including the ones we recommend, with different HAC covariance matrix estimators and found that the size can increase dramatically with the number of instruments. Our main concern with respect to the dimensionality is therefore not so much the appropriateness of the central limit theorem but more so the convergence of the HAC covariance matrix estimators.

6. WRIGHT

One of the main points of Wright's discussion is that the quest for strong instruments in the NKPC is essentially the same problem as the search for good forecasts of inflation and measures of economic slack. It is well known that pooling forecasts typically reduces the mean square prediction error especially when the pooled forecasts are not highly correlated. Wright therefore proposes to use the Greenbook and professional forecasts as additional instruments. These forecasts are judgmental, and therefore not highly correlated with the forecasts that result from autoregressions, which are implicitly used in the GMM estimation procedure. Since the Greenbook forecast is known to improve model-based inflation forecasts, one might expect that it can also improve the identification of the parameters in the NKPC. Figure 1 in Wright's discussion provides evidence of this as it shows that the S-based confidence set of (λ, γ_f) reduces considerably when we use the Greenbook forecast as an additional instrument. It is interesting to extend this analysis further and incorporate more powerful identification robust statistics like the KLM and MQLR statistics.

The second point in Wright's discussion concerns a different model of the NKPC that allows for exogenous time-varying drift in inflation. Cogley and Sbordone (2008) develop such a model and show that its coefficients are time-varying even though the underlying structural parameters, such as the degree of price stickiness, are constant. Since this model is subject to the same possible identification problems, we agree with Wright that it is important to apply identification robust methods for inference in this model, too. The main challenge is not that the Cogley and Sbordone (2008) model has more parameters than the standard NKPC, but rather that its coefficients are time-varying. This makes it hard to apply GMM methods directly. Moreover, as Wright points out, nonstationarity in inflation complicates the asymptotic distribution of the identification robust tests. This is similar to the point made by Mikusheva, which we managed to address in the restricted version of the standard NKPC by using lagged changes in inflation as instruments. However, this trick may not work more generally, so this remains an important open issue for future research.

Finally, Wright argues that it would be useful to have a measure of the strength of identification as a way to assess the reliability of point estimators and standard errors. A similar comment was made by Chao and Swanson. As we already stressed earlier, the identification robust confidence sets provide such a measure of identification. If they are tightly concentrated around the CUE, then the point estimates are reliable, otherwise they are not. Moreover, the fact that the identification robust confidence sets are computationally more demanding than the usual standard error bands should not be used as an argument against their use, since efficient computer programs are readily available. Judging from recent developments in empirical macroeconomics, it seems that leading researchers are not averse to using methods that are even more computationally demanding than the ones we propose.

7. YOGO

Yogo discusses our empirical findings on the NKPC from the perspective of the results documented in the literature. He shows that the 95% confidence sets on the slope coefficient and forward-looking parameter encompass those stated in Galí and Gertler (1999) for both of their normalizations. He states that this is indicative of the weak identification of these parameters with which we clearly concur.

An interesting perspective that Yogo takes from the large confidence sets of the structural price adjusment and backward-looking parameters is that it might be useful to incorporate information from microdata. Yogo shows that usage of a plug-in microdata estimate of the price adjustment parameter reduces the 95% confidence set for the slope and structural backward-looking parameter but not of the forward-looking parameter. This result is of considerable interest and shows a way of improving the precision of the empirical analysis that we find important.

8. ZIVOT AND CHAUDHURI

Zivot and Chaudhuri focus their discussion on the implementation and performance of the Robins test. Figure 3 of our article shows the power curves of the subset S-statistic and projection and Robins tests that are based on the S-statistic. Figure 3 shows that the projection and Robins tests are conservative and have less power than the subset S-statistic.

Our implementation of the Robins test is geared towards the subset S-statistic in the sense that the test attempts to reduce the power loss of the projection-based S-statistic relative to the subset S-statistic. Zivot and Chaudhuri show that our version of the Robins test can be improved upon by gearing the Robins test towards the KLM statistic. In case of two parameters (α, β), the Robins test for testing $H_0: \beta = \beta_0$ would then consist of constructing a $(1 - \nu) \times 100\%$ confidence set for α using $S(\alpha, \beta_0)$

and then, for each value of α inside the $(1 - \nu) \times 100\%$ confidence set, testing the hypothesis $H_0 : \beta = \beta_0$ using a conditional KLM statistic for β, which they define in their discussion and also in Chaudhuri (2007) and Chaudhuri et al. (2007). The power curves reported by Chaudhuri and Zivot in Figure 1 of their discussion result from this implementation of the Robins test and are similar to the ones we reported in Figure 3 of our article except that now the power curve of the Robins test is to be compared with the power curve of the KLM–JKLM combination and the MQLR statistic, which are the most powerful subset statistics. This is an interesting result since it shows that, although the Robins test remains conservative, it can be constructed in a way that is geared towards the most powerful statistic and can be made almost as powerful. This approach therefore offers a general principle for preserving the size of tests on subsets of the parameter with good power properties.

The manner in which the Robins test advocated by Zivot and Chaudhuri operates is similar to the combination of the KLM and JKLM statistic advocated by Kleibergen (2005) to avoid spurious power declines of the KLM statistic. Some issues that are of interest to investigate further for the Robins test are: (i) its performance at distant values of the hypothesized parameter for which the subset statistics resemble identification statistics and (ii) its ease of implementation since the construction of the $(1 - \nu) \times 100\%$ confidence set might be computationally demanding.

ADDITIONAL REFERENCES

Bekker, P. A., and Kleibergen, F. (2003), "Finite-Sample Instrumental Variables Inference Using an Asymptotically Pivotal Statistic," *Econometric Theory*, 19, 744–753.

Cogley, T., and Sbordone, A. M. (2008), "Trend Inflation and Inflation Persistence in the New Keynesian Phillips Curve," *American Economic Review*, 98, 2101–2126.

Hansen, C., Hausman, J., and Newey, W. (2008), "Estimation With Many Instrumental Variables," *Journal of Business & Economic Statistics*, 26, 398–422.

Mavroeidis, S. (2002), "Econometric Issues in Forward-Looking Monetary Models," Ph.D. thesis, Oxford University, Oxford.

Stock, J. H., and Yogo, M. (2003), "Testing for Weak Instruments in Linear IV Regression," Technical Working Paper 284, National Bureau of Economic Research.

Sun, Y., Phillips, P. C. B., and Jin, S. (2008), "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76 (1), 175–194.