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

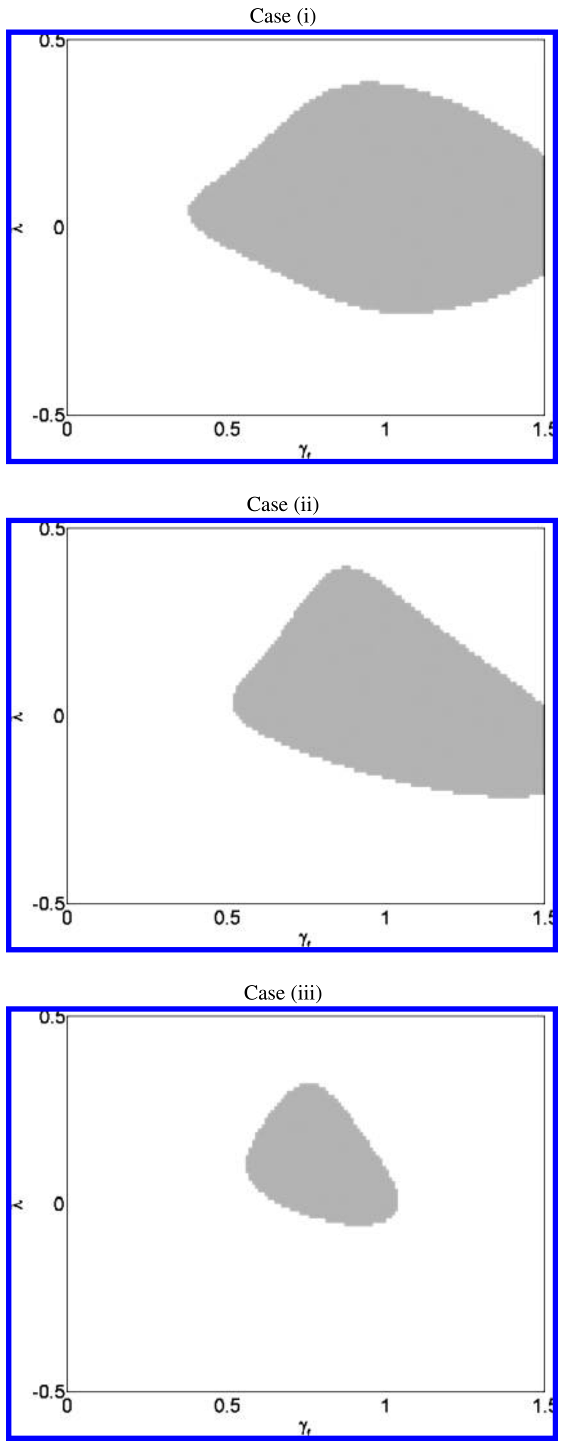

Figure 1. Three alternative confidence sets for the parameters of the NKPC model: Case (i): Kleibergen–Mavroeidis instruments; revised data. Case (ii): Kleibergen–Mavroeidis instruments; real-time inflation data. Case (iii): alternative instruments; real-time inflation data. *Notes*: Cases (i) and (ii) 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

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

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,

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

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

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