Comments on “Weak instrument robust tests in GMM and the new Keynesian Phillips curve” by F. Kleibergen and S. Mavroeidis [*]

Jean-Marie Dufour [†]

First version: August 2008
Revised: September 2008
This version: September 2008
Compiled: August 11, 2009, 11:08pm

This paper has been published in the *Journal of Business and Economic Statistics*, 27 (2009), 3, 318-321.

---

[*] This paper 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).

[†] William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

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*: a 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 which 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 comforting development for econometricians.

I will discuss the paper 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 (2003), Dufour and Taamouti (2005, 2007), Doko Tchatoka and Dufour (2008)] as well as NKPCs [Dufour, Khalaf, and Kichian (2006), Dufour, Khalaf, and Kichian (2007b), Dufour, Khalaf, and Kichian (2007a), Dufour, Khalaf, and Kichian (2008)]. I intend to focus on some pitfalls associated with the econometric methods proposed by the authors as well as 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.

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 which 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 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 come from the fact that instruments may used to solve an endogeneity or an errors-in-variables problem. It is very rare one can or should use all the possible valid instruments.

Consider the standard model:

$$y = Y\beta + X_1\gamma + u\,, \tag{3.1}$$

$$Y = X_1\mathit{\Pi}_1 + X_2\mathit{\Pi}_2 + V\,, \tag{3.2}$$

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, $\mathit{\Pi}_1$ and $\mathit{\Pi}_2$ 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

$$H_0(\beta_0) : \beta = \beta_0\,. \tag{3.3}$$

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

$$y - Y\beta_0 = X_1\theta_1 + X_2\theta_2 + \varepsilon \tag{3.4}$$

where $\theta_1 = \gamma + \mathit{\Pi}_1(\beta - \beta_0)$, $\theta_2 = \mathit{\Pi}_2(\beta - \beta_0)$ and $\varepsilon = u + V(\beta - \beta_0)$, $H_0(\beta_0)$ can be tested by testing

$$H_0' : \theta_2 = 0\,. \tag{3.5}$$

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

$$AR(\beta_0) = \frac{(y - Y\beta_0)'[M(X_1) - M(X)](y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(X)(y - Y\beta_0)/(T - k)} \sim F(k_2,\, T - k)\,, \tag{3.6}$$

which yields the confidence set $C_\beta(\alpha) = \{\beta_0 : AR(\beta_0) \le F_\alpha(k_2,\, T - k)\}$ 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 (3.2) 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 :

$$\mathsf{E}(Y) = X_1\mathit{\Pi}_1 + X_2\mathit{\Pi}_2\,. \tag{3.7}$$

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

Now suppose model (3.2) is in fact incomplete, and a third matrix of instruments does indeed appear in the reduced form for Y :

$$Y = X_1\mathit{\Pi}_1 + X_2\mathit{\Pi}_2 + X_3\mathit{\Pi}_3 + V \tag{3.8}$$

where X_3 is a $T \times k_3$ matrix of explanatory variables (not necessarily strictly exogenous). Equation (3.4) then becomes:

$$y - Y\beta_0 = X_1\mathit{\Delta}_1 + X_2\mathit{\Delta}_2 + X_3\mathit{\Delta}_3 + \varepsilon \tag{3.9}$$

where $\mathit{\Delta}_1 = \gamma + \mathit{\Pi}_1(\beta - \beta_0)$, $\mathit{\Delta}_2 = \mathit{\Pi}_2(\beta - \beta_0)$, $\mathit{\Delta}_3 = \mathit{\Pi}_3(\beta - \beta_0)$ and $\varepsilon = u + V(\beta - \beta_0)$. Since $\mathit{\Delta}_2 = 0$ and $\mathit{\Delta}_3 = 0$ under H_0, it is easy to see that the null distribution of $AR(\beta_0)$ 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 $V_1,\, \dots\, ,\, V_T$ may not follow a Gaussian distribution and may be heteroskedastic. A similar result obtains if

$$Y = g(X_1,\, X_2,\, X_3,\, V,\, \mathit{\Pi}) \tag{3.10}$$

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

$$Y = X_1\mathit{\Pi}_1 + X_2\mathit{\Pi}_2 + V \tag{3.11}$$

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.

	AR	ARS	K	LM	LR	LR1	LR2	AR	ARS	K	LM	LR	LR1	LR2
k_2	(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

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:

$$y = Y_1\beta_1 + Y_2\beta_2 + u\,, \quad (Y_1, Y_2) = X_2\Pi_2 + X_3\delta + (V_1, V_2), \tag{3.12}$$

$$(u_t, V_{1t}, V_{2t})' \overset{i.i.d.}{\sim} N(0,\ \Sigma)\,, \quad \Sigma = \begin{pmatrix} 1 & .8 & .8 \\ .8 & 1 & .3 \\ .8 & .3 & 1 \end{pmatrix}, \tag{3.13}$$

where δ represents the importance of the excluded instrument; a sample of these results is reproduced in Table 1. These show clearly that methods which depend heavily on the specification (3.2) 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 which 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 which are more powerful than AR-type procedures and 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 (2003b, 2003a) and Dufour (2003).

Concerning the GMM procedures used by Kleibergen and Mavroeidis (2009), there is no proof or discussion whether these enjoy robustness to missing instruments. For example, problems similar to "missing instruments" may be induced 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, the assumption (3.2) 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 Mavroeidis (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 be 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 Mavroeidis (2009).

4. Projection methods and subset inference

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

$$\mathsf{P}[\beta \in C_\beta(\alpha)] \geq 1 - \alpha \tag{4.1}$$

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

$$\mathsf{P}\big[g(\beta) \in g\left[C_\beta(\alpha)\right]\big] \geq 1 - \alpha\,. \tag{4.2}$$

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), Dufour and Taamouti (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), Dufour, Khalaf, and Kichian (2007b), Dufour, Khalaf, and Kichian (2007a), Dufour, Khalaf, and Kichian (2008)], 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 the methods proposed by the authors can be applied for that purpose and what results are obtained.

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.

References

ANDERSON, T. W., AND H. RUBIN (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20(1), 46–63.

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

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics, with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1389.

——— (2003): "Identification, Weak Instruments and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36(4), 767–808.

DUFOUR, J.-M., AND J. JASIAK (2001): “Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors,” *International Economic Review*, 42, 815–843.

DUFOUR, J.-M., L. KHALAF, AND M. KICHIAN (2006): “Inflation Dynamics and the New Keynesian Phillips Curve: An Identification Robust Econometric Analysis,” *Journal of Economic Dynamics and Control*, 30(9-10), 1707–1727.

——— (2007a): “Structural Multi-Equation Macroeconomic Models: A System-Based Estimation and Evaluation Approach,” Discussion paper, McGill University (Department of Economics), Montréal, and Bank of Canada, Ottawa, Canada, Montréal, Canada.

——— (2007b): “Which NKPC? Identification-Robust Estimates of Price Stickiness and Real Wage Rigidity,” Discussion paper, McGill University (Department of Economics), Montréal, and Bank of Canada, Ottawa, Canada.

——— (2008): “How Much Do Real Wage Rigidities Matter for Canadian Inflation?,” Discussion paper, McGill University (Department of Economics), Montréal, and Bank of Canada, Ottawa, Canada, Montréal, Canada.

DUFOUR, J.-M., AND M. TAAMOUTI (2003a): “On Methods for Selecting Instruments,” Discussion paper, C.R.D.E., Université de Montréal.

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

——— (2005): “Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments,” *Econometrica*, 73(4), 1351–1365.

——— (2007): “Further Results on Projection-Based Inference in IV Regressions with Weak, Collinear or Missing Instruments,” *Journal of Econometrics*, 139(1), 133–153.

KLEIBERGEN, F. (2002): “Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression,” *Econometrica*, 70(5), 1781–1803.

——— (2005): “Testing Parameters in GMM Without Assuming That They are Identified,” *Econometrica*, 73, 1103–1124.

——— (2007): “Subset Statistics in the Linear IV Regression Model,” Discussion paper, Department of Economics, Brown University, Providence, Rhode Island.

KLEIBERGEN, F., AND S. MAVROEIDIS (2008): “Inference on Subsets of Parameters in GMM Without Assuming Identification,” Discussion paper, Department of Economics, Brown university, Providence, Rhode Island.

——— (2009): “Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve,” *Journal of Business and Economic Statistics*, 27(3), 293–311.

MOREIRA, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71(4), 1027–1048.

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1097–1126.

WANG, J., AND E. ZIVOT (1998): "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66(6), 1389–1404.