

By Anna Mikusheva

We are witnessing a growing awareness among applied researchers about the possibility of having 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 pre-testing procedures, were popular for some 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 pre-test; in particular, the 2SLS results are usually used if the pretest indicates strong instruments. However, pre-testing procedures are not always reliable in controlling size, and are considered an inferior solution to weak identification problem (see for example, 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 pre-testing procedures. The paper 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 LM test of Kleibergen (2002) and the Conditional LR test by Moreira (2003)) satisfying the two conditions stated above. Andrews, Moreira and Stock (2007) studied the optimality properties of such tests under an assumption of homoskedasticity. There is also software (`condivreg` procedure in STATA) that allows for production of robust tests and confidence sets in no time and is easy to use (Mikusheva and Poi (2006)).

The paper we are discussing today is a big step forward in creating robust procedures in a more general situation. The authors consider a 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 were 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, there existed a trade-off between robustness to weak identification and efficiency under strong identification. Such a trade-off creates incentives to use pre-testing. The paper 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 is strong, but would have correct coverage if in some dimensions identification fails. The main message of the discussed paper is that one should not use pre-testing, 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 non-uniform or discontinuous asymptotics.

In the discussion below I would 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, or, 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 non-identification. It means that the closer the nuisance parameter 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. Or, in other words, if the parameter is very close to a non-identification point, the normal distribution provides a very poor description of the finite sample distribution of GMM t-statistics.

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 non-identification 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:

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

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 $\frac{1}{T} \sum_{t=1}^T \pi_{t-1} e_t \rightarrow^d \omega \int_0^1 w(t) dw(t)$. Notice two things: a non-standard 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 for example Bobkoski (1983) and Phillips (1987)). We can also argue that the partial sums of \bar{q}_t may also be non-normal for non-stationary components. In general, a use of persistent instruments may lead to Assumption 1's failure to hold.

One of the solutions suggested in the paper 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. And we are back 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 non-uniform 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 less 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 not very large sample sizes. An asymptotic efficiency oriented Kleibergen and Mavroeidis procedure seems to be more applicable to large cross-section data sets.

References

Anderson, T., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46-63.

Andrews, D.W.K. , M. Moreira and J. Stock (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74(3), 715-752.

Andrews D.W.K. and J.H. Stock (2005) Inference with Weak Instruments, NBER Technical Report.

Bobkovski M.J.(1983): “Hypothesis testing in nonstationary time series,” Unpublished PhD thesis (Dept. of Statistics, University of Wisconsin, Madison)

Kleibergen, F. (2002): “Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression,” *Econometrica*, 70, 1781-1803.

Kleibergen, F. and S. Mavroeidis (2008):“ Inference on subsets of parameters in GMM without assuming identification,” Working Paper, Brown University.

Mikusheva, A. and B. Poi (2006): “Tests and confidence sets with correct size when instruments are potentially weak,” *Stata Journal* , 6(3), 335-47.

Moreira, M. (2003): “A Conditional Likelihood Ratio Test for Structural Models,” *Econometrica*, 71 (4), 1027-1048.

Phillips P.C.B.(1987): “Toward a Unified Asymptotic Theory for Autoregression,” *Biometrika*, 74(3), 535-547.

Stock, J.H. and J.H. Wright (2000):“ GMM with Weak Identification,” *Econometrica*, 68, 1055-1096