A Comment on Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve

Eric Zivot*and Saraswata Chaudhuri†

October 27, 2008

1 Introduction

The paper by Kleibergen and Mavroeidis (2008a), 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 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 referred 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 the MQLR test recommended by

*Professor and Gary Waterman Distinguished Scholar, Department of Economics, University of Washington. Box 353330, Seattle, WA, 98195-3330. email: ezivot@u.washington.edu. Support from the Gary Waterman Distinguished Scholar Fund and the University of Washington Center for Statistics in the Social Sciences Seed Grant is gratefully acknowledged.

†Assistant Professor, Department of Economics, University of North Carolina Chapel Hill. email: saraswata_chaudhuri@unc.edu.

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

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

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 $q_t(\theta) = \text{vec}\left(\frac{\partial f_t(\theta)}{\partial \theta'}\right)$ 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

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix}\bar{f}_t(\theta)\\ \bar{q}_t(\theta)\end{bmatrix}\overset{d}{\to}\begin{bmatrix}\psi_f(\theta)\\ \psi_\theta(\theta)\end{bmatrix}\sim N(0,V(\theta)),\ V(\theta)=\begin{bmatrix}V_{ff}(\theta) & V_{f\theta}(\theta)\\ V_{\theta f}(\theta) & V_{\theta\theta}(\theta)\end{bmatrix}.$$

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

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

where $\hat{D}_T(\theta)=\sum_{t=1}^{T}D_t(\theta)$ and $D_t(\theta)=\text{devec}_k\left[q_t(\theta)-\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_t(\theta)\right]$. For the definition of the devec operator see Chaudhuri (2007).

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-statisic is a score-type statistic based on $Q(\theta)$ and may be expressed as

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

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 showed that irrespective of the identification of α, the S and the 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 (2007), Chaudhuri *et al.* (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 $R_\beta(\theta)$.

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

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

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

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

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 $C_\alpha^S(1-\xi, \beta_0)$ 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 (2007) 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

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

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

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

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

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

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

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

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 (2007) proved that if $C_\alpha(1-\xi,\beta_0)$ is non-empty 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 $C_\alpha(1-\xi,\beta_0) = C_\alpha^K(1-\xi,\beta_0)$ and $R(\theta) = S(\alpha,\beta_0)$ and concluded that the Robins test, proposed in Chaudhuri (2007) and Chaudhuri *et al.* (2008), does not outperform the usual method of projection based on $R(\theta) = S(\alpha,\beta_0)$. However, this is not what Chaudhuri (2007) and Chaudhuri *et al.* (2008) refer to as the Robins test. In the context of GMM, Chaudhuri (2007) and Chaudhuri and Zivot

(2008) recommend using $C_\alpha(1-\xi,\beta) = C_\alpha^S(1-\xi,\beta)$ 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 $C_\alpha^S(1-\xi,\beta_0)$ 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 illustrated 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

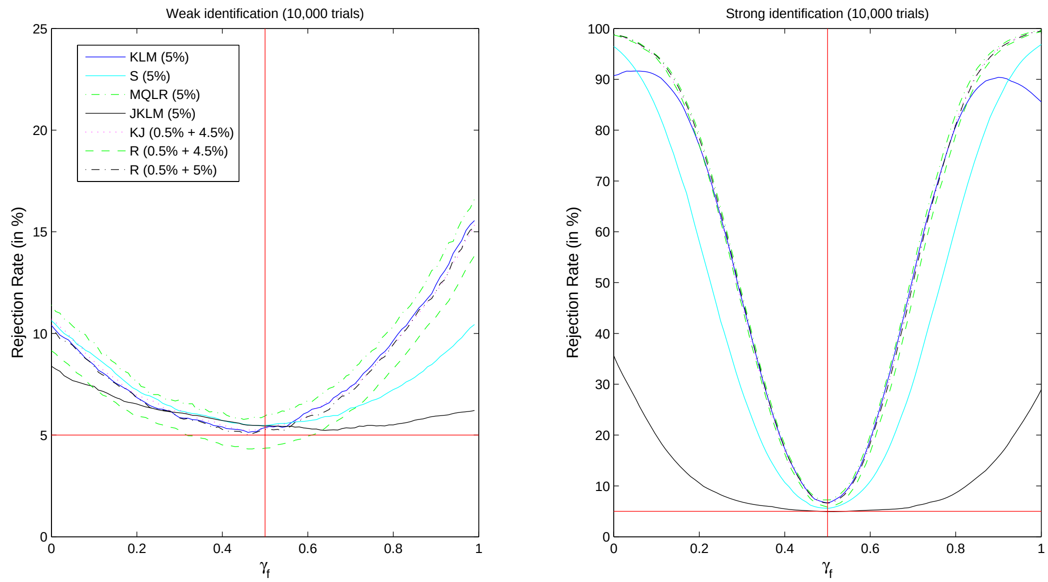

Figure 1: Power curves of 5% level tests for $H_0 : \gamma_f = 0.5$ against $H_1 : \gamma_f \neq 0.5$. The sample size is 1000 and the number of Monte Carlo simulations is 10000.

the new method of projection based on $C_\lambda^K(1-\xi, \gamma_f = 1/2)$ 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 $C_\lambda^S(1-\xi, \gamma_f = 1/2)$ 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 new method of projection actually performs as well as the MQLR and KJ tests recommended by KM. For the strong identification case, use of $C_\lambda^S(1-\xi, \gamma_f = 1/2)$ avoids the spurious decline in power observed for the KLM statistic.

4 Conclusion

KM showed 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 showed 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 real 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.

References

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

[2] 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.

[3] Chaudhuri, S. (2007). "Projection-Type Score Tests for Subsets of Parameters," Phd Thesis, Department of Economics, University of Washington.

[4] Chaudhuri, S. and Zivot, E. (2008). "A New Projection Test for Hypothesis on Subsets of Parameters," unpublished manuscript, Department of Economics, University of North Carolina.

[5] 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, Department of Economics, University of Washington.

[6] Dufour, J.M. (1990). "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58, 475-494.

[7] Dufour, J.M. (1997). "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365-1388.

[8] Dufour, J.M. and Jasiak, J. (2001). "Finite Sample Inference for Simultaneous Equations and Models with Unobserved and Generated Regressors," *International Economic Review*, 42, 815-844.

[9] Dufour, J.M. and Taamouti, M. (2005). "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351-1365.

[10] Dufour, J.M. and Taamouti, M. (2005). "Further Results on Projection-Based Inference in IV regressions with Weak, Collinear or Missing Instruments, *Journal of Econometrics*, 139, 133-153.

[11] Kleibergen, F. (2005). "Testing Parameters In GMM Without Assuming That They Are Identified," *Econometrica*, 73, 1103-1123.

[12] Kleibergen, F. and Mavroeidis, S. (2008a). "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," forthcoming in the *Journal of Business and Economic Statistics*.

[13] Kleibergen, F. and Mavroeidis, S. (2008b). "Inference on Subsets of Parameters in GMM without Assuming Identification," unpublished manuscript, Department of Economics, Brown University.

[14] Robins, J.M. (2004). "Optimal Structural Nested Models for Optimal Sequential Decisions, in Lin, D.Y. and Heagerty, P. (eds.) *Proceedings of the Second Seattle Symposium on Biostatistics*. Springer-Verlag, New York.

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

[16] Stock, J. H. and Wright, J.H. (2000). "GMM with Weak Identification," *Econometrica*, 68, 1055-1096.

[17] van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press.