A new method of projection-based inference in GMM with weakly identified nuisance parameters

Saraswata Chaudhuri* and Eric Zivot†

December 19, 2008

Abstract

Projection-based methods of inference on subsets of parameters are useful for obtaining tests that do not over-reject the true parameter values. However, they are also often criticized for being conservative. We show that the usual method of projection can be modified to obtain tests that are as powerful as the conventional tests for subsets of parameters. Like the usual projection-based methods, one can always put an upper bound to the rate at which the new method over-rejects the true value of the parameters of interest. The new method is described in the context of GMM with possibly weakly identified parameters.

JEL Classification: C12; C13; C30

Keywords: GMM; Weak identification; Efficient score statistic; Nuisance parameters; Confidence intervals

*Corresponding author. Department of Economics, CB 3305, University of North Carolina, Chapel Hill, NC 27519. Telephone: 919-966-3962. Fax: 919-966-4986. Email: saraswata_chaudhuri@unc.edu.

†Department of Economics, Box 353330, University of Washington, Seattle, WA 98195. Telephone: 206-543-6715. Fax: 206-685-7477. Email: ezivot@u.washington.edu.

1 Introduction

In this paper we are concerned with the problem of inference on subsets of parameters in a model where the identifiability of some or all of the parameters is in question. We focus on the framework of generalized method of moments (GMM) with possibly weakly identified parameters and our research can be seen as an extension of the work by Stock and Wright (2000), Dufour and Taamouti (2005), Kleibergen (2005) and Kleibergen and Mavroeidis (2008a).

Stock and Wright (2000) showed that the GMM estimators of the weakly identified parameters in a model are not consistent. While the GMM estimators of the (strongly) identified parameters are consistent, they are, in general, not asymptotically normal if there exists any weakly identified parameter in a model. Subsequently, the usual Wald, likelihood ratio and score statistics are not asymptotically pivotal even when evaluated at the true value of the parameters, and hence cannot be used for inference. However, they showed that the efficient continuous updating GMM (CU-GMM) objective function evaluated at the true value of the parameters is asymptotically pivotal, and proposed the S statistic based on the CU-GMM objective function to jointly test for all the parameters in a model. Kleibergen (2005) proposed a score-type statistic, which he called the K statistic, based on the gradient of the CU-GMM objective function and designed a method to jointly test for all the parameters in a model. This test, which came to be known as the K test, is generally more powerful than the S test based on the S statistic.

For inference on subsets of parameters regardless of the identifiability of the nuisance parameters, one can always use the usual projection technique based on the S and the K statistics to test and obtain confidence regions with (at least) correct asymptotic coverage probability. However, such inferences can be quite conservative. When the nuisance parameters are identified, use of projection is not required and the conservativeness of inference can be avoided by plugging in the CU-GMM estimator of the parameters, restricted by the null hypothesis, to the S and the K statistics. Stock and Wright (2000) and Kleibergen (2005) respectively showed that the resulting S and the K statistics are asymptotically pivotal and can be used to test for subsets of parameters simply by adjusting the critical values of the respective tests. We refer to these tests as the subset-S and the subset-K tests.

Recently Kleibergen and Mavroeidis (2008a) proved that the plug in principle does not over-size the subset-K test (and the generally less powerful subset-S test) even when the nuisance parameters are weakly identified. This is a major development in the weak identification literature. Hence in models with possibly weakly identified parameters, the subset-K test should serve as a benchmark to gauge the performance of any new method of inference in the GMM framework.

The purpose of this paper is to show how the usual methods of projection-based inference in the GMM framework can be modified to eliminate the conservativeness generally associated with them. In the process we also describe how and when this modified projec-

tion principle can be made asymptotically equivalent to the widely used plug in principle in the context of Kleibergen's subset-K test.

To fix ideas, let us denote the parameters of interest by θ_1 and the nuisance parameters by θ_2. We assume that the parameters are vectors of finite dimensions ν_1 and ν_2 respectively. Following the weak identification framework of Stock and Wright (2000), we consider four different cases of weak (partial) identification – WI-Case I with both θ_1 and θ_2 weakly identified, WI-Case II with θ_1 weakly identified but θ_2 (strongly) identified, WI-Case III with θ_1 identified but θ_2 weakly identified, and WI-Case IV with both θ_1 and θ_2 identified (i.e. the standard case).

We propose a projection-based method to test null hypotheses of the form $H_1 : \theta_1 = \theta_{*1}$ such that the test – (i) does not result in an uncontrolled over-rejection of the true value of θ_1 under any of the four cases of weak (partial) identification and (ii) is asymptotically (locally) equivalent to the subset-K test whenever the nuisance parameters θ_2 are identified, i.e. in WI-Cases II and IV.

Our method uses an efficient score equivalent of Kleibergen's K statistic. We call this the efficient K statistic for θ_1 and denote it by $K_{n1}(\theta_1, \theta_2)$ where the subscript n denotes the sample size.[1] We also note that, under the null hypothesis $H_1 : \theta_1 = \theta_{*1}$, a confidence region for the nuisance parameters θ_2 can be obtained by inverting any weak identification robust test; e.g. the S test, the K test or the GMM-MLR test [see Kleibergen (2005)]. We denote such a confidence region generically by $\mathcal{C}_2(1-\zeta, \theta_{*1})$ where $(1-\zeta)$ (bounded away from 0 and 1) gives its (intended) asymptotic coverage probability under the null hypothesis.

Our method rejects the null hypothesis $H_1 : \theta_1 = \theta_{*1}$

(i) if $\mathcal{C}_2(1-\zeta, \theta_{*1}) = \varnothing$, i.e. if the confidence region is empty,

(ii) or if $\inf_{\theta_2 \in \mathcal{C}_2(1-\zeta, \theta_{*1})} K_{n1}(\theta_{*1}, \theta_2) > \chi^2_{\nu_1}(1-\epsilon)$ where $\chi^2_{\nu_1}(1-\epsilon)$ is the $(1-\epsilon)$ quantile of a central chi-square distribution with ν_1 degrees of freedom.

While the use of an empty confidence region $\mathcal{C}_2(1-\zeta, \theta_{*1})$ may seem counterintuitive, it is possible to obtain a null set when the confidence region is obtained by inverting the S test. This occurs when the over-identification restrictions are rejected under the null hypothesis $H_1 : \theta_1 = \theta_{*1}$. In Section 3 we will see that such confidence regions can occur with distinctly positive probability.

This new method can be seen as a two-step procedure. In the first step we construct a confidence region for the nuisance parameters such that the region has (at least) the correct

[1] In the context of maximum likelihood, one can define the efficient score for θ_1 as the part of the score for θ_1 that is orthogonal to the score for θ_2 [see van der Vaart (1998)]. The efficient score statistic for θ_1 is a quadratic form of the efficient score for θ_1 (ideally, with respect to its asymptotic variance). Based on this idea and using the fact that the K statistic is a score-type statistic based on the gradient of the CU-GMM objective function, we define an efficient score equivalent of the K statistic [see Section 2].

asymptotic coverage probability $(1-\zeta)$ under the null hypothesis $H_1 : \theta_1 = \theta_{*1}$. And in the second step we reject the null hypothesis if the infimum (with respect to θ_2 inside the confidence region) of the statistic $K_{n1}(\theta_{*1}, \theta_2)$ is larger than the $\chi^2_{\nu_1}(1-\epsilon)$ critical value. This method is primarily motivated from Robins (2004). Similar techniques of hypothesis testing in the presence of nuisance parameters were also suggested by Dufour (1990), Berger and Boos (1994) and Silvapulle (1996).

Our proposed method is based on projection. However, unlike the usual technique of projection found in the weak identification literature [see, among others, Dufour (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005, 2007)], we project from a subset of the parameter space of the nuisance parameters θ_2, restricted by the null hypothesis $H_1 : \theta_1 = \theta_{*1}$. Such restrictions prove to be crucial in reducing the conservativeness of the methods of inference based on projection, especially in finite samples. Additionally, the use of the efficient K statistic (an analog of the efficient score statistic) in the second step of the test makes it comparable to the subset-K test in WI-Cases II and IV. To highlight these two important facets of the test – the use of (i) the restricted projection and (ii) the efficient K statistic, we will, henceforth, refer to it as the *efficient projection-based K* test.

The new method describes a class of projection-based tests indexed by the choice of the first step confidence region and the choice of ζ and ϵ (for fixed $\zeta + \epsilon$). In Section 2 we discuss the relative merits of such choices.

The main properties of the efficient projection-based K test established in this paper are described below based on the following observations:

(i) Under the null hypothesis $H_1 : \theta_1 = \theta_{*1}$, the region $\mathcal{C}_2(1-\zeta, \theta_{*1})$ (obtained by inverting the S test, the K test or the GMM-MLR test) contains the true value of θ_2 with asymptotic probability $1-\zeta$.

(ii) $K_{n1}(\theta_1, \theta_2)$, evaluated at the true value of the parameters, asymptotically converges to a $\chi^2_{\nu_1}$ distribution.

(iii) For $\sqrt{n}$-local alternatives, when the nuisance parameters are identified, i.e. in WI-Cases II and IV, any point inside $\mathcal{C}_2(1-\zeta, \theta_{*1})$ (provided it is nonempty) is a $\sqrt{n}$-consistent estimator of the true value of θ_2.[2]

(iv) For $\sqrt{n}$-local alternatives, when the nuisance parameters are identified, i.e. in WI-Cases II and IV, $\inf_{\theta_2 \in \mathcal{C}_2(1-\zeta, \theta_{*1})} K_{n1}(\theta_{*1}, \theta_2)$ (provided $\mathcal{C}_2(1-\zeta, \theta_{*1})$ is nonempty) is asymptotically equivalent to the subset-K statistic.

Hence using (i) and (ii), it follows from Bonferroni's inequality that the asymptotic size of the efficient projection-based K test cannot exceed $\zeta + \epsilon$. Furthermore, from (iii) and (iv) and conditional on $\mathcal{C}_2(1-\zeta, \theta_{*1}) \neq \varnothing$, it follows that the efficient projection-based K test is asymptotically equivalent to the level-ϵ subset-K test against $\sqrt{n}$-local alternatives.

[2]Section 2 provides a formal result for confidence regions based on the S test and the K test. [also see Kleibergen (2005)]

If we allow for $\mathcal{C}_2(1-\zeta,\theta_{*1})$ to be empty, then it can be shown that a suitable choice of $\mathcal{C}_2(1-\zeta,\theta_{*1})$ (based on Kleibergen's J-statistic) extends the asymptotic equivalence to the level-$(\zeta+\epsilon)$ subset-K-J test [see Kleibergen (2005)]. These latter properties are particularly remarkable. In view of the recent developments in the weak identification literature, the asymptotic equivalence with the subset-K(-J) test is an important property that vindicates the effectiveness of the new method of projection-based inference proposed in this paper.

While the usefulness of the (usual) projection technique in designing tests that are not over-sized has already been well established in a series of papers by Dufour and his co-authors [see, among others, Dufour (1990), Dufour (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005, 2007)], often the projection-based tests are found to be overly conservative. The new method of projection used in the efficient projection-based K test helps to avoid this problem of conservativeness asymptotically. We provide simulations in a linear instrumental variables (IV) regression to show that the asymptotic gains do not go away in finite samples. Further simulations are also presented in Chaudhuri et al. (2008) and Zivot and Chaudhuri (2008).

The rest of the paper is organized as follows. Section 2 describes the efficient projection-based K test in the context of GMM estimation, Section 3 is a Monte Carlo study in a linear instrumental variables models and shows that the asymptotic results of Section 2 provide a good approximation to the behavior of the efficient projection-based K test in finite samples, and Section 4 gives our conclusions. Proofs of all results are given in the Appendix.

We use the following notations in the rest of the paper. If $A=[A_1,\ldots,A_{bc}]$ is an $a\times bc$ matrix then $vecA=[A_1',\ldots,A_{bc}']'$, $devec_cA'=\left[(A_1,\ldots,A_c)',\ldots,(A_{(b-1)c+1},\ldots,A_{bc})'\right]$ and $\|A\|=\sqrt{\text{trace}(A'A)}$. If A is full column rank then $P(A)=A(A'A)^{-1}A'$ and $N(A)=I_a-P(A)$ where I_a is the $a\times a$ identity matrix. The notation $\mathcal{N}(\mu,\sigma^2)$ is used to denote a normal distribution with mean μ and variance σ^2. If A is a symmetric positive semi-definite matrix then $A^{\frac{1}{2}}$ is the lower-triangular Cholesky factor of A such that $A=A^{\frac{1}{2}}A^{\frac{1}{2}\prime}$. If $A=((A_{ij}))_{i,j=1,2}$ is such that the diagonal blocks A_{11} and A_{22} are non-singular then $A_{ii.j}=A_{ii}-A_{ij}A_{jj}^{-1}A_{ji}$ denotes the Schur complement of A_{jj} for $i\neq j=1,2$. For any random variable X such that $E\|X\|\leq\infty$, let $\overline{X}=X-EX$ where E stands for expectation. Since all the matrices considered here are of finite dimensions, we tend to be less scrupulous with mixing the notations like $\overline{X}=o_p(1)$ and $\|\overline{X}\|=o_p(1)$; both implying that every element of $\overline{X}$ converges in probability to zero. Lastly, we use $\xrightarrow{P}$, $\xrightarrow{d}$ to denote convergence in probability and distribution respectively, $\overset{A}{\sim}$ to denote "asymptotically follows"and the acronym w.p.a.1 for "with probability approaching one".

2 The efficient projection-based K test in GMM

We first describe the assumptions in the GMM framework before introducing the efficient projection-based K test. The moment restrictions defined below can (but need not) be viewed as obtained from the first order conditions of some optimization problem. Let

$g : \mathcal{S} \times \Theta \mapsto \mathbb{R}^k$ be a measurable and twice-continuously differentiable function such that for the sample of observations $\{w_t : t = 1, \ldots, n\}$ we have

$$\begin{aligned} Eg(w_t, \theta) &= 0 \text{ if } \theta = \theta_0, \\ &\neq 0 \text{ if } \theta \neq \theta_0. \end{aligned} \tag{2.1}$$

Equation (2.1) gives $k \geq \nu$ moment restrictions for inference on ν unknown elements of θ and is often referred to as the *global* identification condition. We assume that ν and k are fixed and finite numbers. For simplicity, let $\theta_0 = (\theta_{01}', \theta_{02}')'$ be such that $\theta_{0i} \in \text{interior}(\Theta_i)$ where Θ_i is a compact (ν_i-dimensional) subset of $\mathbb{R}^{\nu_i}$ for $i = 1, 2$. The parameter space $\Theta = \Theta_1 \times \Theta_2$ is a compact subset of $\mathbb{R}^\nu$ where $\nu = \nu_1 + \nu_2$.

In the following we suppress the explicit dependence of the functionals on the observations; for example, $g_t(\theta)$ should be read as $g(w_t, \theta)$. Define $\nabla_\theta g_t(\theta) = \partial g_t(\theta)/\partial\theta'$ and $\nabla_{\theta\theta} g_t(\theta) = \partial vec \nabla_\theta g_t(\theta)/\partial\theta'$.

2.1 Assumptions on the moment vector and weak identification

We make the following assumptions similar to Stock and Wright (2000), Guggenberger and Smith (2005) and Kleibergen (2005). Chaudhuri (2008) contains more discussion on these assumptions.

High level assumptions on the moment restrictions and their first derivatives are summarized under Assumption D.

Assumption D:

D1. There exists an open neighborhood $\mathcal{T} \subset \Theta$ containing θ_0 such that

(i) $n^{-1} \sum_{t=1}^n g_t(\theta)$ is continuously differentiable almost surely, and

(ii) $\sup_{\theta \in \mathcal{T}} \| n^{-1} \sum_{t=1}^n \nabla_\theta g_t(\theta) \|$ is integrable.

D2. $\sup_{\theta \in \Theta} n^{-1} \sum_{t=1}^n \overline{g_t(\theta)} = o_p(1)$,$n^{-1} \sum_{t=1}^n \overline{\nabla_\theta g_t(\theta)} = o_p(1)$, $n^{-1} \sum_{t=1}^n \overline{\nabla_{\theta\theta} g_t(\theta)} = o_p(1)$. $En^{-1} \sum_{t=1}^n \nabla_{\theta\theta} g_t(\theta)$ converges to a continuous and bounded function $L(\theta)$ for $\theta \in \Theta$.

D3. $n^{-1/2} \sum_{t=1}^n \begin{bmatrix} \overline{g_t(\theta_0)} \\ vec \overline{\nabla_\theta g_t(\theta_0)} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Psi_g \\ \Psi_\nabla \end{bmatrix} \sim \mathcal{N}(0, V(\theta_0))$ and the asymptotic variance-covariance matrix $V(\theta) = ((V_{ab}(\theta)))_{a,b=g,\nabla}$ is bounded, continuous, symmetric and positive semi-definite. $V_{gg}(\theta)$ is symmetric, positive definite and differentiable with respect to $\theta \in \Theta$.

D4. There exist $\widehat{V}_{\nabla g}(\theta)$ and a symmetric positive definite matrix $\widehat{V}_{gg}(\theta)$ such that $\widehat{V}_{\nabla g}(\theta) - V_{\nabla g}(\theta) = o_p(1)$ and $\partial vec \widehat{V}_{gg}(\theta)/\partial\theta' - \partial vec V_{gg}(\theta)/\partial\theta' = o_p(1)$ for $\theta \in \Theta$ and $\sup_{\theta \in \Theta} \left(\widehat{V}_{gg}(\theta) - V_{gg}(\theta) \right) = o_p(1)$.

The characterization of weak identification due to Stock and Wright (2000) is summarized under Assumption W.

Assumption W:

$En^{-1}\sum_{t=1}^{n} g_t(\theta_1,\theta_2) = \sum_{i=1}^{2}\left[1_{[\delta_i=1]}m_i(\theta_i) + 1_{[\delta_i=\frac{1}{2}]}n^{-1/2}\tilde{m}_{ni}(\theta_1,\theta_2)\right]$ where for $i=1,2$,

1. $m_i(\theta_{0i}) = 0, m_i(\theta_i) \neq 0$ for $\theta_i \neq \theta_{0i}$, $M_i(\theta_i) = \partial m_i(\theta_i)/\partial\theta_i'$ is continuous and $M_i(\theta_{0i})$ has full column rank.

2. $\sup_{\theta\in\Theta}(\tilde{m}_{ni}(\theta) - \tilde{m}_i(\theta)) = o(1)$, $\tilde{m}_i(\theta_0) = 0$ and $\tilde{m}_i(\theta)$ is continuous in θ and bounded on Θ. For $i,j=1,2$, $\tilde{M}_{n(i,j)}(\theta) = \partial\tilde{m}_{ni}(\theta)/\partial\theta_j'$ converges to some function $\tilde{M}_{(i,j)}(\theta)$.

The non-random indicator functions involving the δ's in Assumption W distinguish between the four cases of weak (partial) identification mentioned in the introduction and summarized in Table 1.[3]

Table 1: Four Cases of Weak (Partial) Identification.

	$\delta_2 = \frac{1}{2}$	$\delta_2 = 1$
$\delta_1 = \frac{1}{2}$	**WI-Case I** θ_1 : weakly identified θ_2 : weakly identified	**WI-Case II** θ_1 : weakly identified θ_2 : (strongly) identified
$\delta_1 = 1$	**WI-Case III** θ_1 : (strongly) identified θ_2 : weakly identified	**WI-Case IV** θ_1 : (strongly) identified θ_2 : (strongly) identified

Assumptions D1, D2 and W imply that for $\theta \in \mathcal{T}$ and for $i=1,2$,

$$\begin{aligned} G_{ni}(\theta) = E\frac{1}{n^{\delta_i}}\nabla_i\sum_{t=1}^{n} g_t(\theta) &= \frac{\partial}{\partial\theta_i'}E\frac{1}{n^{\delta_i}}\sum_{t=1}^{n} g_t(\theta) \\ &= 1_{[\delta_i=1]}M_i(\theta_i) + \sum_{j=1}^{2} 1_{[\delta_j=\frac{1}{2}]}\frac{n^{\delta_j}}{n^{\delta_i}}\tilde{M}_{n(j,i)}(\theta) \end{aligned} \tag{2.2}$$

where $\nabla_1 g_t(\theta)$ and $\nabla_2 g_t(\theta)$ are, respectively, the first ν_1 columns and last ν_2 columns of $\nabla_\theta g_t(\theta)$. Assumption W further implies that for $i,j=1,2$ (and $i \neq j$), $G_{ni}(\theta)$ is continuous

[3]The δ_i's are assigned the values $\frac{1}{2}$ and 1 because, as it can be seen from (2.2), n^{δ_i} will often be used as a suitable scaling factor.

in θ and

$$G_{ni}(\theta) \xrightarrow{P} G_i(\theta) = 1_{[\delta_i=1]} M_i(\theta_i) + 1_{[\delta_i=\frac{1}{2}]} \left[\tilde{M}_{(i,i)}(\theta) + 1_{[\delta_j=\frac{1}{2}]} \tilde{M}_{(j,i)}(\theta) \right],$$

which has full column rank for $\theta \in \theta_{0i} \times \Theta_j$ when θ_i is identified. When both θ_1 and θ_2 are identified, this implies the so-called *local* identification condition; i.e., the expected Jacobian is full column rank at $\theta = \theta_0$ [see Kleibergen (2005)].

It is possible to relax some of our assumptions. For example, since we are only concerned with the asymptotic behavior of the tests against $\sqrt{n}$-local alternatives, it is sufficient if all the assumptions specified for $\theta \in \Theta$ hold in $\mathcal{T}_1 \times \Theta_2$ where $\mathcal{T}_1 \subset \Theta_1$ is an open neighborhood containing θ_{01}.[4] Nevertheless, we made these simplifying assumptions to avoid the secondary details in the exposition which can obscure the basic idea behind the efficient projection-based K test. Assumption D is stated in a somewhat unconventional form so that it can be directly applied to prove our results. However, these assumptions are not different in nature from those in Stock and Wright (2000), Guggenberger and Smith (2005) and Kleibergen (2005). For example, Assumption B in Stock and Wright (2000) states that $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \overline{g_t(\theta)} \Rightarrow \xi(\theta)$ for $\theta \in \Theta$ where $\xi(\theta)$ is a mean-zero Gaussian stochastic process. By definition of weak convergence (denoted by "$\Rightarrow$"), this implies that $\sup_{\theta\in\Theta} \|\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \overline{g_t(\theta)}\| \xrightarrow{d} \sup_{\theta\in\Theta} \|\xi(\theta)\|$ and thus implies that $\sup_{\theta\in\Theta} n^{-1} \sum_{t=1}^{n} \overline{g_t(\theta)} = o_p(1)$ [see Andrews (1994)]. Under this assumption, in order to show the consistency of the CU-GMM estimator of a strongly identified θ_2, restricted by a hypothesized $\sqrt{n}$-local θ_1, it is also required to assume that $\sup_{\theta\in\mathcal{T}_1\times\Theta_2} \|\xi(\theta)\| = O_p(1)$. Instead, we directly assume that $\sup_{\theta\in\Theta} n^{-1} \sum_{t=1}^{n} \overline{g_t(\theta)} = o_p(1)$ (in Assumption D2) and that $V_{gg}(\theta)$ is bounded in Θ (in Assumption D3). Similarly, instead of specifying the form of $\widehat{V}_{gg}(\theta)$ and making assumptions such as Assumption M(ii) in Guggenberger and Smith (2005) to ensure its non-singularity for $\theta \in \mathcal{T}_1 \times \Theta_2$, we directly assume that there exists a matrix $\widehat{V}_{gg}(\theta)$ which is positive definite along with its convergence result (in Assumption D4). Finally, unlike Assumptions M_θ (iii) and (vii) in Guggenberger and Smith (2005), our distributional assumption (in Assumption D3) is local in nature [similar to Kleibergen (2005)] and hence we need to specify assumptions (in Assumption D2) on the second derivative of the moment vector for studying the properties of the tests against $\sqrt{n}$-local alternatives. Assumption D1, which allows us to interchange the order of differentiation and integration, is made for simplicity.

2.2 The subset-K test and the efficient projection-based K test

Kleibergen (2005) pointed out that the failure of the usual score test (based on the efficient two-step GMM) in the presence of weakly identified parameters is due to the (asymptotic) non-zero correlation between the estimator of the expected Jacobian and the moment vector

[4] In fact, to prove that the asymptotic size of the efficient projection-based K test is bounded from above and to show that under WI-Cases II and IV the test is asymptotically equivalent to the infeasible efficient K test (described after Lemma 2.2 and before Theorem 2.3) against $\sqrt{n}$-local alternatives, it is sufficient if the above assumptions hold in $\mathcal{T} \subset \Theta$, an open neighborhood containing θ_0.

(both under suitable scaling). He noted that the CU-GMM offers a natural way of constructing the estimator of the expected Jacobian such that even under weak identification, the (scaled) estimator of the expected Jacobian and the moment vector are asymptotically uncorrelated (and under Assumption D3, asymptotically independent). Kleibergen's K statistic is a quadratic form in the gradient of the CU-GMM objective function.

Using Assumptions D3 and D4, one can define the CU-GMM objective function (with an efficient weighting matrix) as

$$Q_n(\theta) = \frac{1}{2n}\left[\sum_{t=1}^{n} g_t(\theta)\right]' \widehat{V}_{gg}^{-1}(\theta)\left[\sum_{t=1}^{n} g_t(\theta)\right].$$

The gradient of the CU-GMM objective function with respect to θ is given by

$$\nabla_\theta Q_n(\theta) = \frac{\partial Q_n(\theta)}{\partial \theta'} = \frac{1}{n} g_T'(\theta)\widehat{V}_{gg}^{-1}(\theta)\widehat{D}_T(\theta), \tag{2.3}$$

where $g_T(\theta) = \sum_{t=1}^{n} g_t(\theta)$, $\widehat{D}_T(\theta) = \sum_{t=1}^{n} \widehat{D}_t(\theta)$ and $\widehat{D}_t(\theta) = devec_k\left[vec\nabla_\theta g_t(\theta) - \widehat{V}_{\nabla g}(\theta)\right.$ $\left.\widehat{V}_{gg}^{-1}(\theta) g_t(\theta)\right]'$. The CUE (CU-GMM estimator) $\widehat{\theta}_n$ of $\theta \in \Theta$ satisfies the first-order condition

$$\nabla_\theta Q_n(\widehat{\theta}_n) = \left[\nabla_1 Q_n(\widehat{\theta}_n), \nabla_2 Q_n(\widehat{\theta}_n)\right] = \frac{1}{n} g_T'(\widehat{\theta}_n)\widehat{V}_{gg}^{-1}(\widehat{\theta}_n)\widehat{D}_T(\widehat{\theta}_n) = 0,$$

w.p.a.1. Similarly, under the restriction that $\theta_1 = \theta_{*1}$, the CUE $\tilde{\theta}_{n2}(\theta_{*1})$ of θ_2 minimizes $Q_n(\theta_{*1}, \theta_2)$ with respect to $\theta_2 \in \Theta_2$ and hence, w.p.a.1, $\tilde{\theta}_* = (\theta_{*1}', \tilde{\theta}_{n2}'(\theta_{*1}))'$ satisfies the first-order condition

$$\nabla_2 Q_n(\tilde{\theta}_*) = \frac{1}{n} g_T'(\tilde{\theta}_*)\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\widehat{D}_{T2}(\tilde{\theta}_*) = 0, \tag{2.4}$$

where $\widehat{D}_{Ti}(\theta) = \sum_{t=1}^{n} \widehat{D}_{ti}(\theta)$ and $\widehat{D}_{ti}(\theta) = devec_k\left[vec\nabla_i g_t(\theta) - \widehat{V}_{ig}(\theta)\widehat{V}_{gg}^{-1}(\theta) g_t(\theta)\right]'$ for $i = 1, 2$. The above expression uses the partition $\widehat{D}_T(\theta) = \left[\widehat{D}_{T1}(\theta), \widehat{D}_{T2}(\theta)\right]$ and $\widehat{V}_{\nabla g}(\theta) = \left[\widehat{V}_{1g}'(\theta), \widehat{V}_{2g}'(\theta)\right]'$ with respect to θ_1 and θ_2.

Lemma 2.1 *Let $\theta_{*1} = \theta_{01} + d_1/\sqrt{n} \in \Theta_1$ where $d_1 \in \mathbb{R}^{\nu_1}$. Then under Assumptions D and W, $\sqrt{n}(\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02}) = O_p(1)$ in WI-Cases II and IV.*

Lemma 2.1 follows directly from Lemma A1 in Stock and Wright (2000) which shows the $\sqrt{n}$-consistency of the unconstrained estimator of θ_2 in WI-Case II. A formal proof is also given in the Appendix.

Kleibergen's subset-K test rejects the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ at level ϵ if $K_n(\tilde{\theta}_*) >$

$\chi^2_{\nu_1}(1-\epsilon)$ where the K statistic is defined by Kleibergen (2005) as

$$\begin{aligned} K_n(\theta) &= n\left(\nabla_\theta Q_n(\theta)\right)\left[\widehat{D}_T'(\theta)\widehat{V}_{gg}^{-1}(\theta)\widehat{D}_T(\theta)\right]^{-1}\left(\nabla_\theta Q_n(\theta)\right)' \\ &= \frac{1}{n} g_T'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta)P\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_T(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)g_T(\theta). \end{aligned} \tag{2.5}$$

In WI-Cases II and IV and under Assumptions D and W, $K_n(\theta_{01}, \tilde{\theta}_{n2}(\theta_{01})) \overset{A}{\sim} \chi^2_{\nu_1}$. See Theorem 2 in Kleibergen (2005) for the proof under presumably weaker conditions. The limiting $\chi^2_{\nu_1}$ distribution of the K statistic $K_n(\theta_{01}, \tilde{\theta}_{n2}(\theta_{01}))$ in Kleibergen's proof, however, crucially depends on the $\sqrt{n}$-consistency of $\tilde{\theta}_{n2}(\theta_{01})$.

In WI-Cases I and III, $\tilde{\theta}_{n2}(\theta_{01})$ is inconsistent and deducing the properties of the K statistic (and hence the subset-K test) is still an area of ongoing research. It is in these two cases where the literature typically recommends the use of projection techniques. However, as discussed in the introduction, Kleibergen and Mavroeidis (2008a) recently showed that even in WI-Cases I and III, where the nuisance parameters θ_2 are weakly identified, the subset-K test does not over-reject the true values of the parameters of interest.

The above argument in favor of the subset-K test essentially marginalizes the need for projection-based inference in the context of GMM with weakly identified parameters. At the same time, this also provides an excellent benchmark to establish the usefulness of the new method of projection by comparing the efficient projection-based K test against the subset-K test. In the rest of the paper we will show that the efficient projection-based K test performs as good as the subset-K test in terms of size and power.

Now we discuss the construction of the efficient K statistic, which is crucial for the local asymptotic equivalence between the subset-K test and the efficient projection-based K test. The efficient K statistic is the analog of the efficient score statistic in the CU-GMM framework. Because the efficient score statistic is not often explicitly used in the econometrics literature, we first provide a very simple example to fix the idea [also see Bera and Bilias (2001)].

Example of an efficient score statistic

Consider n observations $w_t = (y_t, X_{1t}, X_{2t})$ such that $y_t \overset{\text{indep}}{\sim} (X_{1t}\theta_{01} + X_{2t}\theta_{02}, 1)$ for $t = 1, \ldots, n$. Denote $y = (y_1, \ldots, y_n)'$ and $X_i = (X_{i1}, \ldots, X_{in})'$ for $i = 1, 2$. Assume that $X = [X_1, X_2]$ is non-stochastic and let

$$\frac{1}{n}X'X = \frac{1}{n}\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \to \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \Omega \text{ (be positive definite).}$$

The positive definiteness assumption rules out problems with identification and allows us to focus just on the idea behind the efficient score statistic. Defining the moment restrictions as $Eg_t(\theta_0) = 0$ (and $\neq 0$ otherwise) where $g_t(\theta) = (X_{1t}, X_{2t})'(y_t - X_{1t}\theta_1 - X_{2t}\theta_2)$, it follows

from (2.3) and the Lindberg-Feller CLT that the scaled score (gradient)[5]

$$\begin{aligned}\left[\frac{1}{\sqrt{n}}\nabla_\theta Q_n(\theta_0)\right]' &= \begin{bmatrix} \frac{1}{\sqrt{n}}(\nabla_1 Q_n(\theta_0))' \\ \frac{1}{\sqrt{n}}(\nabla_2 Q_n(\theta_0))' \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{n}}X_1'(y - X_1\theta_{01} - X_2\theta_{02}) \\ -\frac{1}{\sqrt{n}}X_2'(y - X_1\theta_{01} - X_2\theta_{02}) \end{bmatrix} \\ &\overset{A}{\sim} \mathcal{N}(0, \Omega = V_{gg}(\theta)).\end{aligned}$$

The sum of the n observations of the population version of the efficient score (gradient) for θ_1 is $\nabla_{1.2}Q_n^{\text{pop}}(\theta) = \nabla_1 Q_n(\theta) - \nabla_2 Q_n(\theta)\Omega_{22}^{-1}\Omega_{21}$. A sample version replaces the unknown Ω_{ij} terms by $n^{-1}X_i'X_j$ and is defined as $\nabla_{1.2}Q_n(\theta) = \nabla_1 Q_n(\theta) - \nabla_2 Q_n(\theta)(X_2'X_2)^{-1}X_2'X_1$. It follows that

$$\begin{aligned}\begin{bmatrix} \frac{1}{\sqrt{n}}(\nabla_{1.2}Q_n(\theta))' \\ \frac{1}{\sqrt{n}}(\nabla_2 Q_n(\theta))' \end{bmatrix} &= \begin{bmatrix} -\frac{1}{\sqrt{n}}X_1'N(X_2)(y - X_1\theta_1 - X_2\theta_2) \\ -\frac{1}{\sqrt{n}}X_2'(y - X_1\theta_1 - X_2\theta_2) \end{bmatrix} \\ &\overset{A}{\sim} \mathcal{N}\left(\begin{bmatrix} -\Omega_{11.2}(\theta_{01} - \theta_1) \\ -\Omega_{21}(\theta_{01} - \theta_1) - \Omega_{22}(\theta_{02} - \theta_2) \end{bmatrix}, \begin{bmatrix} \Omega_{11.2} & 0 \\ 0 & \Omega_{22} \end{bmatrix}\right),\end{aligned}$$

i.e. the (scaled) efficient score for θ_1 is asymptotically independent of the (scaled) score for θ_2 for any value of the nuisance parameters. An efficient score statistic for testing $H_1 : \theta_1 = \theta_{*1}$ can be constructed as

$$\begin{aligned}\text{LM}_n^{\text{eff}}(\theta_{*1}) &= \frac{1}{n}(\nabla_{1.2}Q_n(\theta_{*1}, \theta_2))\widehat{\Omega}_{11.2}^{-1}(\nabla_{1.2}Q_n(\theta_{*1}, \theta_2))' \\ &= (y - X_1\theta_{*1} - X_2\theta_2)'P(N(X_2)X_1)(y - X_1\theta_{*1} - X_2\theta_2).\end{aligned} \tag{2.6}$$

For θ_{*1} as defined in Lemma 2.1, i.e. for $\sqrt{n}$ local alternatives, $\text{LM}_n^{\text{eff}}(\theta_{*1}) \overset{A}{\sim} \chi_1^2$ with non-centrality parameter equal to $d_1'\Omega_{11.2}d_1$. Since $N(X_2)X_2 = 0$, the efficient score statistic for θ_1 does not depend on θ_2 at all. While this nice property does not completely carry through in a general nonlinear model, we show in Lemma 2.2 that the asymptotic distribution of the efficient score (K) statistic for θ_1 is the same for all θ_2 in the $\sqrt{n}$-neighborhood of the true value θ_{02}, and that proves sufficient to establish the asymptotic properties of the efficient projection-based K test.

The efficient K statistic

In the general context of CU-GMM based on (2.1), we can use the principle illustrated in the example above and define the estimated efficient score (gradient) for θ_1 as

[5] In this linear regression setup, $k = \nu$ and $-n^{-1}\widehat{D}_T(\theta) = n^{-1}X'X \to \Omega = V_{gg}(\theta)$. This simplifies the expressions considerably compared to those to be introduced in the general setup. For example, the objective function and its gradient with respect to θ now simplify to

$$\begin{aligned}Q_n(\theta) &= \frac{1}{2n}g_T'(\theta)\widehat{V}_{gg}(\theta)g_T(\theta) = \frac{1}{2}(y - X\theta)'P(X)(y - X\theta) \text{ and} \\ \nabla_\theta Q_n(\theta) &= \frac{1}{n}g_T'(\theta)\widehat{V}_{gg}(\theta)\widehat{D}_T(\theta) = -(y - X\theta)'X.\end{aligned}$$

$$\nabla_{1.2}Q_n(\theta) = \frac{1}{n} g_T'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta) N\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T2}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T1}(\theta).$$

Finally, using the same idea as in (2.6), we define the efficient score version of the K statistic, i.e. the efficient K statistic, as

$$\begin{aligned} K_{n1}(\theta) &= n\left(\nabla_{1.2}Q_n(\theta)\right)\left(\widehat{D}_{T1}'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta) N\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T2}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T1}(\theta)\right)^{-1}\left(\nabla_{1.2}Q_n(\theta)\right)' \\ &= \frac{1}{n} g_T'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta) P\left(N\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T2}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T1}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta) g_T(\theta). \end{aligned} \tag{2.7}$$

If the restricted CUE $\tilde{\theta}_*$ is such that $\nabla_2 Q_n(\tilde{\theta}_*) = 0$ (and not just w.p.a.1 as in (2.4)), then it follows that $K_n(\tilde{\theta}_*) = K_{n1}(\tilde{\theta}_*)$ once we note that the top-left $\nu_1 \times \nu_1$ block of $\left[\widehat{D}_T'(\theta)\widehat{V}_{gg}^{-1}(\theta)\widehat{D}_T(\theta)\right]^{-1}$ is $\left[\widehat{D}_{T1}'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta) N\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T2}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T1}(\theta)\right]^{-1}$. Hence Kleibergen's (subset-)K statistic can also be seen as a (normalized) quadratic form of the estimated efficient score for θ_1 evaluated at the restricted CUE $\tilde{\theta}_*$.

The following lemma gives the asymptotic properties of the efficient K statistic and some related statistics under $\sqrt{n}$-local alternatives.

Lemma 2.2 *Let $\theta_{ni} = \theta_{0i} + d_i/\sqrt{n} \in \Theta_i$ where $d_i \in \mathbb{R}^{\nu_i}$ for $i = 1, 2$; and let $\theta_n = (\theta_{n1}', \theta_{n2}')'$ and $d_\theta = (d_1', d_2')'$. Define $\Psi_{\nabla.g} = \Psi_\nabla - V_{\nabla g}(\theta_0)V_{gg}^{-1}(\theta_0)\Psi_g$. Let $\Psi_{\nabla.g}$ and $L(\theta)$ be partitioned with respect to θ_1 and θ_2 such that $\Psi_{\nabla.g} = \left[\Psi_{1.g}', \Psi_{2.g}'\right]'$ and $L(\theta) = [L_1'(\theta), L_2'(\theta)]'$. Then under Assumptions D and W,*

(i) $K_{n1}(\theta_n) \xrightarrow{d} \mathbb{B}' P\left(N\left(\mathbb{A}_2\right)\mathbb{A}_1\right)\mathbb{B}$

(ii) $\tilde{K}_{n2}(\theta_n) = n\left(\nabla_2 Q_n(\theta_n)\right)\left(\widehat{D}_{T2}'(\theta_n)\widehat{V}_{gg}^{-1}(\theta_n)\widehat{D}_{T2}(\theta_n)\right)^{-1}\left(\nabla_2 Q_n(\theta_n)\right)' \xrightarrow{d} \mathbb{B}' P\left(\mathbb{A}_2\right)\mathbb{B}$

(iii) $S_n(\theta_n) = 2Q_n(\theta_n) \xrightarrow{d} \mathbb{B}'\mathbb{B}$

where $\mathbb{A}_i = V_{gg}^{-\frac{1}{2}'}(\theta_0)\left[G_i(\theta_0) + (1 - 1_{[\delta_i=1]})devec_k\left[\Psi_{i.g} + L_i(\theta_0)d_\theta\right]'\right]$ for $i = 1, 2$ and $\mathbb{B} = V_{gg}^{-\frac{1}{2}'}(\theta_0)\left[\Psi_g + \sum_{i=1}^2 1_{[\delta_i=1]}M_i(\theta_{0i})d_i\right]$.

Lemma 2.2 is proved in the Appendix. We make the following remarks:

(i) $K_{n1}(\theta_0) \overset{A}{\sim} \chi^2_{\nu_1}$ in WI-Cases I-IV. Further, using Lemma 2.1 it follows that $K_n(\theta_{n1}, \tilde{\theta}_{n2}(\theta_{n1})) = K_{n1}(\theta_n) + o_p(1) = K_{n1}(\theta_{n1}, \theta_{02}) + o_p(1)$ in WI-Cases II and IV.

(ii) $\tilde{K}_{n2}(\theta_{*1}, \theta_{*2})$ is the K statistic for testing $H_2 : \theta_2 = \theta_{*2}$ when θ_1 is assumed to be equal to θ_{*1} (and hence no longer considered an unknown parameter). Because $\tilde{K}_{n2}(\theta_0) \overset{A}{\sim} \chi^2_{\nu_2}$, when the true value of θ_1 is known *a priori* then the test that rejects

$H_2 : \theta_2 = \theta_{*2}$ if $\tilde{K}_{n2}(\theta_{01}, \theta_{*2}) > \chi^2_{\nu_2}(1-\zeta)$ has asymptotic size ζ. In WI-Cases II and IV, $\tilde{K}_{n2}(\theta_n)$ converges to a non-central $\chi^2_{\nu_2}$ distribution with non-centrality parameter

$$\left[\sum_{i=1}^{2} 1_{[\delta_i=1]} M_i(\theta_{0i}) d_i\right]' V_{gg}^{-\frac{1}{2}}(\theta_0) P\left(V_{gg}^{-\frac{1}{2}\prime}(\theta_0) M_2(\theta_{02})\right) V_{gg}^{-\frac{1}{2}\prime}(\theta_0) \left[\sum_{i=1}^{2} 1_{[\delta_i=1]} M_i(\theta_{0i}) d_i\right]$$

which, under (2.1), can be finite only in the $\sqrt{n}$-neighborhood of θ_{02}.

(iii) $S_n(\theta_n)$ is the S statistic proposed by Stock and Wright (2000). In WI-Cases II and IV, $S_n(\theta_n)$ converges to a non-central χ^2_k distribution with non-centrality parameter

$$\left[\sum_{i=1}^{2} 1_{[\delta_i=1]} M_i(\theta_{0i}) d_i\right]' V_{gg}^{-1}(\theta_0) \left[\sum_{i=1}^{2} 1_{[\delta_i=1]} M_i(\theta_{0i}) d_i\right]$$

which, under (2.1), can be finite only in the $\sqrt{n}$-neighborhood of θ_{02}. When θ_{01} is known *a priori*, then the test that rejects $H_2 : \theta_2 = \theta_{*2}$ if $S_n(\theta_{01}, \theta_{*2}) > \chi^2_k(1-\zeta)$ has asymptotic size ζ.

The efficient projection-based K test

The results of Lemma 2.2 show that in WI-Cases II and IV (when θ_2 is identified), and against $\sqrt{n}$-local alternatives, the level-ϵ subset-K test for $H_1 : \theta_1 = \theta_{*1}$ is asymptotically equivalent to the *infeasible efficient K test* that rejects $H_1 : \theta_1 = \theta_{*1}$ at level ϵ when $K_{n1}(\theta_{*1}, \theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)$. The latter test uses the unknown true value of the nuisance parameters θ_2 and hence is infeasible. Based on these observations, we define and describe the efficient projection-based K test for the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ in Theorem 2.3.

Theorem 2.3 *Let $\theta_{*1} = \theta_{01} + d_1/\sqrt{n} \in \Theta_1$ where $d_1 \in \mathbb{R}^{\nu_1}$. For any $\theta_1 \in \Theta_1$, define $\mathcal{C}_2(1-\zeta, \theta_1) = \{\theta_2 \in \Theta_2 : S_n(\theta_1, \theta_2) \leq \chi^2_k(1-\zeta)\}$. Define the rejection rule for the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ using the efficient projection-based K test by the random variable $\phi_n(\theta_{*1}) \equiv \phi_n(\theta_{*1}; w_1, \ldots, w_n)$ such that*

$$\phi_n(\theta_{*1}) = \begin{cases} 1 & \text{if } \mathcal{C}_2(1-\zeta, \theta_{*1}) = \varnothing \quad \text{or if} \quad \inf_{\theta_{*2} \in \mathcal{C}_2(1-\zeta, \theta_{*1})} K_{n1}(\theta_{*1}, \theta_{*2}) > \chi^2_{n_1}(1-\epsilon) \\ 0 & \text{otherwise.} \end{cases}$$

Under Assumptions D and W,

(i) $\lim_{n\to\infty} E_{\theta_{01}} \phi_n(\theta_{01}) \leq \zeta + \epsilon$;

(ii) additionally if $\mathcal{C}_2(1-\zeta, \theta_{*1}) \neq \varnothing$, then in WI-Cases II and IV,

$$\lim_{n\to\infty} \left[E_{\theta_{01}} \phi_n(\theta_{*1}) - Pr_{\theta_{01}}\left[K_{n1}(\theta_{*1}, \theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)\right]\right] = 0.$$

Theorem 2.3 is the main result of the paper. Part(i) shows that the asymptotic size of this test is bounded from above by $\zeta + \epsilon$. In WI-Cases I and III (when θ_1 is weakly identified),

ζ and ϵ can be chosen such that the desired level of the test is not exceeded. Part(ii) shows that in WI-Cases II and IV, the choice of ζ becomes asymptotically irrelevant if the first step confidence region for θ_2 is nonempty. In such cases the efficient projection-based K test is asymptotically equivalent to the infeasible efficient K test that rejects $H_1 : \theta_1 = \theta_{*1}$ at level ϵ if $K_{n1}(\theta_{*1}, \theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)$.

A conservative $(1-\zeta-\epsilon) \times 100\%$ asymptotic confidence region for θ_1 can be obtained by inverting the efficient projection-based K test as

$$\begin{aligned}&\{\theta_1 \in \Theta_1 : \phi_n(\theta_1) = 0\}\\ &= \left\{\theta_1 \in \Theta_1 : \mathcal{C}_2(1-\zeta, \theta_1) \neq \varnothing, \inf_{\theta_2 \in \mathcal{C}_2(1-\zeta,\theta_1)} K_{n1}(\theta_1, \theta_2) \leq \chi^2_{\nu_1}(1-\epsilon)\right\}.\end{aligned}$$

In WI-Cases II and IV, the region's coverage (and length) is asymptotically equivalent to the asymptotic coverage (and length) of the infeasible region $\left\{\theta_1 \in \Theta_1 : K_{n1}(\theta_1, \theta_{02}) \leq \chi^2_{\nu_1}(1-\epsilon)\right\}$ if $\mathcal{C}_2(1-\zeta, \theta_1) \neq \varnothing$ for all $\theta_1 \in \Theta$.

The asymptotic equivalence under WI-Cases II and IV naturally extends to the subset-K test (see the discussion preceding Theorem 2.3). On the other hand, if $\mathcal{C}_2(1-\zeta, \theta_{*1})$ is empty with positive probability, the asymptotic equivalence in WI-Cases II and IV may not hold – the efficient projection-based K test can be more powerful than the infeasible efficient K test and the subset-K test at the cost of its asymptotic size exceeding ϵ and belonging to the interval $(\epsilon, \zeta+\epsilon)$.

The definition of the confidence region $\mathcal{C}_2(1-\zeta, \theta_{*1})$ in the statement of Theorem 2.3 has a major advantage. The underlying S test concurrently tests the model-specification (i.e. $Eg_t(\theta_{*1}, \theta_2) = 0$) and thus, under equation (2.1), rules out the spurious decline of power at all local-minima and saddle points (unless present in the $\sqrt{n}$-neighborhood of θ_0) of the objective function which is typical to tests based on the K statistic. Kleibergen proposed a two-step method, the subset-K-J test, to avoid the spurious decline in power. The subset-K-J test rejects the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ at level $\zeta + \epsilon$

$$\begin{aligned}&(i) \text{ if } K_n(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1})) > \chi^2_{\nu_1}(1-\epsilon)\\ &(ii) \text{ or if } J_n(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1})) = S_n(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1})) - K_n(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1})) > \chi^2_{k-\nu}(1-\zeta)\end{aligned}$$

[see Kleibergen (2005) and Kleibergen and Mavroeidis (2008a)]. Our choice of $\mathcal{C}_2(1-\zeta, \theta_{*1})$ achieves the same goal. Simulation results in Zivot and Chaudhuri (2008), based on the setting of Kleibergen and Mavroeidis (2008b), suggest that the two methods are practically indistinguishable when the nuisance parameters are identified.[6] Given that the usual projection-based tests for subsets of parameters are often criticized for being needlessly conservative, these observations certainly justify the new method of projection-

[6] Another choice, although not recommended, for the first step confidence region $\mathcal{C}_2(1-\zeta, \theta_{*1})$ is the region $\mathcal{C}_2^K(1-\zeta, \theta_{*1}) = \left\{\theta_{*2} \in \Theta_2 : \tilde{K}_{n2}(\theta_{*1}, \theta_{*2}) \leq \chi^2_{\nu_2}(1-\zeta)\right\}$ [see Lemma 2.2(ii)]. By definition of $\tilde{\theta}_{n2}(\theta_{*1})$, the statistic $\tilde{K}_{n2}(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1})) = 0$ w.p.a.1 and hence, such a first step confidence region for θ_2

based inference proposed in this paper.

Comparison with the usual method of projection

The basic requirements for using the new method of projection are – (i) a confidence region $\mathcal{C}_2(1-\zeta,\theta_1)$ for the nuisance parameters θ_2 with asymptotic coverage probability $(1-\zeta)$ when $\theta_1=\theta_{01}$, and (ii) a statistic $R_n(\theta_1,\theta_2)$ that is asymptotically pivotal when $\theta_1=\theta_{01}$ and $\theta_2=\theta_{02}$. The usual method of projection does not have the first requirement because it involves a projection from the entire parameter space Θ_2 of θ_2. The usual method of projection-based tests would reject the null hypothesis $H_1:\theta_1=\theta_{*1}$ if $\inf_{\theta_{*2}\in\Theta_2} R_n(\theta_{*1},\theta_{*2})$ is greater than the same critical value (used in the new method). It is clear from this observation that the usual method is going to be more conservative than the new method. However, the comparison is probably not fair because the usual method only allows for an asymptotic size of (at most) ϵ, where as this is (at most) $(\zeta+\epsilon)$ for the new method. The difference in the asymptotic size is the price for the reduction in conservativeness by restricting the unknown nuisance parameters θ_2 in $R_n(\theta_1,\theta_2)$ to vary only in a small neighborhood $\mathcal{C}_2(1-\zeta,\theta_1)$ possibly containing the true value θ_{02}.

This is where the choice of $R_n(\theta_1,\theta_2)=K_{n1}(\theta_1,\theta_2)$, the efficient K statistic, becomes important – it ensures that the price in terms of the increased asymptotic size ζ disappears asymptotically whenever the nuisance parameters are identified. In such cases, one can always obtain a nonempty confidence region in the first step and thus *restrict the asymptotic size of the new projection-based test not to exceed* ϵ without any loss of asymptotic power (as compared to the infeasible efficient K test or the subset-K test). In particular, when the nuisance parameters are identified, the efficient K statistic makes the new method of projection-based test (locally) asymptotically equivalent to the level-ϵ infeasible efficient K test, which under standard conditions (i.e. WI-Case IV) is asymptotically equivalent to the usual (level-ϵ) Wald, likelihood ratio and score tests, and thus attains local asymptotic optimality among level-ϵ tests for the null hypothesis $H_1:\theta_1=\theta_{*1}$ that treat the nuisance parameters θ_2 as unknown.

We are not aware of any other form of the second step test statistic that would ensure that the asymptotic size of the new (restricted) projection-based test does not exceed ϵ when the nuisance parameters are identified and subsequently allow for a comparison with other tests (including the usual projection-based) tests that are designed to have size less than equal to ϵ.

Given the superior performance of the GMM-MLR test [see Kleibergen (2005) and Kleibergen and Mavroeidis (2008b)], it is natural to consider an extension of the new

is nonempty w.p.a.1. This further implies that

$$K_n(\theta_{*1},\tilde{\theta}_{n2}(\theta_{*1}))\geq \inf_{\theta_{*2}\in\mathcal{C}_2^K(1-\zeta,\theta_{*1})} K_{n1}(\theta_*)$$

and hence the power (size) of the subset K-test for $H_1:\theta_1=\theta_{*1}$ dominates the power (size) of the efficient projection-based K test when this region is used in the first step.

method of projection to the (quasi) likelihood ratio type tests. However, because the use of the efficient K statistic is crucial for proving the desirable properties of the new test, it is not obvious how such an extension can be made possible.

Finally it should be noted that, if it is possible to obtain a $\sqrt{n}$-consistent point estimator for a subset of the nuisance parameters, the computational cost of the efficient projection-based K test can be reduced substantially by using this estimator and restricting the search for the infimum of the efficient K statistic to the remaining nuisance parameters only.

3 Simulation study in a linear IV regression

In this section we present a Monte Carlo study in a linear instrumental variables model and show that the asymptotic results from Section 2 provide a good approximation to the behavior of the efficient projection-based K test in finite samples. The framework is similar to that of Kleibergen (2004) and Zivot et al. (2006).

Model

Consider the following model:

$$y = X_1\theta_{01} + X_2\theta_{02} + u \tag{3.1}$$

$$X_1 = Z\Pi_1 + \eta_1 \tag{3.2}$$

$$X_2 = Z\Pi_2 + \eta_2 \tag{3.3}$$

where y, X_1 and X_2 are respectively the $n \times 1$, $n \times \nu_1$ and $n \times \nu_2$ matrices of endogenous variables; u, η_1 and η_2 are respectively the $n\times 1$, $n\times\nu_1$ and $n\times\nu_2$ matrices of the unobserved correlated structural errors. Z is the non-stochastic and full-column rank $n \times k$ matrix of instruments, and the order condition $k \geq \nu = \nu_1+\nu_2$ is assumed to be satisfied. We assume that the structural errors are i.i.d., i.e. the t-th row of u, η_1 and η_2 is described by

$$(u_t, \eta_{1t}, \eta_{2t}) \overset{\text{i.i.d.}}{\sim} \left(0, \Sigma = \begin{bmatrix} \sigma_{uu} & \sigma_{u1} & \sigma_{u2} \\ \sigma_{1u} & \sigma_{11} & \sigma_{12} \\ \sigma_{2u} & \sigma_{21} & \sigma_{22} \end{bmatrix}\right) \quad \text{for} \quad t = 1, \ldots, n.$$

The parameters of interest are θ_1. Under the above assumptions, the parameters Π_1 and Π_2 can be consistently estimated as long as $\lim_{n\to\infty} n^{-1}Z'Z$ is positive definite, and the parameters Σ can be consistently estimated if, in addition, θ_{01} and θ_{02} (or their consistent estimators) are known. Hence we can partial out the parameters Π_1, Π_2 and Σ, and treat only θ_2 as the nuisance parameters for which a first step confidence region needs to be constructed in the efficient projection-based K test. This reduces the computational cost considerably. The model described in (3.1) – (3.3) is a special case of the general GMM framework and all the results from Section 2 apply in this setup [see, for example, Chaudhuri (2008)]. In the rest of this section we will see that the finite sample rejection rates of the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ are very similar for the subset-K test and the

efficient projection-based K test, especially when the nuisance parameters are identified.

Monte Carlo specifications

The true values of the structural coefficients are taken as $\theta_{01} = 1$ and $\theta_{02} = 10$. The structural errors $[u, \eta_1, \eta_2]$ are generated by drawing $n = 100$ independent random samples from $\mathcal{N}(0, \Sigma)$ where

$$\Sigma = \begin{pmatrix} 1 & \rho_{u1} & \rho_{u2} \\ \rho_{u1} & 1 & 0 \\ \rho_{u2} & 0 & 1 \end{pmatrix}. \tag{3.4}$$

If η_1 and η_2 are correlated, the level of endogeneity of the regressor X_1 depends on the correlations between [η_1 and u], [η_1 and η_2] and [η_2 and u]. Our choice of Σ in (3.4) simplifies the setup by ensuring that the level of endogeneity of X_1 depends only on ρ_{u1} and similarly the level of endogeneity of X_2 depends only on ρ_{u2}. We make three different choices for the pair (ρ_{u1}, ρ_{u2}): $(\rho_{u1}, \rho_{u2}) = (0.5, 0.5), (0.1, 0.99), (0.99, 0.1)$. X_1 and X_2 are moderately (and equally) endogenous in the first case, X_1 is highly endogenous and X_2 is mildly endogenous in the second case, X_1 is mildly endogenous and X_2 is highly endogenous in the third case. We refer to the corresponding covariance matrices of the structural errors as Σ_1, Σ_2 and Σ_3 respectively.

The instruments Z are generated by drawing n independent random samples from $\mathcal{N}(0, I_{k-1})$ independently of the structural errors and appending the matrix with an $n \times 1$ column vector $(1, \ldots, 1)'$. We consider three different choices of the number of instruments: $k = 2, 4, 20$. The first choice gives a just-identified model and the latter two give over-identified models. The large value $k = 20$ is taken because when the nuisance parameters are unidentified, the limiting null distribution of the subset-K statistic converges to a $\chi^2_{\nu_1}$ distribution from below as $k \to \infty$ and $n \to \infty$ such that $k/n \to 0$, and hence the power of the subset-K test is likely to be the largest for this choice [see Kleibergen (2007)].

To our knowledge, there is no universally accepted measure of instrumental relevance for individual structural coefficients in a linear IV model with more than one endogenous regressor. However, for a model with a single endogenous regressor, the instruments are considered weak for the structural coefficient if the concentration parameter is less than 10 [see Staiger and Stock (1997)]. We follow Zivot et al. (2006) and choose the parameters Π_1 and Π_2 such that the concentration matrix μ, as defined by Stock and Yogo (2005), is diagonal where for $i = 1, 2$, the i-th diagonal element μ_i corresponds to the concentration parameter for θ_i. The weak instrument setup for the experiment is summarized in Table 2.

Results

The results reported are based on 10,000 Monte Carlo trials. The instrument matrix Z is kept fixed over the trials. We compute the empirical rejection rates of the efficient projection-based K test, the subset-K test and the projection-based S test for a grid of θ_{*1} values around the true value θ_{01}. For the efficient projection-based K test we consider

Table 2: Four Cases of Weak Instruments.

	$\mu_2 = 1$	$\mu_2 = 10$
$\mu_1 = 1$	**WI-Case I** θ_1 : weak instrument θ_2 : weak instrument	**WI-Case II** θ_1 : weak instrument θ_2 : strong instrument
$\mu_1 = 10$	**WI-Case III** θ_1 : strong instrument θ_2 : weak instrument	**WI-Case IV** θ_1 : strong instrument θ_2 : strong instrument

$\zeta = 0.01$ and $\zeta = 0.05$ in the construction of the first step confidence region $C_2(1-\zeta, \theta_{*1})$, and use $\epsilon = 0.05$ for the test in the second step. In the figures we refer to the resulting tests with these values as "New Test (1% + 5%)" and "New Test (5% + 5%)", respectively. For the subset-K and projection-based S tests we use a nominal size $\epsilon = 0.05$. In our linear IV setup, the S test is the Anderson-Rubin (AR) test and we compute the projection-based AR test using the results of Dufour and Taamouti (2005). For brevity, we only present a subset of the results in Table 3 and Figures 1 – 3 to illustrate our main points. The full set of results are available in the supplement appendix.

[**Insert Figures 1 – 3 and Table - 3 around here.**]

In general across the four WI cases, the empirical rejection rates of the subset-K test and the efficient projection based K tests are very close and the projection-based AR test is the most conservative.

In the cases when the nuisance parameter θ_2 is (strongly) identified (see WI-Case II and WI-Case IV in the right panel of the figures), Theorem 2.3 states that the efficient projection-based K tests and the subset-K test are asymptotically equivalent provided $C_2(1-\zeta, \theta_{*1}) \neq \varnothing$. This result is borne out in our simulations as illustrated in Figure 1. The difference in behavior between the efficient projection-based K tests for the two choices of ζ arises solely from the difference in the frequency of $C_2(1-\zeta, \theta_{*1}) = \varnothing$, which is summarized in Table 3. To see this result in the figures, consider the point where the null hypothesis is true (i.e., $\theta_{*1} - \theta_{01} = 0$). In Figures 2 and 3 the empirical rejection rates of the efficient projection-based K tests exceed the nominal size $\epsilon = 0.05$. From Table 3 it can be seen that the amount by which they exceed the nominal size ϵ is (roughly) the same as the frequencies (probabilities) of $C_2(1-\zeta, \theta_{*1}) = \varnothing$. This result also explains the difference between the rejection rate of the level-ϵ subset-K test and the efficient projection-based K tests. Empty confidence regions occur when the over-identification restrictions are rejected under the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ by the AR test in the first step. Table 3 shows that the probability of such occurrences increases with the order of over-identification of

the restricted (by $H_1 : \theta_1 = \theta_{*1}$) model. The occurrence of empty confidence regions for θ_2 in the first step increases the rate at which the efficient projection-based K tests reject the null hypothesis $H_1 : \theta_1 = \theta_{*1}$. As can be seen from Figures 2 and 3, the benefit of such an increased rejection rate in terms of increase in power (relative to the subset-K test) probably outweighs its cost from increase in size (above $\epsilon = 0.05$).

When the nuisance parameter θ_2 is weakly identified (see WI-Case I and WI-Case II in the left panel of the figures), Theorem 2.3 states that the asymptotic size of the efficient projection-based K test is bounded from above by $(\zeta + \epsilon)$. The simulations show that, even with the relatively large choice $\zeta = 0.05$, the efficient projection-based K test does not tend to over-reject provided the number of instruments is not too large relative to the sample size and/or the level of endogeneity of the regressor X_2 associated with the nuisance parameter θ_2 is not extremely high.

Overall, the simulations corroborate the asymptotic properties of the efficient projection-based K test summarized in Theorem 2.3. In a comment to Kleibergen and Mavroeidis (2008b), Zivot and Chaudhuri (2008) presented Monte Carlo evidence that showed the efficient projection-based K test performed similarly to the level-ϵ subset-GMM-MLR test and the level-$(\zeta + \epsilon)$ subset-K-J test in a simulation design calibrated to mimic data used to estimate a typical new Keynesian Phillips curve.

4 Conclusion

In this paper we questioned the common perception that the projection-based tests are conservative, and subsequently showed that proper use of projection techniques could lead to tests that are comparable to the tests based on the plug in principle.

We proposed a new projection-based test for subsets of parameters in the context of GMM with possibly weakly identified parameters. We showed that two simple modifications to the usual projection-based tests can considerably reduce their conservativeness. The modifications – (i) a restricted projection and (ii) the use of an efficient score equivalent of Kleibergen's K statistic, allowed us to design a test that can be asymptotically equivalent to Kleibergen's subset-K test whenever the nuisance parameters are (strongly) identified and at the same time would not ever result in an uncontrolled over-rejection of the true value of the parameters of interest.

While we introduced the new method of projection in the context of GMM estimation, the method is more generally applicable to any estimation technique that admits a score-type statistic whose distribution is asymptotically pivotal when evaluated at the true values of the parameters. For example, Chaudhuri et al. (2008) applied the new method of projection to inference on subsets of parameters in a split-sample two-stage-least-squares context and Chaudhuri (2008) described the method in the general extremum estimation context. Further applications of the new method are the subject of our future research.

A Appendix

Proof of Lemma 2.1: The following proof follows from Stock and Wright (2000) with minor modifications. Note that in WI-Cases II and IV,

(a) $En^{-1}g_T(\theta) = 1_{[\delta_1=1]}m_1(\theta_1) + 1_{[\delta_1=\frac{1}{2}]}\frac{1}{\sqrt{n}}\tilde{m}_{n1}(\theta) + m_2(\theta_2)$ where $m_1(\theta_1) \to m_1(\theta_{01})$ for $\theta_1 \to \theta_{01}$ and $\tilde{m}_{n1}(\theta) \to \tilde{m}_1(\theta)$ uniformly in $\theta \in \Theta$, and

(b) $\widehat{V}_{gg}^{-1}(\theta) \xrightarrow{P} V_{gg}^{-1}(\theta)$ uniformly where $V_{gg}^{-1}(\theta)$ is positive definite, continuous and bounded in $\theta \in \Theta$.

Therefore, it follows from Assumption W that in WI-Cases II and IV,

$$\frac{1}{n}Q_n(\theta_{*1}, \theta_2) \xrightarrow{P} m_2'(\theta_2)V_{gg}^{-1}(\theta_{01}, \theta_2)m_2(\theta_2)$$

uniformly in $\theta_2 \in \Theta_2$. The right-hand side is zero if and only if $\theta_2 = \theta_{02}$ and hence continuity of the argmin operator gives

$$\tilde{\theta}_{n2}(\theta_{*1}) \xrightarrow{P} \theta_{02}. \tag{A.1}$$

Let $\tilde{\theta}_* = (\theta_{*1}', \tilde{\theta}_{n2}'(\theta_{*1}))'$ and $\theta_{*0} = (\theta_{*1}', \theta_{02}')'$. By definition of CUE $\tilde{\theta}_{n2}(\theta_{*1})$,

$$\begin{aligned} 0 \geq Q_n(\tilde{\theta}_*) - Q_n(\theta_{*0}) \quad &= \quad \left[n^{-1/2}\nabla_\theta g_T(\bar{\theta})(\tilde{\theta}_* - \theta_0)\right]' \widehat{V}_{gg}^{-1}(\tilde{\theta}_*) \left[n^{-1/2}\nabla_\theta g_T(\bar{\theta})(\tilde{\theta}_* - \theta_0)\right] \\ &\quad + \Delta_{1n} + 2\left[n^{-1/2}\nabla_\theta g_T(\bar{\theta})(\tilde{\theta}_* - \theta_0)\right]' \widehat{V}_{gg}^{-1}(\tilde{\theta}_*)n^{-1/2}g_T(\theta_0) \end{aligned} \tag{A.2}$$

where the mean-value $\bar{\theta} \in \Theta$ is such that $\|\bar{\theta} - \theta_0\| \leq \|\tilde{\theta}_* - \theta_0\| = o_p(1)$ and

$$\Delta_{1n} = n^{-1/2}g_T'(\theta_0)\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)n^{-1/2}g_T(\theta_0) - n^{-1/2}g_T'(\theta_{*0})\widehat{V}_{gg}^{-1}(\theta_{*0})n^{-1/2}g_T(\theta_{*0}).$$

For notational convenience define $\mathcal{M} \equiv \mathcal{M}(\tilde{\theta}_*, \bar{\theta}, \theta_0) = \left[n^{-1/2}\nabla_\theta g_T(\bar{\theta})(\tilde{\theta}_* - \theta_0)\right]$. For any square matrix A let mineval(A) denote its minimum eigen-value. Note that,

$$\begin{aligned} &\text{(a)} \qquad \mathcal{M}'\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\mathcal{M} \geq \|\mathcal{M}\|^2 \text{mineval}\left(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\right) \text{ and} \\ &\text{(b)} \qquad \mathcal{M}'\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)g_T(\theta_0)/\sqrt{n} \geq -\|\mathcal{M}\| \left\|\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)g_T(\theta_0)/\sqrt{n}\right\|, \end{aligned}$$

the latter following from the Cauchy-Schwartz inequality.

Now define $\Delta_{2n} = \dfrac{\left\|\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)g_T(\theta_0)/\sqrt{n}\right\|}{\text{mineval}\left(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\right)}$ and $\Delta_{3n} = \dfrac{\Delta_{1n}}{\text{mineval}\left(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\right)}$. Therefore, dividing (A.2) by mineval$\left(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\right)$, we get,

$$\|\mathcal{M}\|^2 - 2\|\mathcal{M}\|\Delta_{2n} + \Delta_{3n} \leq 0$$

which implies that $\Delta_{2n} - \sqrt{\Delta_{2n}^2 - \Delta_{3n}} \leq \|\mathcal{M}\| \leq \Delta_{2n} + \sqrt{\Delta_{2n}^2 - \Delta_{3n}}$.

Noting that $\|\bar{\theta} - \theta_0\| = o(1)$, Assumptions D1 and W give $n^{-1}\nabla_i g_T(\bar{\theta}) \to 1_{[\delta_i=1]} M_i(\theta_{0i})$ for $i = 1, 2$. Since $\|d_1\| = O(1)$, it is clear that $\sqrt{n}\left(\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02}\right) = O_p(1)$ if Δ_{2n} and Δ_{3n} are $O_p(1)$.

Now we verify that Δ_{2n} and Δ_{3n} are $O_p(1)$. First note that under Assumption D,

$$\Delta_{2n} \leq \frac{\sup_\theta \left\|\widehat{V}_{gg}^{-1}(\theta) g_T(\theta_0)/\sqrt{n}\right\|}{\inf_\theta \text{mineval}\left(\widehat{V}_{gg}^{-1}(\theta)\right)} \xrightarrow{d} \frac{\sup_\theta \left\|V_{gg}^{-1}(\theta)\Psi_g\right\|}{\inf_\theta \text{mineval}\left(V_{gg}^{-1}(\theta)\right)} = O_p(1). \tag{A.3}$$

Again, noting that, for some $\bar{\theta}_{10} = (\bar{\theta}_1', \theta_{02}')'$ such that $\|\bar{\theta}_{10} - \theta_0\| \leq \|\theta_{*0} - \theta_0\| = O(1/\sqrt{n})$, i.e. for some $\bar{\theta}_1 = \theta_{01} + \bar{d}_1/\sqrt{n}$ where $\|\bar{d}_1\| \leq \|d_1\| = O(1)$,

$$\begin{aligned}
&|\Delta_{1n}| \\
&= \left|\frac{g_T'(\theta_0)}{\sqrt{n}}\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\frac{g_T(\theta_0)}{\sqrt{n}} - \left[\frac{g_T(\theta_0)}{\sqrt{n}} + \frac{\nabla_1 g_T(\bar{\theta}_{10})}{n}\bar{d}_1\right]' \widehat{V}_{gg}^{-1}(\theta_{*0})\left[\frac{g_T(\theta_0)}{\sqrt{n}} + \frac{\nabla_1 g_T(\bar{\theta}_{10})}{n}\bar{d}_1\right]\right| \\
&\leq \left|\frac{g_T'(\theta_0)}{\sqrt{n}}\left[\widehat{V}_{gg}^{-1}(\tilde{\theta}_*) - \widehat{V}_{gg}^{-1}(\theta_{*0})\right]\frac{g_T(\theta_0)}{\sqrt{n}}\right| + 2\left|\frac{g_T'(\theta_0)}{\sqrt{n}}\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\frac{\nabla_1 g_T(\bar{\theta}_{10})}{n}\bar{d}_1\right| \\
&\quad + \left|\bar{d}_1'\frac{\nabla_1 g_T'(\bar{\theta}_{10})}{n}\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\frac{\nabla_1 g_T(\bar{\theta}_{10})}{n}\bar{d}_1\right| \\
&\xrightarrow{d} 2 \times 1_{[\delta_1=1]}\left[\left|\Psi_g V_{gg}^{-1}(\theta_0) M_1(\theta_{01})\bar{d}_1\right| + \left|\bar{d}_1' M_1'(\theta_{01}) V_{gg}^{-1}(\theta_0) M_1(\theta_{01})\bar{d}_1\right|\right] = O_p(1)
\end{aligned}$$

follows from Assumptions D and W. Since $V_{gg}^{-1}(\theta)$ is positive definite, similar arguments as in (A.3) give $|\Delta_{3n}| = O_p(1)$. Hence $\sqrt{n}\|\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02}\| = O_p(1)$. ■

Lemmas A.1 are A.2 will be helpful in getting the rest of the results.

Lemma A.1 *Let $\widehat{a}_n(.)$ and $a(.)$ be $p_a \times p$ and $\widehat{b}_n(.)$ and $b(.)$ be $p \times p_b$ finite-dimensional matrices. Let $\theta_0 \in$interior(Θ) where Θ is compact. Then following results hold as $n \to \infty$:*

(i) Let $\widehat{a}_n(\theta) - a_n(\theta) = o_p(1)$ and $a_n(\theta) - a(\theta) = o(1)$ for $\theta \in \Theta$. Then $\widehat{a}_n(\theta_n) - a(\theta_0) = o_p(1)$ if $a(\theta)$ is continuous at θ_0 and if $\theta_n - \theta_0 = o_p(1)$.

(ii) In addition, let $\widehat{b}_n(\theta) - b_n(\theta) = o_p(1)$ and $b_n(\theta) - b(\theta) = o(1)$ for $\theta \in \Theta$. If $a(\theta)$ and $b(\theta)$ are bounded on Θ, then $\widehat{a}_n(\theta_n)\widehat{b}_n(\theta_n) - a(\theta_0)b(\theta_0) = o_p(1)$ if $a(\theta)$ and $b(\theta)$ are continuous at θ_0 and if $\theta_n - \theta_0 = o_p(1)$.

Sketch of Proof: (i) Using the Triangle inequality, the result follows once we note that for n large enough $\theta_n \in \Theta$ w.p.a.1 and $\|\widehat{a}_n(\theta_n) - a(\theta_0)\| \leq \|\widehat{a}_n(\theta) - a_n(\theta)\| + \|a_n(\theta) - a(\theta)\| + \|a(\theta) - a(\theta_0)\| = o_p(1)$.

(ii) For $x = a, b$, let us define the index $\mathcal{I}_x = \{(i,j) : i = 1, \ldots, p_x \text{ and } j = 1, \ldots, p\}$ and let $\sup_{\theta\in\Theta} \max_{(i,j)\in\mathcal{I}_x} x_{(i,j)}(\theta) \leq R_x = O(1)$. Then the result follows using the same technique as in (i) once we note that the Triangle inequality and the Cauchy-Schwartz inequality give $\|\widehat{a}_n(\theta)\widehat{b}_n(\theta) - a(\theta)b(\theta)\| \leq \|\widehat{a}_n(\theta) - a(\theta)\|\|\widehat{b}_n(\theta) - b(\theta)\| + \sqrt{p_a p}|R_a|\|\widehat{b}_n(\theta) - b(\theta)\| + \|\widehat{a}_n(\theta) - a(\theta)\|\sqrt{pp_b}|R_b| = o_p(1)$. ■

Lemma A.2 *Let $\theta_{ni} = \theta_{0i} + d_i/\sqrt{n} \in \Theta_i$ where $d_i \in \mathbb{R}^{\nu_i}$ for $i = 1, 2$; and let $\theta_n = (\theta'_{n1}, \theta'_{n2})'$ and $d_\theta = (d'_1, d'_2)'$. Define $\Psi_{\nabla.g} = \Psi_\nabla - V_{\nabla g}(\theta_0)V_{gg}^{-1}(\theta_0)\Psi_g$. Let $\Psi_{\nabla.g}$ and $L(\theta)$ be partitioned with respect to θ_1 and θ_2 such that $\Psi_{\nabla.g} = \left[\Psi'_{1.g}, \Psi'_{2.g}\right]'$ and $L(\theta) = [L'_1(\theta), L'_2(\theta)]'$. Define $\widehat{h}_{Ti}(\theta) = \sum_{t=1}^n \widehat{h}_{ti}(\theta)$ where $\widehat{h}_{ti}(\theta) = \left[vec\nabla_i g_t(\theta) - \widehat{V}_{ig}(\theta)\widehat{V}_{gg}^{-1}(\theta)g_t(\theta)\right]$ for $i = 1, 2$. Under Assumptions D and W,*

$$\begin{bmatrix} n^{-1/2}g_T(\theta_n) \\ n^{-\delta_1}\widehat{h}_{T1}(\theta_n) \\ n^{-\delta_2}\widehat{h}_{T2}(\theta_n) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Psi_g + \sum_{i=1}^2 1_{[\delta_i=1]}M_i(\theta_{0i})d_i \\ vecG_1(\theta_0) + (1 - 1_{[\delta_1=1]})\left[\Psi_{1.g} + L_1(\theta_0)d_\theta\right] \\ vecG_2(\theta_0) + (1 - 1_{[\delta_2=1]})\left[\Psi_{2.g} + L_2(\theta_0)d_\theta\right] \end{bmatrix}.$$

Proof: Define $V_{\nabla\nabla.g}(\theta) = V_{\nabla\nabla}(\theta) - V_{\nabla g}(\theta)V_{gg}^{-1}(\theta)V_{g\nabla}(\theta)$. Following the obvious partition with respect to θ_1 and θ_2, let $V_{\nabla g} = [V'_{1g}, V'_{2g}]'$, $V_{\nabla.g} = [V'_{1.g}, V'_{2.g}]'$ and for $i = 1, 2$, let $h_{Ti}(\theta) = \sum_{t=1}^n h_{ti}(\theta)$ where $h_{ti}(\theta) = \left[vec\nabla_i g_t(\theta) - V_{ig}(\theta_0)V_{gg}^{-1}(\theta_0)g_t(\theta)\right]$. Letting $h_T(\theta) = [h'_{T1}(\theta), h'_{T2}(\theta)]'$, Assumptions D and W give

$$\frac{1}{\sqrt{n}}\begin{bmatrix} g_T(\theta_0) \\ h_T(\theta_0) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Psi_g \\ \Psi_{\nabla.g} \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} V_{gg}(\theta_0) & 0 \\ 0 & V_{\nabla\nabla.g}(\theta_0) \end{bmatrix}\right) \text{ and hence}$$

$$\frac{1}{\sqrt{n}}g_T(\theta_0) \xrightarrow{d} \Psi_g \text{ and for } i = 1, 2 \ \frac{1}{n^{\delta_i}}h_{Ti}(\theta_0) \xrightarrow{d} vecG_i(\theta_0) + (1 - 1_{[\delta_i=1]})\Psi_{i.g}. \tag{A.4}$$

A mean-value expansion of gives $\frac{1}{\sqrt{n}}g_T(\theta_n) = \frac{1}{\sqrt{n}}g_T(\theta_0) + \frac{1}{n}\nabla_\theta g_T(\bar{\theta})d_\theta$ for some $\bar{\theta}$ such that $\|\bar{\theta} - \theta_0\| \leq \|\theta_n - \theta_0\| = O(1/\sqrt{n})$. Hence, using (A.4) and Assumption W, we get

$$\frac{1}{\sqrt{n}}g_T(\theta_n) = \Psi_g + \sum_{i=1}^2 1_{[\delta_i=1]}M_i(\theta_{0i})d_i + o_p(1).$$

Using Lemma A.1 and the fact that continuity is preserved by matrix inversion, for $i = 1, 2$,

$$\begin{aligned} \frac{1}{n^{\delta_i}}\widehat{h}_{Ti}(\theta_n) &= \frac{1}{n^{\delta_i}}\left[vec\nabla_i g_T(\theta_n) - \widehat{V}_{ig}(\theta_n)\widehat{V}_{gg}^{-1}(\theta_n)g_T(\theta_n)\right] \\ &= \frac{1}{n^{\delta_i}}\left[vec\nabla_i g_T(\theta_n) - V_{ig}(\theta_0)V_{gg}^{-1}(\theta_0)g_T(\theta_n)\right] + o_p(1) \\ &= \frac{1}{n^{\delta}}h_{Ti}(\theta_n) + o_p(1) \\ &= \frac{1}{n^{\delta_i}}h_{Ti}(\theta_0) + \frac{1}{n^{\delta_i+\frac{1}{2}}}\nabla_\theta h_{Ti}(\bar{\theta})d_\theta + o_p(1) \end{aligned}$$

for some $\bar{\theta}$ such that $\|\bar{\theta}-\theta_0\| \leq \|\theta_n-\theta_0\| = O(1/\sqrt{n})$ following from a mean-value expansion of $h_{Ti}(\theta_n)$. Hence Assumption D2, Lemma A.1 and (A.4) give for $i = 1, 2$,

$$\frac{1}{n^{\delta_i}}\widehat{h}_{Ti}(\theta_n) = \frac{1}{n^{\delta i}}h_{Ti}(\theta_0)+(1-1_{[\delta_i=1]})L_i(\theta_0)d_\theta+o_p(1) \xrightarrow{d} vecG_i(\theta_0)+(1-1_{[\delta_i=1]})\left[\Psi_{i.g} + L_i(\theta_0)d_\theta\right] . \blacksquare$$

Proof of Lemma 2.2: Define $A_{ni} = \widehat{V}_{gg}^{-\frac{1}{2}'}(\theta_T)\widehat{D}_{Ti}(\theta_n)$ for $i = 1, 2$. It follows from Lemma A.2 that $\frac{1}{n^{\delta_i}}A_{ni} \xrightarrow{d} V_{gg}^{-\frac{1}{2}'}(\theta_0)\left[G_i(\theta_0) + (1 - 1_{[\delta_i=1]})devec_k\left[\Psi_{i.g} + L_i(\theta_0)d_\theta\right]'\right]$ and

$$(A_{n1}'N(A_{n2})A_{n1})^{-\frac{1}{2}'} A_{n1}'N(A_{n2})\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta_n)\frac{g_T(\theta_n)}{\sqrt{n}} \xrightarrow{d} (\mathbb{A}_1'N(\mathbb{A}_2)\mathbb{A}_1)^{-\frac{1}{2}'} \mathbb{A}_1'N(\mathbb{A}_2)\mathbb{B}, \tag{A.5}$$

$$(A_{n2}'A_{n2})^{-\frac{1}{2}'} A_{n2}'\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta_n)\frac{g_T(\theta_n)}{\sqrt{n}} \xrightarrow{d} (\mathbb{A}_2'\mathbb{A}_2)^{-\frac{1}{2}'} \mathbb{A}_2'\mathbb{B}, \text{ and} \tag{A.6}$$

$$\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta_n)\frac{g_T(\theta_n)}{\sqrt{n}} \xrightarrow{d} \mathbb{B} \tag{A.7}$$

Lemma 2.2 follows directly from (A.5), (A.6) and (A.7). $\blacksquare$

Proof of Theorem 2.3: In the following, whenever we refer to $\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_1)} K_{n1}(\theta_1, \theta_{*2})$, it is implied that $\mathcal{C}_2(1 - \zeta, \theta_1)$ is nonempty.

(i) From Lemma 2.2(iii), it is clear that $\mathcal{C}_2(1 - \zeta, \theta_{01})$ contains θ_{02} with probability $(1 - \zeta)$. Therefore, using Lemma 2.2(i), it follows that the asymptotic size of the efficient projection-based K test is

$$\begin{aligned}
& \lim_{n\to\infty} E_{\theta_{01}}\phi_n(\theta_{01}) \\
= & \lim_{n\to\infty} Pr_{\theta_{01}}\left[\{\mathcal{C}_2(1-\zeta,\theta_{01}) = \varnothing\} \cup \{\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{01})} K_{n1}(\theta_{01},\theta_{*2}) > \chi^2_{\nu_1}(1-\epsilon)\}\right] \\
\leq & \; 1 - \lim_{n\to\infty} Pr_{\theta_{01}}\left[\{\theta_{02} \in \mathcal{C}_2(1-\zeta,\theta_{01})\} \cap \{\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{01})} K_{n1}(\theta_{01},\theta_{*2}) \leq \chi^2_{\nu_1}(1-\epsilon)\}\right] \\
= & \; 1 - \lim_{n\to\infty} Pr_{\theta_{01}}\left[\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{01})} K_{n1}(\theta_{01},\theta_{*2}) \leq \chi^2_{\nu_1}(1-\epsilon) \mid \theta_{02} \in \mathcal{C}_2(1-\zeta,\theta_{01})\right] \\
& \times \lim_{n\to\infty} Pr_{\theta_{01}}\left[\, \theta_{02} \in \mathcal{C}_2(1-\zeta,\theta_{01})\right] \\
\leq & \; 1 - \lim_{n\to\infty} Pr_{\theta_{01}}\left[K_{n1}(\theta_{01},\theta_{02}) \leq \chi^2_{\nu_1}(1-\epsilon)\right] \lim_{n\to\infty} Pr_{\theta_{01}}\left[\, \theta_{02} \in \mathcal{C}_2(1-\zeta,\theta_{01})\right] \\
\leq & \; 1 - (1-\epsilon)(1-\zeta) \\
\leq & \; \zeta + \epsilon.
\end{aligned}$$

(ii) Lemma 2.2(iii) also implies that in WI-Cases II and IV, $\mathcal{C}_2(1 - \epsilon, \theta_{*1})$ is contained in the $\sqrt{n}$-neighborhood of θ_{02} w.p.a.1 under the conditions of the Theorem. Hence $\theta_2^{\inf}(\theta_{*1})$, where the infimum $\inf_{\theta_{*2}\in\mathcal{C}_2(1-\zeta,\theta_{*1})} K_{n1}(\theta_{*1}, \theta_{*2})$ is attained, is also in the $\sqrt{n}$-neighborhood of θ_{02}. Hence Lemma 2.2(i) directly applies and gives the local asymptotic equivalence of the tests. $\blacksquare$

Acknowledgements

Discussions with James Robins and Thomas Richardson were very important for this paper. We thank Eric Renault, Frank Kleibergen, Jean-Marie Dufour and Sophocles Mavroeidis for their helpful comments and feedback. We also thank the seminar participants at NCSU, UNC Chapel Hill, Queen's and Waterloo and the session participants in the ESRC Study Group Conference (2007), Midwest Econometrics Conference (2007) and the Winter meeting of the Econometric Society (2008) for their feedback. Eric Zivot gratefully acknowledges the financial support from the Gary Waterman Distinguished Scholar Fund.

References

Andrews, D. W. K. (1994). Empirical Process Methods in Econometrics. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume IV, chapter 37, pages 2247–2294. Elsevier Science Publisher.

Bera, A. K. and Bilias, Y. (2001). Rao's score, Neyman's $C(\alpha)$ and Silvey's LM tests: an essay on historical developments and some new results. *Journal of Statistical Planning and Inference*, 97: 9–44.

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 Paramaters*. PhD thesis, University of Washington.

Chaudhuri, S., Richardson, T., Robins, J., and Zivot, E. (2008). A New Projection-Type Split-Sample Score Tests in Linear Instrumental Variables Regression. University of Washington Working Paper.

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

Dufour, J. M. (1997). Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, 65: 1365–1388.

Dufour, J. M. and Jasiak, J. (2001). Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors. *International Economic Review*, 42: 815–843.

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

Dufour, J. M. and Taamouti, M. (2007). Further Results on Projection-Based Inference in IV Regressions with Weak, Collinear or Missing Instruments. *Journal of Econometrics*, 139: 133–153.

Guggenberger, P. and Smith, R. (2005). Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification. *Econometric Theory*, 21: 667–709.

Kleibergen, F. (2004). Testing Subsets of Parameters In The Instrumental Variables Regression Model. *The Review of Economics and Statistics*, 86: 418–423.

Kleibergen, F. (2005). Testing Parameters In GMM Without Assuming That They Are Identified. *Econometrica*, 73: 1103–1123.

Kleibergen, F. (2007). Subset statistic in the linear IV regression model. Technical report, Brown University.

Kleibergen, F. and Mavroeidis, S. (2008a). Inference on Subsets of Parameters in GMM without Assuming Identification. Technical report, Brown University Working Paper.

Kleibergen, F. and Mavroeidis, S. (2008b). Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve. Technical report, Brown University Working Paper.

Robins, J. M. (2004). Optimal Structural Nested Models for Optimal Sequential Decisions. In Lin, D. Y. and Heagerty, P., editors, *Proceedings of the Second Seattle Symposium on Biostatistics*. New York: Springer.

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

Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65: 557–586.

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

Stock, J. H. and Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In Andrews, D. W. K. and Stock, J. H., editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge University Press.

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

Zivot, E. and Chaudhuri, S. (2008). A Comment on Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve. University of Washingon Working Paper.

Zivot, E., Startz, R., and Nelson, C. (2006). Inference in Weakly Identified Instrumental Variables Regression. In Corbae, D., Durlauf, S. N., and Hansen, B. E., editors, *Frontiers in Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, pages 125–166. Cambridge University Press.

B Tables and Figures

n	k	Σ	WI-Case I		WI-Case II		WI-Case III		WI-Case IV	
			$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$	$\zeta = 1\%$	$\zeta = 5\%$
10^2	2	Σ_1	0	0.09	0.28	1.27	0.01	0.18	0.23	1.28
10^2	2	Σ_2	0.27	1.43	0.35	1.48	0.27	1.51	0.30	1.58
10^2	2	Σ_3	0	0.12	0.27	1.22	0.01	0.11	0.21	1.11
10^2	4	Σ_1	0.01	0.46	0.47	2.40	0.03	0.45	0.56	2.70
10^2	4	Σ_2	0.56	2.75	0.61	2.72	0.62	2.91	0.66	3.04
10^2	4	Σ_3	0.01	0.27	0.44	2.26	0.03	0.39	0.53	2.58
10^2	20	Σ_1	0.78	3.44	1.58	5.91	0.87	3.76	1.88	6.05
10^2	20	Σ_2	1.84	6.35	1.63	6.05	1.98	6.22	1.92	6.30
10^2	20	Σ_3	0.42	2.80	1.55	5.80	0.57	3.03	1.86	5.89
10^3	2	Σ_1	0.03	0.26	0.18	1.27	0.02	0.26	0.18	1.22
10^3	2	Σ_2	0.19	1.33	0.27	1.40	0.31	1.63	0.22	1.41
10^3	2	Σ_3	0.01	0.21	0.15	1.19	0.02	0.26	0.16	1.13
10^3	4	Σ_1	0.01	0.50	0.42	2.33	0.03	0.56	0.22	2.10
10^3	4	Σ_2	0.35	2.46	0.49	2.44	0.44	2.16	0.32	2.25
10^3	4	Σ_3	0.03	0.41	0.41	2.24	0.01	0.34	0.25	2.00
10^3	20	Σ_1	0.27	2.45	0.65	3.69	0.32	2.24	0.80	3.77
10^3	20	Σ_2	0.72	3.82	0.68	3.92	0.76	3.58	0.86	3.91
10^3	20	Σ_3	0.19	1.94	0.65	3.63	0.21	1.69	0.80	3.78
10^4	2	Σ_1	0	0.17	0.19	1.27	0.01	0.18	0.15	1.31
10^4	2	Σ_2	0.27	1.49	0.22	1.50	0.29	1.45	0.23	1.52
10^4	2	Σ_3	0	0.14	0.18	1.28	0	0.09	0.14	1.28
10^4	4	Σ_1	0.04	0.35	0.32	2.08	0.02	0.50	0.22	2.01
10^4	4	Σ_2	0.41	2.15	0.36	2.41	0.28	2.29	0.25	2.18
10^4	4	Σ_3	0.03	0.31	0.31	2.05	0.03	0.46	0.22	1.88
10^4	20	Σ_1	0.27	2.11	0.59	3.41	0.22	2.11	0.58	3.74
10^4	20	Σ_2	0.69	3.54	0.65	3.56	0.65	3.40	0.63	3.83
10^4	20	Σ_3	0.17	1.68	0.59	3.36	0.19	1.71	0.56	3.69

Table 3: % of times $\mathcal{C}_2(100 - \zeta, \theta_{01})$ is Empty!

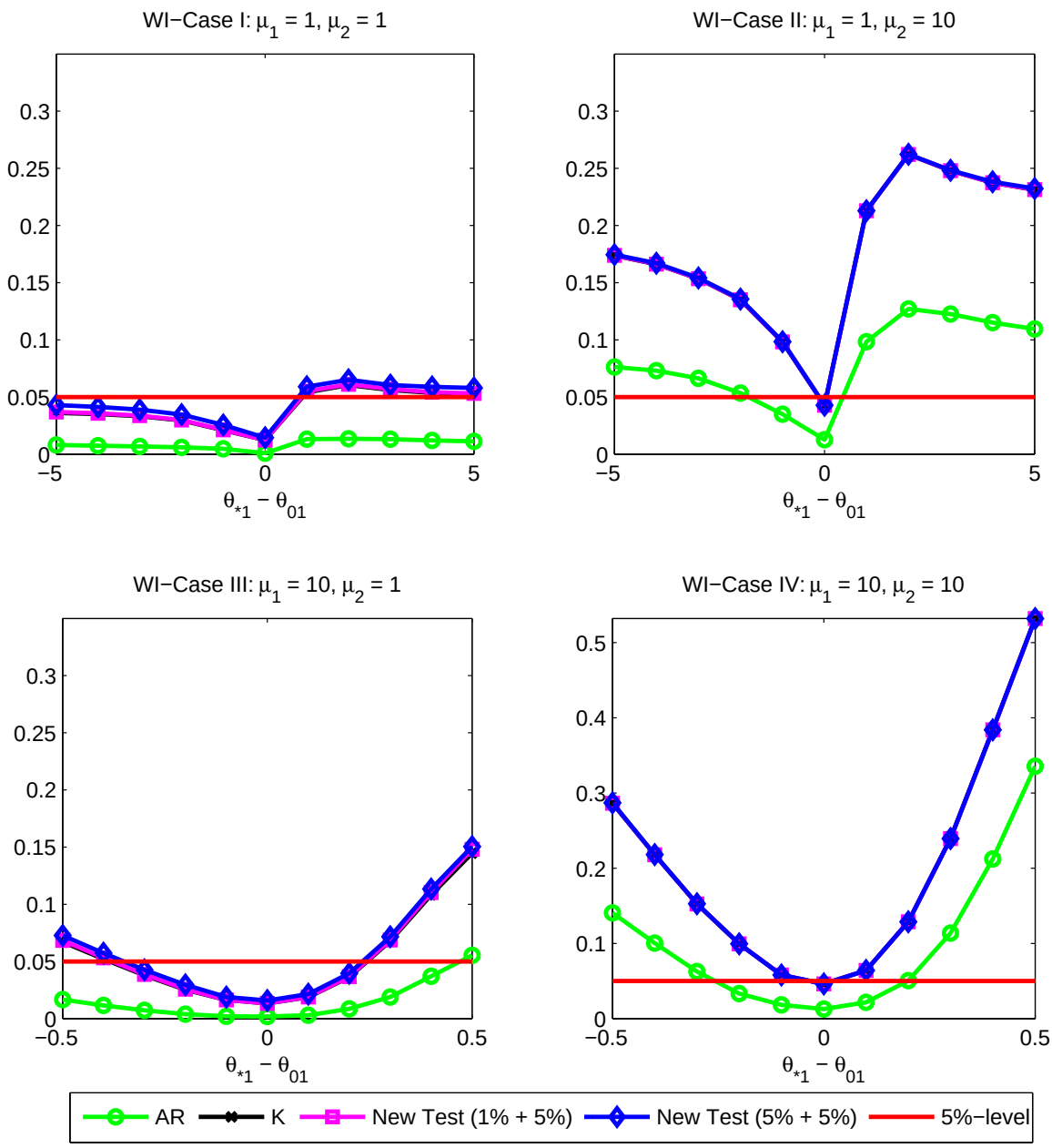

Figure 1: Sample Size = 100, Number of Instruments = 2, $\rho_{u1} = 0.5$, $\rho_{u2} = 0.5$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

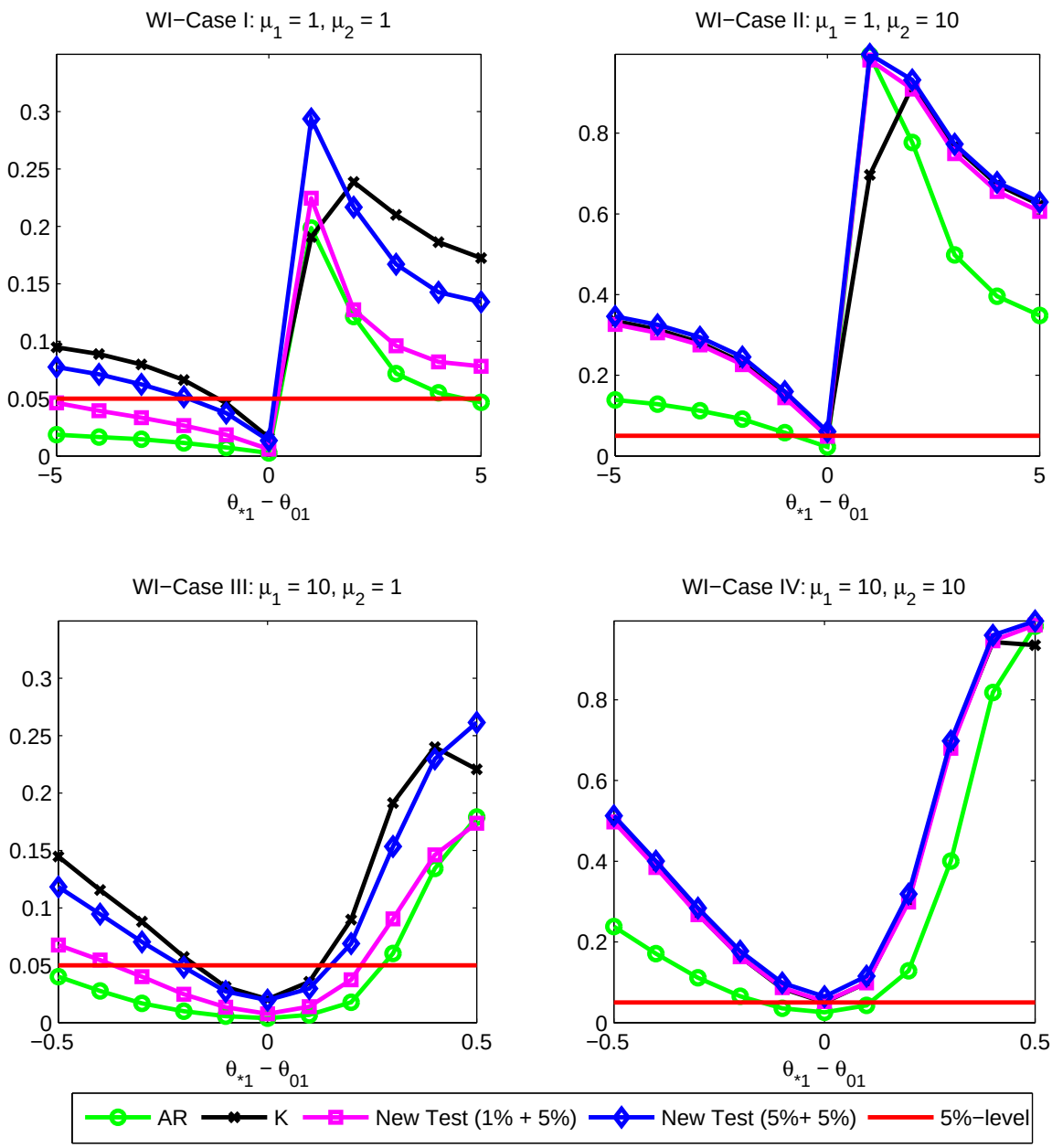

Figure 2: Sample Size = 100, Number of Instruments = 4, $\rho_{u1} = 0.99$, $\rho_{u2} = 0.1$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

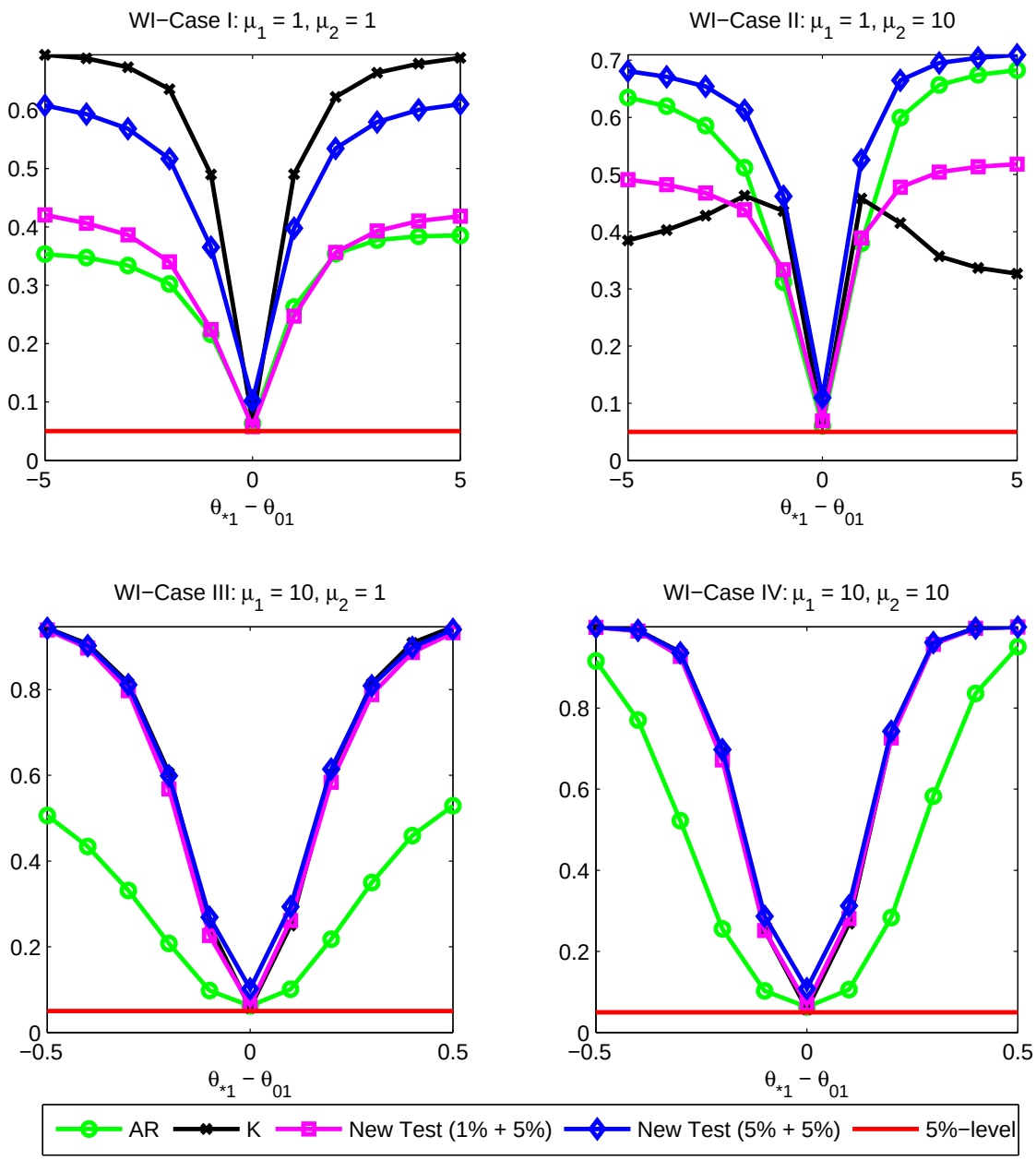

Figure 3: Sample Size = 100, Number of Instruments = 20, $\rho_{u1} = 0.1$, $\rho_{u2} = 0.99$ and $\rho_{12} = 0$. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.