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

Saraswata Chaudhuri* and Eric Zivot†

August 5, 2009

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 overly conservative. We show that under the standard regularity conditions the usual method of projection can be modified to obtain tests that are as powerful as the conventional plug-in-based tests for subsets of parameters. 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

We are concerned with the problem of inference on subsets of parameters using the generalized method of moments (GMM) when the identifiability of some or all of the parameters is in question. We focus on parameters θ, whose unknown "true value" θ_0 satisfies moment restrictions of the form

$$\begin{aligned} E[g(w_t,\theta)] &= 0 \text{ if } \theta = \theta_0 \\ &\neq 0 \text{ if } \theta \neq \theta_0 \end{aligned} \tag{1.1}$$

where $g : \mathcal{S} \times \Theta \mapsto \mathbb{R}^k$ is a measurable function, Θ is the parameter space, $\{w_t \in \mathcal{S} : t = 1, \ldots, n\}$ is the sample of observations from the sample space $\mathcal{S}$, and E is the expectation with respect to a probability measure $\mathcal{P}$ that considers θ_0 as the true value of θ. We model the identification failure of the elements of θ using the weak identification framework of Stock and Wright (2000).

GMM estimation is inconsistent if the elements of θ are weakly identified. Subsequently, the usual Wald, likelihood ratio and score statistics are not asymptotically pivotal and cannot be used for inference on the unknown true value θ_0. As a motivating example, consider an over-identified linear instrumental variables (IV) regression with moment restrictions given by (1.1) where $g(w_t,\theta) = Z_t(y_t - X_t'\theta)$, $E[X_t(y_t - X_t'\theta_0)] \neq 0$ and $w_t = (y_t, X_t', Z_t')'$ for $t = 1, \ldots, n$. The key nuisance parameter here is the one that determines the column-rank failure of $E[Z_t X_t']$, which, in turn, determines the identifiability of θ. Under the weak instrument(identification) framework that models this rank failure by specifying $E[Z_t X_t'] = O(1/\sqrt{n})$ (and $E[Z_t Z_t'] = O(1)$), the asymptotic distributions of the GMM estimator and the corresponding Wald, likelihood ratio and score statistics depend on the other unknown nuisance parameters $E[X_t(y_t - X_t'\theta_0)]$ and hence cannot be used for inference on θ_0 [see Phillips (1989), Staiger and Stock (1997), Wang and Zivot (1998)].

Fortunately there have been some advances recently in developing test statistics whose asymptotic distribution is robust to weak identification. The S statistic based on the efficient continuous updating GMM (CU-GMM) objective function and the K statistic based on a quadratic form of the gradient of the same objective function are asymptotically pivotal at $\theta = \theta_0$ [see Stock and Wright (2000) and Kleibergen (2005)]. Hence the corresponding tests, the S and the K tests, can be used to correctly test for hypotheses of the form $\theta = \theta_0$ irrespective of weak identification.

The focus of this paper is inference on subsets of parameters. To fix ideas, let $\theta = (\theta_1', \theta_2')'$ and let θ_1 be the subset of interest. We treat θ_2 as nuisance parameters. 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).

Regardless of the identifiability of θ_1 and θ_2, i.e. under WI-Cases I-IV, one can always use the usual projection technique based on the K and the S statistics to test for the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ and invert these tests to obtain confidence regions with (at least) correct asymptotic coverage probability. However, such inferences can be quite conservative.

The purpose of our paper is to show how the usual projection-based methods in the GMM framework can be modified to reduce the conservativeness generally associated with them. In particular, we propose a modification of the projection-based K test that is generally more powerful than the usual projection-based K test and at the same time does not result in an uncontrolled over-rejection of the true value of θ_1 under WI-Cases I-IV. The new test, which we call the *efficient projection-based K test*, is motivated from Robins (2004).

The standard way of avoiding the conservativeness of the usual projection-based methods is to use the plug-in principle. Two such methods are the subset-S and the subset-K tests that plug-in the CU-GMM estimator of θ_2 (restricted by the null hypothesis) to the S and the K statistics, and adjust the critical values of the tests suitably [see Kleibergen (2005)]. Recently Kleibergen and Mavroeidis (2009) showed that these tests never over-reject the true value of θ_1 irrespective of the identifiability of θ_1 and θ_2 (i.e. under WI-Cases I-IV). This is a major development in the weak identification literature. Simulations show that the subset-K test is generally more powerful than the subset-S test. Hence in models with possibly weakly identified parameters, the subset-K test should provide a good benchmark to gauge the performance of the efficient projection-based K test (the new test) in the GMM framework. We show that the new test can be made asymptotically (locally) equivalent to the subset-K test whenever θ_2 is identified, i.e. in WI-Cases II and IV.

Our results have a general implication beyond the paradigm of weak identification. 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-based inference presented in this paper helps to reduce this problem of conservativeness asymptotically. Moreover, the results show that the projection principle, if applied in the way proposed in this paper, can also lead to methods of inference that can, under standard circumstances, be made asymptotically equivalent to the (generally more powerful) methods based on the plug-in principle.

The rest of the paper is organized as follows. Section 2 describes the efficient projection-based K test, Section 3 is a Monte Carlo study in a linear instrumental variables regression 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 relegated to the Appendix. This paper draws heavily from Chaudhuri (2008) to which we frequently refer the readers for additional results and discussions. (web link – http://www.unc.edu/~saraswat/saraswata_thesis.pdf)

Notations:

Definitions of all the notations used are collected here. $O_p(1)$, $o_p(1)$, $\xrightarrow{P}$, $\xrightarrow{d}$, $\overset{A}{\sim}$ and $:=$ respectively denote "bounded in probability", "converges in probability to zero", "converges in probability", "converges in distribution", "asymptotically follows"and "defined as". If $A = [A_1, \ldots, A_{bc}]$ is an $a \times bc$ matrix then $vec\ A := [A_1', \ldots, A_{bc}']'$, $devec_c A' := \left[(A_1, \ldots, A_c)', \ldots, (A_{(b-1)c+1}, \ldots, A_{bc})'\right]$ 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. If A is symmetric and positive semi-definite then $A^{\frac{1}{2}}$ is the lower-triangular Cholesky factor of A such that $A = A^{\frac{1}{2}}A^{\frac{1}{2}'}$. If $A = ((A_{ij}))_{i,j=1,2}$ is such that the diagonal blocks A_{11} and A_{22} are non-singular then $A_{ii.j} := A_{ii} - A_{ij}A_{jj}^{-1}A_{ji}$ is the Schur complement of A_{jj} for $i \neq j = 1, 2$. If $\mathcal{W}$ is a closed set then $\mathcal{W}^{\text{int}} := \text{interior}(\mathcal{W})$. Lastly, let $\overline{X}_t := X_t - E[X_t]$ for any random variable X_t, and denote the derivatives with respect to θ as $\triangledown_\theta g(w_t, \theta) := \partial/\partial\theta'(g(w_t, \theta))$ and $\triangledown_{\theta\theta} g(w_t, \theta) := \partial/\partial\theta'(vec \triangledown_\theta g(w_t, \theta))$ if they exist.

2 The efficient projection-based K test in GMM

We make a set of high level assumptions following Stock and Wright (2000), Guggenberger and Smith (2005) and Kleibergen (2005) to describe the GMM framework based on the moment restrictions in (1.1). Chaudhuri (2008) contains more discussion on these assumptions.

Assumption Θ: [*assumptions on the parameter space*]
Equation (1.1) gives $k \geq \nu$ moment restrictions on the $\nu \times 1$ parameter vector θ. Let $\theta_0 = (\theta_{01}', \theta_{02}')'$ be such that $\theta_{0i} \in \Theta_i^{\text{int}}$ where Θ_i is a ν_i-dimensional compact 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 $\nu = \nu_1 + \nu_2$.

Hereafter we suppress the explicit dependence of the functionals on the observations for notational convenience; for example, $g_t(\theta)$ should be read as $g(w_t, \theta)$, $\nabla_\theta g_t(\theta)$ as $\nabla_\theta g(w_t, \theta)$ and so on.

Assumption D: [*assumptions on the moment vector and its derivatives*]

D1. (i) $g_t(\theta)$ is twice continuously differentiable in $\theta \in \Theta^{\text{int}}$.
(ii) $\partial/\partial\theta'(E[g_t(\theta)]) = E[\nabla_\theta g_t(\theta)]$ for $\theta \in \mathcal{T} \subset \Theta$, where $\mathcal{T}$ is a non-shrinking open neighborhood of θ_0. [1]

D2. (i) $\sup_{\theta \in \Theta} n^{-1} \sum_{t=1}^n \overline{g_t(\theta)} = o_p(1)$.
(ii) $n^{-1} \sum_{t=1}^n \overline{\nabla_\theta g_t(\theta)} = o_p(1)$ for $\theta \in \Theta^{\text{int}}$.
(iii) $n^{-1} \sum_{t=1}^n \overline{\nabla_{\theta\theta} g_t(\theta)} = o_p(1)$ for $\theta \in \Theta^{\text{int}}$ and $E[n^{-1} \sum_{t=1}^n \nabla_{\theta\theta} g_t(\theta)]$ converges to a continuous and bounded function $L(\theta)$.

D3. The joint asymptotic distribution of the moment vector and its first derivative is given by

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{bmatrix} \overline{g_t(\theta_0)} \\ vec\ \overline{\nabla_\theta g_t(\theta_0)} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Psi_g \\ \Psi_\nabla \end{bmatrix} \sim \mathcal{N}\left(0, V(\theta_0) = \begin{bmatrix} V_{gg}(\theta_0) & V_{g\nabla}(\theta_0) \\ V_{\nabla g}(\theta_0) & V_{\nabla\nabla}(\theta_0) \end{bmatrix}\right).$$

$V(\theta)$ is a bounded, continuous, symmetric positive semi-definite matrix. $V_{gg}(\theta)$ is positive definite and differentiable with respect to $\theta \in \Theta^{\text{int}}$.

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

Assumption W: [*assumptions on the weak identification*]
$E\ n^{-1} \sum_{t=1}^n g_t(\theta_1, \theta_2) = \sum_{i=1}^2 \left[1_{[\delta_i = 1]} m_i(\theta_i) + 1_{[\delta_i = \frac{1}{2}]} n^{-1/2} \tilde{m}_{ni}(\theta_1, \theta_2)\right]$ where for $i = 1, 2$,

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

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

[1]See Lemma 3.6 in Newey and McFadden (1994) for sufficient conditions. This is trivially satisfied in linear IV regressions. Neighborhoods such as $\mathcal{T}$ could be allowed to shrink to θ_0 at a rate slower than $\sqrt{n}$, but is not done here for simplicity [also see Assumption O].

The δ's are the key nuisance parameters in this model and the non-random indicator functions involving them in Assumption W are used to distinguish between the four cases of weak (partial) identification mentioned in the introduction and summarized in Table 1. (δ_i's are assigned the values 1/2 and 1 because n^{δ_i} will often be used as a suitable scaling factor, for example in (2.1).)

	$\delta_2 = \frac{1}{2}$	$\delta_2 = 1$
$\delta_1 = \frac{1}{2}$	**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

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

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

$$G_{ni}(\theta) := E\frac{1}{n^{\delta_i}}\nabla_i \sum_{t=1}^{n} g_t(\theta) = 1_{[\delta_i=1]}M_i(\theta_i) + \sum_{j=1}^{2} 1_{[\delta_j=\frac{1}{2}]}\frac{n^{\delta_j}}{n^{\delta_i}}\tilde{M}_{nj}^{(i)}(\theta) \tag{2.1}$$

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 in θ and

$$G_{ni}(\theta) \xrightarrow{P} G_i(\theta) := 1_{[\delta_i=1]}M_i(\theta_i) + 1_{[\delta_i=\frac{1}{2}]}\left[\tilde{M}_i^{(i)}(\theta) + 1_{[\delta_j=\frac{1}{2}]}\tilde{M}_j^{(i)}(\theta)\right], \tag{2.2}$$

which has full column-rank for $\theta \in \theta_{0i} \times \Theta_j$ when θ_i is identified (local identification condition).

2.1 The subset-K test and the efficient projection-based K test for $H_1 : \theta_1 = \theta_{*1}$

Weak identification robust tests for the entire parameter θ are those whose nominal size equals the asymptotic size irrespective of the δ's, i.e. under WI-Cases I-IV. We relax this criterion a little when testing the hypothesis $H_1 : \theta_1 = \theta_{*1}$ on subsets of parameters and allow the nominal size to be less than the asymptotic size while prefering them to be equal whenever possible because that would typically lead to gains in power. The tests considered here are based on the CU-GMM. Using Assumptions D3 and D4, we 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 K statistic is based on the gradient of the CU-GMM objective function with respect to θ which 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 $\widehat{D}_t(\theta) := devec_k[vec\, \nabla_\theta g_t(\theta) - \widehat{V}_{\nabla g}(\theta)\widehat{V}_{gg}^{-1}(\theta)g_t(\theta)]'$.

The subset-K test

Kleibergen's subset-K test rejects the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ at level ϵ if $K_n(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1})) > \chi^2_{\nu_1}(1-\epsilon)$ where $\tilde{\theta}_{n2}(\theta_{*1}) := \arg\min_{\theta_2 \in \Theta_2} Q_n(\theta_{*1}, \theta_2)$ is the continuous updating GMM estimator (CUE) of θ_2 under the restriction $\theta_1 = \theta_{*1}$, and the K statistic is defined as

$$\begin{aligned} K_n(\theta) &:= n\left(\nabla_\theta Q_n(\theta)\right)\left[\widehat{D}_T'(\theta)\widehat{V}_{gg}^{-1}(\theta)\widehat{D}_T(\theta)\right]^{-1}\left(\nabla_\theta Q_n(\theta)\right)' \\ &= \frac{1}{n} g_T'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta) P\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_T(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta) g_T(\theta). \end{aligned} \tag{2.4}$$

Kleibergen (2005) showed that $K_n(\theta_{01}, \tilde{\theta}_{n2}(\theta_{01})) \overset{A}{\sim} \chi^2_{\nu_1}$ in WI-Cases II and IV and under Assumptions Θ, D, W and $\partial/\partial\theta_2'(Q_n(\theta_1, \tilde{\theta}_{n2}(\theta_1))) = 0$ (see O1. below). The limiting $\chi^2_{\nu_1}$ distribution of $K_n(\theta_{01}, \tilde{\theta}_{n2}(\theta_{01}))$ in Kleibergen's proof and its behavior under local alternatives crucially depend on the $\sqrt{n}$-consistency of the restricted CUE of θ_2, i.e. on the result of Lemma 2.1.

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

In WI-Cases I and III, when $\tilde{\theta}_{n2}(\theta_{01})$ is inconsistent, and the limiting distribution of $K_n(\theta_{01}, \tilde{\theta}_{n2}(\theta_{01}))$ is not $\chi^2_{\nu_1}$. It is in these two cases where the literature typically recommends the use of projection.

However, as discussed in the introduction, Kleibergen and Mavroeidis (2009) recently showed that even in WI-Cases I and III, i.e. even when θ_2 is weakly identified, the subset-K test does not over-reject the true value of θ_1. This is important because not only, by virtue of the score principle, the subset-K test is likely to enjoy some local optimality properties under standard regularity conditions, but also it is guaranteed to never over-reject the true value of the parameters of interest. Hence the subset-K test provides a benchmark against which we can compare the performance of the efficient projection-based K test to establish the latter's usefulness.

Other assumptions are also required to show that the efficient projection-based K test and the subset-K test are asymptotically (locally) equivalent under WI-Cases II and IV.

Assumption O:

O1. $\tilde{\theta}_{n2}(\theta_1) := \arg\min_{\theta_2 \in \Theta_2} Q_n(\theta_1, \theta_2)$ satisfies the first order condition $\partial/\partial\theta_2'(Q_n(\theta_1, \tilde{\theta}_{n2}(\theta_1))) = 0$ for any $\theta_1 \in \mathcal{T}_1 \subset \Theta_1$ where $\mathcal{T}_1$ is a non-shrinking open neighborhood of θ_{01}.

O2. (i) $\partial/\partial\theta'(vec\ \nabla_\theta E[g_t(\theta)]) = E[\nabla_{\theta\theta} g_t(\theta)]$ for $\theta \in \mathcal{T} \subset \Theta$, where $\mathcal{T}$ is a non-shrinking open neighborhood of θ_0.
(ii) $\tilde{M}_{n1}^{(1,2)}(\theta) := \partial/\partial\theta_2'(vec\ \tilde{M}_{n1}^{(1)}(\theta))$ is $o(\sqrt{n})$ for $\theta \in \mathcal{T}$.

Assumption O1 is similar to (part of) Assumption 4 in Kleibergen and Mavroeidis (2009) and is required to show that for large sample sizes, the subset-K statistic is the same as the efficient K statistic (see the discussion following (2.5)). Assumption O2 summarizes a set of sufficient conditions required for the top right $\nu_1 k \times \nu_2$ block of $L(\theta_0)$ to be equal to zero. While this is satisfied trivially in linear IV regressions, in more general models this condition is required for further local asymptotic equivalence under WI-Case II (i.e. when the subset of interest θ_1 is weakly identified). These assumptions were not made explicitly in Chaudhuri (2008).

The efficient K statistic

The use of the efficient K statistic is crucial to establish the local asymptotic equivalence between the subset-K test and the efficient projection-based K test when θ_2 is identified. The efficient K statistic is the CU-GMM counterpart of the efficient score statistic from the likelihood based inference, and its construction is based on the $C(\alpha)$ principle of Neyman (1959). More intuition and detailed discussion, that are omitted here for brevity, are given in Chaudhuri (2008).

In the context of CU-GMM, we use Assumption D3 and define the estimated efficient score (gradient) for θ_1 such that, at $\theta = \theta_0$, it is asymptotically independent of the score (gradient) for θ_2 (both under suitable scaling). Accordingly, the estimated efficient score for θ_1 is

$$\begin{aligned}\nabla_{1.2}Q_n(\theta) &:= \nabla_1 Q_n(\theta) - \nabla_2 Q_n(\theta)\left[\widehat{D}_{T2}'(\theta)\widehat{V}_{gg}^{-1}(\theta)\widehat{D}_{T2}(\theta)\right]^{-1}\widehat{D}_{T2}'(\theta)\widehat{V}_{gg}^{-1}(\theta)\widehat{D}_{T1}(\theta)\\ &= \frac{1}{n}g_T'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta)N\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T2}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T1}(\theta),\end{aligned}$$

where $\nabla_i Q_n(\theta) := \partial/\partial\theta_i'(Q_n(\theta))$ for $i = 1, 2$. This uses the partition $\widehat{D}_T(\theta) = [\widehat{D}_{T1}(\theta), \widehat{D}_{T2}(\theta)]$ and $\widehat{V}_{\nabla g}(\theta) = [\widehat{V}_{1g}'(\theta), \widehat{V}_{2g}'(\theta)]'$ with respect to θ_1 and θ_2. Now, using Neyman's $C(\alpha)$ principle we define the efficient score version of the K statistic, i.e. the efficient K statistic, as

$$\begin{aligned}K_{n1}(\theta) &:= n\left(\nabla_{1.2}Q_n(\theta)\right)\underbrace{\left(\widehat{D}_{T1}'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta)N\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T2}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T1}(\theta)\right)^{-1}}_{=\Upsilon\text{ (say)}}\left(\nabla_{1.2}Q_n(\theta)\right)'\\ &= \frac{1}{n}g_T'(\theta)\widehat{V}_{gg}^{-\frac{1}{2}}(\theta)P\left(N\left(\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T2}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)\widehat{D}_{T1}(\theta)\right)\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta)g_T(\theta). \qquad (2.5)\end{aligned}$$

Assumption O1 implies that $\tilde{\theta}_{n2}(\theta_1) \in \Theta_2^{\text{int}}$ for $\theta_1 \in \mathcal{T}_1$. Hence, noting that Υ is the top left $\nu_1 \times \nu_1$ block of $[\widehat{D}_T'(\theta)\widehat{V}_{gg}^{-1}(\theta)\widehat{D}_T(\theta)]^{-1}$, it follows that $K_n(\theta_1, \tilde{\theta}_{n2}(\theta_1)) = K_{n1}(\theta_1, \tilde{\theta}_{n2}(\theta_1))$. Therefore, Kleibergen's (subset-)K statistic can also be interpreted as a (normalized) quadratic form of the estimated efficient score for θ_1 in which the restricted CUE $\tilde{\theta}_{n2}(\theta_1)$ is plugged in for θ_2.

Lemma 2.2 *Let $\theta_{n1} = \theta_{01} + d_1/\sqrt{n} \in \Theta_1^{int}$ where $\|d_1\| = O(1)$ and let $\theta_{n2} = \theta_{02} + d_2/\sqrt{n} \in \Theta_2^{int}$ where $\|d_2\| = O_p(1)$. Denote $\theta_n = (\theta_{n1}', \theta_{n2}')'$ and $d_\theta = (d_1', d_2')'$. Then the following results hold under Assumptions Θ, D and W:*

(i) in WI-Cases I-IV, $K_{n1}(\theta_{01}, \theta_{02}) \overset{A}{\sim} \chi^2_{\nu_1}$.

(ii) in WI-Case IV, $K_{n1}(\theta_{n1}, \theta_{02}) = K_{n1}(\theta_{n1}, \theta_{n2}) + o_p(1) \overset{A}{\sim} \chi^2_{\nu_1}(ncp = \eta(d_1)'\eta(d_1))$ where $\eta(d_1)$ can be expressed as $P(N(V_{gg}^{-\frac{1}{2}'}(\theta_0)M_2(\theta_{02}))V_{gg}^{-\frac{1}{2}'}(\theta_0)M_1(\theta_{01}))V_{gg}^{-\frac{1}{2}'}(\theta_0)M_1(\theta_{01})d_1$. [2]

(iii) in WI-Case IV, $K_{n1}(\theta_{n1}, \theta_{n2}) = K_n(\theta_{n1}, \tilde{\theta}_{n2}(\theta_{n1})) + o_p(1)$, if additionally Assumption O1 is satisfied.

(iv) in WI-Case II, $K_{n1}(\theta_{n1}, \theta_{02}) = K_{n1}(\theta_{n1}, \theta_{n2}) + o_p(1) = K_n(\theta_{n1}, \tilde{\theta}_{n2}(\theta_{n1})) + o_p(1)$, if additionally Assumptions O1 and O2 are satisfied.

[2] Alternative representation of $\eta(d_1)$ uses the fact that $P(A)P(A) = P(A) = [(A'A)^{-1/2'}A']'[(A'A)^{-1/2'}A']$.

Remarks: The results in Lemma 2.2 are used to discuss of the properties of the efficient projection-based K test in Theorems 2.3 and 2.4. For now, we proceed after noting the following:

(i) The importance of using the efficient K statistic is most prominently displayed by the asymptotic equivalence in (ii) and (iv). These results imply that as long as the unknown nuisance parameter θ_2 is replaced by any $\sqrt{n}$-consistent estimator, the efficient K (or score) statistic is asymptotically equivalent to its "infeasible form" that replaces θ_2 by its unknown true value θ_{02}. In other words, small deviations ($O(1/\sqrt{n})$) from the unknown true value of the nuisance parameters θ_2 do not affect the asymptotic behavior of the efficient K statistic for θ_1. [3,4] Other statistics, including other forms of the K or score statistic, generally do not have this nice property [see Chaudhuri (2008)].

(ii) The asymptotic distribution of the statistics in (iv) is not (non-central) χ^2 [also see Lemma A.3(i.a) in the Appendix].

The efficient projection-based K test

The usual method of projection, based on the efficient K statistic would reject the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ if $\inf_{\theta_{*2}\in\Theta_2} K_{n1}(\theta_{*1}, \theta_{*2}) > \chi^2_{\nu_1}(1-\epsilon)$ and, by virtue of the result in Lemma 2.2(i), the asymptotic size of this test can never exceed ϵ. However, this test does not utilize the other important properties described in Lemma 2.2 (ii) -(iv), which could help to reduce its conservativeness. The efficient projection-based K test, to be described now, precisely serves this purpose.

We describe two different versions of the efficient projection-based K tests in Theorems 2.3 and 2.4 respectively. The purpose of the first version is to show that the efficient projection-based K test can be designed such that it is never over-sized and, at the same time, in WI-Cases II and IV it is asymptotically equivalent to the subset-K test and the infeasible efficient K test. The second version compromises a little bit (but not beyond control) in terms of the asymptotic size but it has other important properties which, in our opinion, make it useful for practical purposes.

Theorem 2.3 *Let* $\theta_{*1} = \theta_{01} + d_1/\sqrt{n} \in \Theta_1^{int}$ *where* $\|d_1\| = O(1)$. *Let there exist a sequence of regions* $\{\mathcal{C}_{2n}(\theta_{*1})\}_n \subset \Theta_2$, *possibly dependent on* θ_{*1} *and* n, *such that it contains* $\tilde{\theta}_{n2}(\theta_{*1}) := \arg\min_{\theta_2\in\Theta_2} Q_n(\theta_{*1}, \theta_2)$, *and in WI-Cases II and IV,* $\sup_{\theta_{*2}\in\mathcal{C}_{2n}(\theta_{*1})} \sqrt{n}\|\theta_{*2} - \theta_{02}\| = O_p(1)$. *Define the rejection rule of the efficient projection-based K test for the null hypothesis* $H_1 : \theta_1 = \theta_{*1}$ *by the random variable* $\phi_n(\theta_{*1}) \equiv \phi_n(\theta_{*1}; w_1, \ldots, w_n)$ *such that*

$$\phi_n(\theta_{*1}) := \begin{cases} 1 & \textit{if} \ \ \inf_{\theta_{*2}\in\mathcal{C}_{2n}^{int}(\theta_{*1})} K_{n1}(\theta_{*1}, \theta_{*2}) > \chi^2_{\nu_1}(1-\epsilon) \\ 0 & \textit{otherwise.} \end{cases}$$

Then the following results hold under Assumptions Θ, D, W and O1:

(i) the asymptotic size of the efficient projection-based K test cannot ever exceed ϵ

(ii) if additionally Assumption O2 is satisfied, then under WI-Cases II and IV the efficient projection-based K test is (locally) asymptotically equivalent to the infeasible efficient K test

[3] Use of the K statistic is important because unlike the usual score statistic in the GMM context, the K statistic happens to be asymptotically pivotal under WI-Cases I-IV if θ_1 and θ_2 are replaced by their unknown true values.

[4] The asymptotic equivalence with the subset-K statistic in WI-Cases II and IV is simply an artifact of this important result because in those two cases the CUE of θ_2 (both unrestricted and restricted) happens to be a $\sqrt{n}$-consistent estimator [see Lemma A1 in Stock and Wright (2000) and Lemma 2.1].

that rejects $H_1 : \theta_1 = \theta_{*1}$ if $K_{n1}(\theta_{*1}, \theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)$ and the subset-K test that rejects $H_1 : \theta_1 = \theta_{*1}$ if $K_{n1}(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1})) > \chi^2_{\nu_1}(1-\epsilon)$.

Remarks:

(i) While for all sample size n, $K_n(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1}) \geq \inf_{\theta_{*2}\in\mathcal{C}_{2n}(\theta_{*1})} K_{n1}(\theta_{*1}, \theta_{*2})$ implying that the subset-K test rejects the null hypothesis at least as often as the efficient projection-based K test, this difference vanishes asymptotically in WI-Cases II and IV.

(ii) The same phenomenon is true while comparing the infeasible efficient K test with the efficient projection-based K test if we replace the condition that $\tilde{\theta}_{n2}(\theta_{*1}) \in \mathcal{C}_{2n}(\theta_{*1})$ by $\theta_{02} \in \mathcal{C}_{2n}(\theta_{*1})$ [see Lemma 2.2]. However, since θ_{02} is unknown, it is typically easier to construct a region with the former restriction rather than the latter. For example, *even* a naive (Cartesian product form) Wald confidence region (based on the restriction $H_1 : \theta_1 = \theta_{*1}$) given by

$$\mathcal{C}^W_{2n}(1-\zeta, \theta_{*1}) := \left\{\theta_2 \in \Theta_2 : \theta_2 \in \times_{i=1}^{\nu_2} \left[\tilde{\theta}_{n2,i}(\theta_{*1}) \pm n^{-1/2} z_{1-\zeta/2}\mathrm{se}_i(\theta_{*1})\right]\right\} \tag{2.6}$$

satisfies the restrictions imposed on $\mathcal{C}_{2n}(\theta_{*1})$. In the definition of $\mathcal{C}^W_{2n}(1-\zeta, \theta_{*1})$, $z_{1-\zeta/2}$ is the $(1-\zeta/2)$-th quantile of $N(0,1)$, $\tilde{\theta}_{n2,i}(\theta_{*1})$ is the i-th element of $\tilde{\theta}_{n2}(\theta_{*2})$ and $\mathrm{se}_i(\theta_{*1})$ is the positive square root of the i-th diagonal element of $[n^{-2}\nabla_2 g'_T(\theta)\widehat{V}^{-1}_{gg}(\theta)\nabla_2 g_T(\theta)]^{-1}$ (evaluated at $\theta = (\theta'_{*1}, \tilde{\theta}'_{n2}(\theta_{*1})')$. [5] The choice of ζ in $\mathcal{C}^W_{2n}(1-\zeta, \theta_{*1})$ does not matter asymptotically as long as ζ is bounded away from 0 (at $\zeta = 1$ and 0, the new test is respectively identical to the subset-K test and the usual projection test based on the efficient K statistic).

(iii) One could possibly use confidence regions based on the weak identification robust tests like the K test and the GMM-MLR test, both of which always contain $\tilde{\theta}_{n2}(\theta_{*1})$ [see Kleibergen (2005)]. However, we do not pursue it further because these regions make the computation extremely difficult and, at the same time, do not offer any theoretical advantage over the region given in (2.6).

Theorem 2.4 *Let $\theta_{*1} = \theta_{01} + d_1/\sqrt{n} \in \Theta^{int}_1$ where $\|d_1\| = O(1)$. For any $\theta_1 \in \Theta_1$, define a confidence region for θ_2 as $\mathcal{C}^S_{2n}(1-\zeta, \theta_1) := \{\theta_2 \in \Theta_2 : S_n(\theta_1, \theta_2) \leq \chi^2_k(1-\zeta)\}$ where $S_n(\theta_1, \theta_2) := 2Q_n(\theta_1, \theta_2)$. Define the rejection rule of the efficient projection-based K test for the null hypothesis $H_1 : \theta_1 = \theta_{*1}$ by the random variable $\phi_n(\theta_{*1}) \equiv \phi_n(\theta_{*1}; w_1, \ldots, w_n)$ such that*

$$\phi_n(\theta_{*1}) := \begin{cases} 1 & \text{if } \mathcal{C}^S_{2n}(1-\zeta, \theta_{*1}) = \varnothing \text{ or if } \inf_{\theta_{*2}\in\mathcal{C}^{S,int}_{2n}(1-\zeta,\theta_{*1})} K_{n1}(\theta_{*1}, \theta_{*2}) > \chi^2_{\nu_1}(1-\epsilon) \\ 0 & \text{otherwise.} \end{cases}$$

Then the following results hold under Assumptions Θ, D and W:

(i) the asymptotic size of the efficient projection-based K test is bounded from above by $\zeta + \epsilon$ under WI-Cases I-IV

(ii) if additionally Assumption O is satisfied and $\mathcal{C}^S_{2n}(1-\zeta, \theta_{*1}) \neq \varnothing$, then under WI-Cases II and IV the efficient projection-based K test is (locally) asymptotically equivalent to the infeasible efficient K test that rejects $H_1 : \theta_1 = \theta_{*1}$ if $K_{n1}(\theta_{*1}, \theta_{02}) > \chi^2_{\nu_1}(1-\epsilon)$ and the subset-K test that rejects $H_1 : \theta_1 = \theta_{*1}$ if $K_{n1}(\theta_{*1}, \tilde{\theta}_{n2}(\theta_{*1})) > \chi^2_{\nu_1}(1-\epsilon)$.

[5] In WI-Cases II and IV, it does not matter asymptotically if se_i is computed from $[n^{-2}\nabla_2 g'_T(\theta)\widehat{V}^{-1}_{gg}(\theta)\nabla_2 g_T(\theta)]^{-1}$ or $[n^{-2}\widehat{D}'_{T2}(\theta)\widehat{V}^{-1}_{gg}(\theta)\widehat{D}_{T2}(\theta)]^{-1}$, the former is of course easier. The usual Wald confidence region (restricted by $H_1 : \theta_1 = \theta_{*2}$) also satisfies the requirements of the theorem [also see Conniffe (2001)].

Remarks:

(i) By construction, the efficient K statistic is identically equal to zero at the local extrema and the saddle points of the objective function $Q_n(\theta_1, \theta_2)$. This causes a spurious decline in power (at these points) of the efficient projection-based K test (this is also true for the other forms of the K test). While conditions such as Assumption 4 in Kleibergen and Mavroeidis (2009) can rule out the local extrema in terms of θ_2, it will probably be too restrictive to impose such conditions for both θ_1 and θ_2. The use of $\mathcal{C}^S_{2n}(1-\zeta, \theta_{*1})$ has a major advantage in this regard. Since the S test (based on which $\mathcal{C}^S_{2n}(1-\zeta, \theta_{*1})$ is obtained) concurrently tests for the moment restrictions given in (1.1), such local extrema and saddle points (outside the $\sqrt{n}$-neighborhood of θ_0) are automatically ruled out if the moment restrictions are true. In that sense, this version of the new test has similar functionality as the subset-K-J test and hence important for practical purposes [see Kleibergen (2005)].

(ii) Unlike Theorem 2.3, here the asymptotic size of the efficient projection-based K test can exceed ϵ. This happens because the first step confidence region $\mathcal{C}^S_{2n}(1-\zeta, \theta_{01})$ can be empty. However, since this region is also guaranteed to contain θ_{02} with probability approaching $1-\zeta$, one can always choose ϵ and ζ so that the asymptotic size does not exceed its desired level [see Lemma A.3 in the Appendix]. On the other hand, the same phenomenon of empty $\mathcal{C}^S_{2n}(1-\zeta, \theta_{*1})$ also drives up the asymptotic power of the new test against the false values of θ_1. Simulation results in the next section suggest that the gain in power can be quite significant and this benefit probably outweighs the cost in terms of a minor increase in size.

(iii) The upper bound $\zeta + \epsilon$ on the asymptotic size is obtained using Bonferroni's inequality. However, under the conditions of the theorem and provided that $\mathcal{C}^S_{2n}(1-\zeta, \theta_{01}) \neq \varnothing$, the choice of $\zeta(\neq 0)$ does not matter in WI-Cases II and IV, and the asymptotic size of the test cannot exceed ϵ. This is an important difference from the usual Bonferroni-type tests and this is possible because of the properties of the efficient K statistic as described in Lemma 2.2 [see Moon and Schorfheide (2009) and the references therein].

Corollary 2.5 *When the asymptotic equivalence with the subset-K test holds, the efficient projection-based K test is asymptotically at least as powerful as the usual projection-based K test that rejects* $H_1 : \theta_1 = \theta_{*1}$ *if* $\inf_{\theta_2 \in \Theta_2} K_n(\theta_{*1}, \theta_2) > \chi^2_\nu(1-\epsilon)$.

Finally a word on implementation: If it is possible to obtain a $\sqrt{n}$-consistent point estimator for some elements of θ_2, 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 only to the confidence region for the remaining elements of θ_2.

3 Simulation study in a linear IV regression

We present a Monte Carlo study in a linear instrumental variables regression 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). See Chaudhuri (2008) for discussions on the Monte Carlo specification.

We draw $w_t = (y_t, X_{1t}, X_{2t}, Z_t')'$ for $t = 1, \dots, n = 100$ from the data generating process:

$$\left\{ \begin{array}{c} y_t = X_{1t}\theta_{01} + X_{2t}\theta_{02} + u_t, \\ X_{1t} = Z_t'\Pi_1 + U_{1t}, \\ X_{2t} = Z_t'\Pi_2 + U_{2t} \end{array} \right\} \text{ where } (u_t, U_{1t}, U_{2t}) \overset{\text{i.i.d}}{\sim} \mathcal{N}\left(0, \Sigma = \begin{bmatrix} 1 & \rho_{u1} & \rho_{u2} \\ \rho_{u1} & 1 & 0 \\ \rho_{u2} & 0 & 1 \end{bmatrix}\right)$$

is independent of $Z_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, I_k)$. The true values of the structural coefficients are $\theta_{01} = 1$ and $\theta_{02} = 10$. We make three different choices for the pair (ρ_{u1}, ρ_{u2}): $(0.5, 0.5), (0.1, 0.99), (0.99, 0.1)$ and denote the corresponding Σ by Σ_1, Σ_2 and Σ_3 respectively. We consider two different choices of the number of instruments: $k = 2, 4$. Π_1 and Π_2 are chosen 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. Weak identification (instruments) is characterized by $\mu_i = 1$ and strong identification (instruments) by $\mu_i = 10$.

The results reported are based on 10,000 Monte Carlo trials. As before, θ_1 is the coefficient of interest and θ_2 is the nuisance coefficient. We compute the empirical rejection rates of the efficient projection-based K test, the subset-K test and the projection-based AR test for a grid of θ_{*1} values around the true value θ_{01}. [6] Two representative plots are given in Figures 1 and 2, which respectively correspond to high and moderate levels of endogeneity of the regressor X_2 associated with the nuisance structural coefficient θ_2. More simulation results leading to similar observations can be found in Chaudhuri (2008). For the efficient projection-based K test we consider $\zeta = 1\%$ and $\zeta = 5\%$ in the construction of the first step confidence region $C_{2n}^S(1 - \zeta, \theta_{*1})$, and use $\epsilon = 5\%$ 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 $\epsilon = 5\%$.

In the cases when the nuisance structural coefficient θ_2 is (strongly) identified (see WI-Cases II and IV in the right panel of Figure 2), the rate at which the efficient projection-based K test (5% + 5%) rejects the *true* value of θ_1 is slightly greater than $\epsilon = 5\%$ (but less than $\zeta + \epsilon = 10\%$). This arises solely from the frequency of $C_{2n}^S(1 - \zeta, \theta_{*1}) = \varnothing$, which is summarized in Table 2. [7] However, the same phenomenon also increases the rate at which the efficient projection-based K test rejects the *false* values of θ_1. As can be seen from the figure, 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 $\epsilon = 5\%$).

When the nuisance structural coefficient θ_2 is weakly identified (see WI-Cases I and III in the left panel of the figures), Theorem 2.4 states that the asymptotic size of the efficient projection-based K test is bounded from above by $(\zeta + \epsilon)$. However, the simulations show that, even with the relatively large choice $\zeta = 5\%$, the efficient projection-based K test does not tend to over-reject (beyond $\epsilon = 5\%$) the true value of θ_1 in these specifications.

[6] Dufour (2008) pointed out the the S test which, in our linear IV setup is the Anderson-Rubin (AR) test, can often be more robust in terms of size. See Dufour and Taamouti (2005) for the projection-based AR test.

[7] 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. On the other hand, AR confidence regions can also be the entire parameter space Θ_2 (or unbounded if Θ_2 is allowed to be unbounded). However, as also shown in Table- 2, when θ_2 is identified the frequency with which such events can occur decreases with sample size; and this happens even when Θ_2 is allowed to be unbounded.

Overall, the empirical rejection rates of the subset-K test and the efficient projection based K tests corroborate the theoretical results of the last section. In a comment to Kleibergen and Mavroeidis (2008), Zivot and Chaudhuri (2008) presented Monte Carlo evidence that showed the efficient projection-based K test also performed similarly when compared to the other weak identification robust tests, 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. (As expected, it performs much better than the JKLM test in terms of power.) [8]

Finally, as Kleibergen and Mavroeidis (2009) noted, we also observe that the power of all these tests are governed by the least identified parameter. Hence, in WI-Cases I-III, all these tests have poor power (and naturally so, because the data do not contain enough information for a precise inference), whereas in the standard case, i.e. in WI-Case IV, the subset-K and the efficient projection-based K test have good and similar power even with 100 observations. (The projection-based AR test is more conservative.)

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 can lead to tests that are comparable to the tests based on the plug-in principle.

In the context of GMM with possibly weakly identified parameters, we proposed a new projection-based test for subsets of parameters and described two different versions of the test. The first version is never over-sized and at the same time is asymptotically equivalent to Kleibergen's subset-K test when the nuisance parameters are identified. However, it does not have any practical advantage over the subset-K test. The second version, on the other hand, compromises a little in terms of size but has some other desirable properties that make it useful for practical purposes. In our opinion it is the second version of the test (as described in Theorem 2.4) that should be used when it is required to use the projection-based methods of inference.

In essence, this new projection-based test is a two-step procedure. In the first step we construct a confidence region for the nuisance parameters θ_2 such that the region has (at least) the correct 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 where ν_1 is the dimension of θ_1.

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.

[8] At the suggestion of a referee we reproduce Figure 3 from the Figure 1 in Zivot and Chaudhuri (2008). The data generating process is $\pi_t = \lambda x_t + \gamma_f E_t[\pi_{t+1}] + u_t$, $x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + v_t$ and $\pi_{t+1} = (\alpha_0\rho_1 + \alpha_1)x_t + \alpha_0\rho_2 x_{t-1} + \eta_{t+1}$ where $\eta_t = u_t + \alpha_0 v_t$ and v_t are jointly normal with unit variances and correlation $\rho_{\eta v} = 0.2$.

A Appendix

Proof of Lemma 2.1: This proof follows from Stock and Wright (2000) with minor modifications. For the sake of generality, here it is not assumed that $\nabla_2 Q_n(\theta_{*1}, \tilde{\theta}_{2n}(\theta_{*1})) = 0$ and this makes the second part of the proof (i.e. the part following the proof of consistency) less simple. To prove consistency of $\tilde{\theta}_{2n}(\theta_{*1})$, note that in WI-Cases II and IV,

(i) $E[n^{-1}g_T(\theta)] = 1_{[\delta_1=1]}m_1(\theta_1) + 1_{[\delta_1=\frac{1}{2}]}n^{-1/2}\tilde{m}_{n1}(\theta) + m_2(\theta_2)$ 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

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

It follows from Assumption W that in WI-Cases II and IV, $n^{-1}Q_n(\theta_{*1}, \theta_2) \xrightarrow{P} m_2'(\theta_2)V_{gg}^{-1}(\theta_{01}, \theta_2)m_2(\theta_2)$ uniformly in $\theta_2 \in \Theta_2$. The probability limit is zero if and only if $\theta_2 = \theta_{02}$ and hence continuity of the argmin operator gives $\tilde{\theta}_{n2}(\theta_{*1}) \xrightarrow{P} \theta_{02}$.

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

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

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 $\mathcal{M} \equiv \mathcal{M}(\tilde{\theta}_*, \bar{\theta}, \theta_0) := n^{-1/2}\nabla_\theta g_T(\bar{\theta})(\tilde{\theta}_* - \theta_0)$. For any square matrix A let mineval(A) denote its minimum eigen value. Note that, $\mathcal{M}'\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)\mathcal{M} \geq \|\mathcal{M}\|^2\text{mineval}(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*))$ and $\mathcal{M}'\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)n^{-1/2}g_T(\theta_0) \geq -\|\mathcal{M}\|\|\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)n^{-1/2}g_T(\theta_0)\|$ (by the Cauchy-Schwartz inequality).

Now define $\Delta_{2n} := \|\widehat{V}_{gg}^{-1}(\tilde{\theta}_*)n^{-1/2}g_T(\theta_0)\|/\text{mineval}(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*))$ and $\Delta_{3n} := \Delta_{1n}/\text{mineval}(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*))$. Therefore, dividing (A.1) by mineval$(\widehat{V}_{gg}^{-1}(\tilde{\theta}_*))$, 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 D2 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 $\sqrt{n}\|\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02}\| = O_p(1)$ 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,

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

Again, noting that, for some $\bar{\theta}_{10} = (\bar{\theta}_1', \theta_{02}')'$ such that $\sqrt{n}\|\bar{\theta}_{10} - \theta_0\| \leq \sqrt{n}\|\theta_{*0} - \theta_0\| = O(1)$, 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 Θ, D and W. Since $V_{gg}^{-1}(\theta)$ is positive definite, similar arguments as in (A.2) give $\|\Delta_{3n}\| = O_p(1)$. Hence $\sqrt{n}\|\tilde{\theta}_{n2}(\theta_{*1}) - \theta_{02}\| = O_p(1)$. ■

Lemmas A.1, A.2 A.3 will be helpful in getting the other results. In the following we always consider θ_{n1} in the $\sqrt{n}$-neighborhood of θ_{01} (and sometimes, additionally, θ_n in the $\sqrt{n}$-neighborhood of θ_0). Hence for n large enough, such θ_{n1} (and θ_n) will belong to the neighborhood $\mathcal{T}_1$ (and $\mathcal{T}$) implying that the assumptions involving these neighborhoods of θ_{01} (and θ_0) are satisfied.

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 $\theta_0 \in \Theta^{int}$ where Θ is compact. Then the 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 $\|\widehat{a}_n(\theta_n) - a(\theta_0)\| \leq \|\widehat{a}_n(\theta) - a_n(\theta)\| + \|a_n(\theta) - a(\theta)\| + \|a(\theta) - a(\theta_0)\| = o_p(1)$.
(ii) For $x = a, b$, 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 $\|\widehat{a}_n(\theta)\widehat{b}_n(\theta) - a(\theta)b(\theta)\| \leq \|\widehat{a}_n(\theta) - a(\theta)\|\|\widehat{b}_n(\theta) - b(\theta)\| + \sqrt{p_a p}|R_a|\|\widehat{b}_n(\theta) - b(\theta)\| + \|\widehat{a}_n(\theta) - a(\theta)\|\sqrt{p p_b}|R_b| = o_p(1)$. ■

Lemma A.2 *Let $\theta_{n1} = \theta_{01} + d_1/\sqrt{n} \in \Theta_1^{int}$ where $\|d_1\| = O(1)$ and let $\theta_{n2} = \theta_{02} + d_2/\sqrt{n} \in \Theta_2^{int}$ almost surely where $\|d_2\| = O_p(1)$. Denote $\theta_n = (\theta_{n1}', \theta_{n2}')'$ 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 $\Psi_{\nabla.g} = [\Psi_{1.g}', \Psi_{2.g}']'$ and $L(\theta) = [L_1'(\theta), L_2'(\theta)]'$. Define $\widehat{h}_{Ti}(\theta) := \sum_{t=1}^n \widehat{h}_{ti}(\theta)$ where $\widehat{h}_{ti}(\theta) := vec\ \nabla_i g_t(\theta) - \widehat{V}_{ig}(\theta) \widehat{V}_{gg}^{-1}(\theta) g_t(\theta)$ for $i = 1, 2$. Then the following result holds under Assumptions Θ, D and W:*

$$
\begin{bmatrix} n^{-1/2} g_T(\theta_n) \\ n^{-\delta_1} \widehat{h}_{T1}(\theta_n) \\ n^{-\delta_2} \widehat{h}_{T2}(\theta_n) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Psi_g + \sum_{i=1}^2 1_{[\delta_i=1]} M_i(\theta_{0i}) d_i \\ vec\ G_1(\theta_0) + (1 - 1_{[\delta_1=1]}) \left[\Psi_{1.g} + L_1(\theta_0) d_\theta \right] \\ vec\ G_2(\theta_0) + (1 - 1_{[\delta_2=1]}) \left[\Psi_{2.g} + L_2(\theta_0) d_\theta \right] \end{bmatrix}.
$$

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 $h_{ti}(\theta) = vec\ \nabla_i g_t(\theta) - V_{ig}(\theta_0)V_{gg}^{-1}(\theta_0)g_t(\theta)$. Letting $h_T(\theta) = [h'_{T1}(\theta), h'_{T2}(\theta)]'$, Assumptions Θ, D and W give

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

$$\frac{1}{\sqrt{n}} g_T(\theta_0) \xrightarrow{d} \Psi_g \ \text{ and for } i = 1, 2 \ \ \frac{1}{n^{\delta_i}} h_{Ti}(\theta_0) \xrightarrow{d} vec\ G_i(\theta_0) + (1 - 1_{[\delta_i=1]})\Psi_{i.g}. \tag{A.3}$$

Hence, using (A.3) and Assumption W, after a mean-value expansion (for some mean-value $\bar{\theta}$ such that $\sqrt{n}\|\bar{\theta} - \theta_0\| \le \sqrt{n}\|\theta_n - \theta_0\| = O(1)$), we get

$$\begin{aligned} \frac{1}{\sqrt{n}} g_T(\theta_n) &= \frac{1}{\sqrt{n}} g_T(\theta_0) + \frac{1}{n}\nabla_\theta g_T(\bar{\theta})d_\theta = \frac{1}{\sqrt{n}} g_T(\theta_0) + \sum_{i=1}^{2} 1_{[\delta_i=1]} M_i(\theta_{0i})d_i + o_p(1) \quad \text{(A.4)} \\ &\xrightarrow{d} \Psi_g + \sum_{i=1}^{2} 1_{[\delta_i=1]} M_i(\theta_{0i})d_i. \end{aligned}$$

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 $\sqrt{n}\|\bar{\theta} - \theta_0\| \le \sqrt{n}\|\theta_n - \theta_0\| = O(1)$ following from a mean-value expansion of $h_{Ti}(\theta_n)$. Hence Assumption D2, Lemma A.1 and (A.3) give for $i = 1, 2$,

$$\begin{aligned} \frac{1}{n^{\delta_i}}\widehat{h}_{Ti}(\theta_n) &= \frac{1}{n^{\delta i}} h_{Ti}(\theta_0) + (1 - 1_{[\delta_i=1]})L_i(\theta_0)d_\theta + o_p(1) \quad \text{(A.5)} \\ &\xrightarrow{d} vec\ G_i(\theta_0) + (1 - 1_{[\delta_i=1]})\left[\Psi_{i.g} + L_i(\theta_0)d_\theta\right]. \ \blacksquare \end{aligned}$$

Lemma A.3 *Let $\theta_{n1} = \theta_{01} + d_1/\sqrt{n} \in \Theta_1^{int}$ where $\|d_1\| = O(1)$ and let $\theta_{n2} = \theta_{02} + d_2/\sqrt{n} \in \Theta_2^{int}$ almost surely where $\|d_2\| = O_p(1)$. Denote $\theta_n = (\theta'_{n1}, \theta'_{n2})'$ 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 $\Psi_{\nabla.g} = [\Psi'_{1.g}, \Psi'_{2.g}]'$ and $L(\theta) = [L'_1(\theta), L'_2(\theta)]'$. Then the following result holds under Assumptions Θ, D and W:*

(i) (a) $K_{n1}(\theta_n) \xrightarrow{d} \mathbb{B}'(d_\theta)P\left(N\left(\mathbb{A}_2(d_\theta)\right)\mathbb{A}_1(d_\theta)\right)\mathbb{B}(d_\theta)$
(b) $S_n(\theta_n) \xrightarrow{d} \mathbb{B}'(d_\theta)\mathbb{B}(d_\theta)$
where $\mathbb{A}_i(d_\theta) := V_{gg}^{-\frac{1}{2}'}(\theta_0)[G_i(\theta_0) + (1 - 1_{[\delta_i=1]})devec_k[\Psi_{i.g} + L_i(\theta_0)d_\theta]']$ for $i = 1, 2$ and $\mathbb{B}(d_\theta) := V_{gg}^{-\frac{1}{2}'}(\theta_0)\ [\Psi_g + \sum_{i=1}^{2} 1_{[\delta_i=1]} M_i(\theta_{0i})d_i]$.

(ii) Let $\theta_2^\dagger = \theta_{02} + d_2^\dagger/n^p \in \Theta_2^{int}$ almost surely where $p \ge 0$ and bounded away from 1/2 by a fixed real number r (say). Then plim $n^{2p-1}S_n(\theta_{n1}, \theta_2^\dagger) > 0$ in WI-Cases II and IV.

Proof: (i) Note that $\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta_n)n^{-1/2}g_T(\theta_n) \xrightarrow{d} \mathbb{B}(d_\theta)$. Define $A_{ni} := \widehat{V}_{gg}^{-\frac{1}{2}'}(\theta_n)\widehat{D}_{Ti}(\theta_n)$ for $i = 1, 2$. Lemma A.2 gives $n^{-\delta_i}A_{ni} \xrightarrow{d} V_{gg}^{-\frac{1}{2}'}(\theta_0)\left[G_i(\theta_0) + (1 - 1_{[\delta_i=1]})devec_k\left[\Psi_{i.g} + L_i(\theta_0)d_\theta\right]'\right] = \mathbb{A}_i(d_\theta)$. The result in (a) follows once we note that

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

The result in (b) directly follows from the fact that $\widehat{V}_{gg}^{-\frac{1}{2}'}(\theta_n)n^{-1/2}g_T(\theta_n) \xrightarrow{d} \mathbb{B}(d_\theta)$.

(ii) It is instructive to prove this result in two parts: part 1 with $p = 0$, i.e. for fixed deviation from θ_{02} and part 2 for $p \in (0, r < 1/2)$. Part 1 ($p = 0$): From (i) in the proof of Lemma 2.1 and Assumption D(i) it follows that in WI-Cases II and IV, $n^{-1}g_T(\theta_{n1}, \theta_2) \xrightarrow{P} E[n^{-1}g_T(\theta)] \to m_2(\theta_2)$, which is not equal to zero using (1.1) unless $\theta_2 \to \theta_{02}$. Therefore, noting that $\widehat{V}_{gg}(\theta_{n1}, \theta_2) \xrightarrow{P} V_{gg}(\theta_{01}, \theta_2)$, which is positive definite and bounded (Assumption D4), it follows that the probability limit of $n^{-1}S_n(\theta_{n1}, \theta_2) := [n^{-1}g_T(\theta_{n1}, \theta_2)]'\widehat{V}_{gg}^{-1}(\theta_{n1}, \theta_2)[n^{-1}g_T(\theta_{n1}, \theta_2)]$ is positive.
Part 2 ($p \in (0, r < 1/2)$) : In the following we do a slight abuse of the notation. Typically the subscripts i and j have till now been assigned for θ_1 and θ_2. However, here we have let i and j denote elements of θ_1 and θ_2. Note that, by construction, $d_2^\dagger$ is bounded (in probability). Now, a second order expansion gives

$$\frac{n^{p-1/2}}{n^{1/2}}g_T(\theta_{n1}, \theta_2) = \underbrace{\frac{n^{p-1/2}}{n^{1/2}}g_T(\theta_{01}, \theta_{02})}_{=T_{n1}} + \underbrace{\frac{n^{p-1/2}}{n}\nabla_1 g_T(\theta_{01}, \theta_{02})d_1}_{=T_{n2}} + \underbrace{\frac{1}{n}\nabla_2 g_T(\theta_{01}, \theta_{02})d_2^\dagger}_{=T_{n3}} + \underbrace{\sum_{l=1}^{3} R_{nl}(\bar{\theta})}_{=T_{n4}}$$

for some $\bar{\theta}$ such that $\|\bar{\theta} - \theta_0\| \leq \|(\theta'_{n1}, \theta'_2)' - \theta_0\|$. Now, partitioning $L_i(\theta) = [L_{i1}(\theta), L_{i2}(\theta)]$ with respect to θ_1 and θ_2 for $i = 1, 2$ we note the following using Assumption D2(iii):

$$\begin{aligned}
R_{n1}(\bar{\theta}) &= \frac{n^{p-1/2}}{2n^{3/2}}\sum_{i=1}^{\nu_1}\sum_{j=1}^{\nu_1}\frac{\partial^2}{\partial\theta_{1i}\partial\theta_{1j}}\nabla_{(ij)}g_T(\bar{\theta})d_{1i}d_{1j} = \frac{n^{p-1/2}}{2\sqrt{n}}\sum_{i=1}^{\nu_1}\sum_{j=1}^{\nu_1}L_{11}(\theta_{01}, \bar{\theta}_2)d_{1i}d_{1j} + o_p(1) \xrightarrow{P} 0, \\
R_{n2}(\bar{\theta}) &= \frac{n^{p-1/2}}{2n^{1+p}}\sum_{i=1}^{\nu_1}\sum_{j=1}^{\nu_2}\frac{\partial^2}{\partial\theta_{1i}\partial\theta_{2j}}\nabla_{(ij)}g_T(\bar{\theta})d_{1i}d_{2j} = \frac{1}{2n^{1/2}}\sum_{i=1}^{\nu_1}\sum_{j=1}^{\nu_2}L_{12}(\theta_{01}, \bar{\theta}_2)d_{1i}d_{1j} + o_p(1) \xrightarrow{P} 0, \\
R_{n3}(\bar{\theta}) &= \frac{n^{p-1/2}}{2n^{1/2+2p}}\sum_{i=1}^{\nu_2}\sum_{j=1}^{\nu_2}\frac{\partial^2}{\partial\theta_{2i}\partial\theta_{2j}}\nabla_{(ij)}g_T(\bar{\theta})d_{2i}d_{2j} = \frac{1}{2n^p}\sum_{i=1}^{\nu_2}\sum_{j=1}^{\nu_2}L_{22}(\theta_{01}, \bar{\theta}_2)d_{1i}d_{1j} + o_p(1) \xrightarrow{P} 0.
\end{aligned}$$

Finally noting that $T_{n1} \xrightarrow{P} 0$ (using (1.1) and Assumption D2(i)), $T_{n2} \xrightarrow{P} 0$ (using Assumptions W and D3) and $T_{n3} \xrightarrow{P} M_2(\theta_{02})d_2^\dagger \neq 0$, it follows using similar arguments as in Part 1 that $n^{2p-1}S_n(\theta_{n1}, \theta_2^\dagger) \xrightarrow{P} d_2^{\dagger'}M'_2(\theta_{02})V_{gg}^{-1}(\theta_{01}, \theta_2^\dagger)M_2(\theta_{02})d_2^\dagger > 0$. Therefore, since $p < 1/2$ is bounded away from 1/2, it follows that $S_n(\theta_{n1}, \theta_2^\dagger) \to \infty$ whenever $\theta_2^\dagger$ is outside the $\sqrt{n}$-neighborhood of θ_{02}. ■

Remarks: We make the following remarks about the S statistic, which we eventually use to obtain the first step confidence region for the efficient projection-based K test in Theorem 2.4.

(i) $S_n(\theta_n)$ is the S statistic proposed by Stock and Wright (2000). Under WI-Cases I-IV, letting

$d_\theta = 0$ in Lemma A.3(i.b), $S_n(\theta_0) \xrightarrow{d} \Psi_g' V_{gg}^{-1}(\theta_0)\Psi_g \sim \chi_k^2$ using Assumption D3. Hence if θ_{01} is known *a priori*, the test that rejects $H_2 : \theta_2 = \theta_{*2}$ if $S_n(\theta_{01}, \theta_{*2}) > \chi_k^2(1-\zeta)$ has asymptotic size ζ [same as in Stock and Wright (2000)]. Therefore, a confidence region defined by $\mathcal{C}_{2n}^S(1-\zeta, \theta_1) := \{\theta_2 : S_n(\theta_1, \theta_2) \leq \chi_k^2(1-\zeta)\}$ has the (uniform) asymptotic coverage probability $1-\zeta$ for θ_{02} when $\theta_1 = \theta_{01}$. This is true under all the four cases of weak identification determined by the nuisance parameters δ_1 and δ_2.

(ii) In WI-Cases II and IV, $S_n(\theta_n)$ is $O_p(1)$ in the $\sqrt{n}$-neighborhood of θ_0 and converges to a non-central χ_k^2 distribution with a finite non-centrality parameter

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

In addition, the result in Lemma A.3(ii) says that the S statistic is not finite in WI-Cases II and IV when evaluated outside the $\sqrt{n}$-neighborhood of θ_{02}. Therefore, in WI-Cases II and IV the confidence region $\mathcal{C}_{2n}^S(1-\zeta, \theta_{n1})$ is contained in the $\sqrt{n}$-neighborhood of θ_{02}.

Proof of Lemma 2.2: (i) Consider the asymptotic distribution of $K_{n1}(\theta_{n1}, \theta_{n2})$ in Lemma A.3(i.a). At $\theta = \theta_0$, $\mathbb{B}(0) = V_{gg}^{-\frac{1}{2}'}(\theta_0)\Psi_g$ is independent of $\mathbb{A}_i(0) = V_{gg}^{-\frac{1}{2}'}(\theta_0)\left[G_i(\theta_0) + (1 - 1_{[\delta_i=1]}) devec_k \Psi_{i.g}\right]$ for $i = 1, 2$ (see Lemma A.2). Therefore, conditional on $\Psi_{1.g}$ and $\Psi_{2.g}$, $K_{n1}(\theta_0) \overset{A}{\sim} \chi^2$ distribution with degrees of freedom equal to ν_1, i.e. the rank (trace) of the idempotent matrix $P\left(N\left(\mathbb{A}_2(0)\right)\mathbb{A}_1(0)\right)$. Because of the independence of Ψ_g with $\Psi_{i.g}$ for $i = 1, 2$, this result also holds unconditionally. It is important to note that the result holds for any θ_0 satisfying the moment restrictions in (1.1) and Assumptions Θ and D, and under all the four cases of weak identification determined by the nuisance parameters δ_1 and δ_2. [9]

In the proof of (ii)-(iv) we use the same notations (particularly those involving the h's) as in Lemma A.2. Let us first note that under WI-Cases II and IV, we have

(a) $n^{-1/2} g_T(\theta_n) = n^{-1/2} g_T(\theta_0) + 1_{[\delta_1=1]} M_1(\theta_{01}) d_1 + M_2(\theta_{02}) d_2 + o_p(1)$ using (A.4).

(b) $n^{-\delta_1} \widehat{h}_{T1}(\theta_n) = n^{-\delta_1} h_{T1}(\theta_0) + (1 - 1_{[\delta_1=1]}) L_1(\theta_0) d_\theta + o_p(1)$ using (A.5).

(c) $n^{-1} \widehat{h}_{T2}(\theta_n) = n^{-1} h_{T2}(\theta_0) + o_p(1) = E\, n^{-1} \sum_{t=1}^n vec\, \nabla_2 g_t(\theta_0) - V_{2g}(\theta_0) V_{gg}^{-1}(\theta_0) E\, n^{-1} \sum_{t=1}^n g_t(\theta_0) + o_p(1) = vec\ M_2(\theta_{02}) + o_p(1)$ using (A.5), Assumption W and (2.2) respectively.

(d) $N(\hat{V}_{gg}^{-1/2'}(\theta_n) devec_k(n^{-1}\widehat{h}_{T2}(\theta_n))')\hat{V}_{gg}^{-1/2'}(\theta_n) n^{-1/2} g_T(\theta_n) = N(V_{gg}^{-1/2'}(\theta_0) M_2(\theta_{02})) V_{gg}^{-1/2'}(\theta_0)$ $[n^{-1/2} g_T(\theta_0) + 1_{[\delta_1=1]} M_1(\theta_{01}) d_1]$ from (a) and (c) (and using Assumption D4) since, by definition, $N(V_{gg}^{-1/2'}(\theta_0) M_2(\theta_{02})) V_{gg}^{-1/2'}(\theta_0) M_2(\theta_{02}) d_2 = 0$.

In addition, under WI-Case IV, (b) can be modified in the same way as (c) to get

(e) $n^{-\delta_1} \widehat{h}_{T1}(\theta_n) = vec\ M_1(\theta_{01}) + o_p(1)$.

[9] We are grateful to the referee for suggesting to add explanations to this crucial result.

(ii) Therefore, under WI-Case IV, and using (a), (c), (d), (e) we have $K_{n1}(\theta_n) = K_{n1}(\theta_{n1}, \theta_{02}) + o_p(1) = \Xi_n + o_p(1)$ where

$$\begin{aligned}\Xi_n &= \left[\frac{1}{\sqrt{n}} g_T(\theta_0) + M_1(\theta_{01}) d_1\right]' V_{gg}^{-\frac{1}{2}}(\theta_0) P\left(N\left(V_{gg}^{-\frac{1}{2}'}(\theta_0) M_2(\theta_{02})\right) V_{gg}^{-\frac{1}{2}'}(\theta_0) M_2(\theta_{02})\right) \\ &\quad \times V_{gg}^{-\frac{1}{2}'}(\theta_0)\left[\frac{1}{\sqrt{n}} g_T(\theta_0) + M_1(\theta_{01}) d_1\right] \\ &\overset{A}{\sim} \chi^2_{\nu_1}(\eta(d_1)'\eta(d_1)).\end{aligned}$$

(iii) Furthermore, under Assumption O1 and for n large enough such that $\theta_{n1} \in \mathcal{T}_1$, we have $\nabla_2 Q_n(\theta_{n1}, \tilde{\theta}_{n2}(\theta_{n1})) = 0$, which further implies that for large n, $K_n(\theta_{n1}, \tilde{\theta}_{n2}(\theta_{n1})) = K_{n1}(\theta_{n1}, \tilde{\theta}_{n2}(\theta_{n1}))$. Lemma 2.1 implies that $\tilde{\theta}_{n2}(\theta_{n1})$ is $\sqrt{n}$-consistent for θ_{02} under WI-Case IV, and hence the result follows directly from (i).

(iv) From the expression of $n^{-\delta_1}\widehat{h}_{T1}(\theta_n)$ in (b) it can be seen that even under WI-Case II, the results from (ii) and (iii) will hold (although the unconditional limiting distribution will no longer be χ^2) if the last ν_2 columns of $L_1(\theta_0)$ are zeros, which will make the right hand side of (b) independent of d_2 (i.e. $\sqrt{n}$ times the deviation from θ_{02}). We verify this now. Denoting them $L_{12}(\theta_0)$, under WI-Case II the last ν_2 columns of $L_1(\theta_0)$ are

$$\begin{aligned}L_{12}(\theta_0) &= \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta_2'} vec \frac{\partial}{\partial\theta_1'} g_t(\theta)\right]_{\theta_0} = \lim_{n\to\infty} \frac{\partial}{\partial\theta_1'}\frac{\partial}{\partial\theta_2'} E\left[\frac{1}{n}\sum_{i=1}^{n} vec\ g_t(\theta)\right]_{\theta_0} \quad (A.6) \\ &= \lim_{n\to\infty} \frac{\partial}{\partial\theta_1'}\left[M_2(\theta_2) + \frac{1}{\sqrt{n}}\tilde{M}_{n1}^{(2)}(\theta_1, \theta_2)\right]_{\theta_0} = 0 \quad (A.7)\end{aligned}$$

where (A.6) follows from Assumptions D1.(ii) and O2(i) and (A.7) follows from Assumption O2(ii). For the parts involving $n^{-1/2}g_T(\theta_n)$ and $n^{-1}\widehat{h}_{T2}(\theta_n)$, the same arguments as in (i) and (ii) apply. Hence the proof follows noting that Lemma 2.1 implies that $\tilde{\theta}_{n2}(\theta_{n1})$ is $\sqrt{n}$-consistent for θ_{02} also under WI-Case II. ■

Proof of Theorem 2.3: $\mathcal{C}_{2n}(\theta_{*1})$ always contains $\tilde{\theta}_{2n}(\theta_{*1})$ and hence $\inf_{\theta_{*2}\in\mathcal{C}_{2n}(\theta_{*1})} K_{n1}(\theta_{*1}, \theta_{*2})$ always exits.

(i) Therefore, by definition, $\inf_{\theta_{*2}\in\mathcal{C}_{2n}(\theta_{*1})} K_{n1}(\theta_{*1}, \theta_{*2}) \leq K_{n1}(\theta_{*1}, \tilde{\theta}_{2n}(\theta_{*1}))$, which under the conditions of the theorem is equal to $K_n(\theta_{*1}, \tilde{\theta}_{2n}(\theta_{*1}))$. Now, Theorem 12 in Kleibergen and Mavroeidis (2009) shows that the asymptotic distribution of $K_n(\theta_{01}, \tilde{\theta}_{2n}(\theta_{01}))$ is always bounded from above by a (central) $\chi^2_{\nu_1}$ distribution. Therefore, transitivity implies that the asymptotic size of the efficient projection-based K test is always bounded from above by the asymptotic size of the subset-K test, which in turn, cannot exceed ϵ [because of Theorem 12 in Kleibergen and Mavroeidis (2009)].

(ii) Recall that $\theta_{*1} = \theta_{01} + d_1/\sqrt{n}$. Now consider any arbitrary sequence $\{\theta_{2n} \in \mathcal{C}_{2n}(\theta_{*1})\}_n$. By the conditions of the theorem, the n-th element from this arbitrary sequence can be expressed as $\theta_{02} + d_2/\sqrt{n}$ for some $d_2 = O_p(1)$. Therefore, the n-th element of the sequence $\{\theta_2^{\inf}(\theta_{*1}) \in \mathcal{C}_{2n}(\theta_{*1})\}_n$, where the infimum $\inf_{\theta_{*2}\in\mathcal{C}_{2n}(\theta_{*1})} K_{n1}(\theta_{*1}, \theta_{*2})$ is attained, can also be expressed as $\theta_{02} + d_2^{\inf}/\sqrt{n}$ for some $d_2^{\inf} = O_p(1)$. Hence the result follows from Lemma 2.2(ii)-(iv). ■

Proof of Theorem 2.4: In the following, whenever we refer to $\inf_{\theta_{*2}\in\mathcal{C}^S_{2n}(1-\zeta,\theta_1)} K_{n1}(\theta_1,\theta_{*2})$, it is implied that $\mathcal{C}^S_{2n}(1-\zeta,\theta_1)$ is nonempty.

(i) For $\theta_0 \in$ interior(Θ) and under WI-Cases I-IV, the confidence region $\mathcal{C}^S_{2n}(1-\zeta,\theta_{01})$ has an uniform asymptotic coverage probability $(1-\zeta)$ [see Remark (i) following the proof of Lemma A.3]. Again, from Lemma 2.2(i) we know that $K_{n1}(\theta_{01},\theta_{02}) \overset{A}{\sim} \chi^2_{\nu_1}$ under WI-Cases I-IV. This limiting distribution does not depend on $\theta_0 = (\theta'_{01},\theta'_{02})$ and holds uniformly for all $\theta_0 \in \Theta^{\text{int}}$ satisfying (1.1) and all the assumptions. Therefore, whenever θ_{02} is contained in $\mathcal{C}^S_{2n}(1-\zeta,\theta_{01})$ (which happens with asymptotic probability $1-\zeta$), it follows that

$$\inf_{\theta_{*2}\in\mathcal{C}^S_{2n}(1-\zeta,\theta_{01})} K_{n1}(\theta_{01},\theta_{*2}) \le K_{n1}(\theta_{01},\theta_{02}) \le \chi^2_{\nu_1}(1-\epsilon)$$

with probability approaching $1-\epsilon$ uniformly across WI-Cases I-IV and for all $\theta_{02} \in \Theta_2^{\text{int}}$ such that $\theta_0 = (\theta'_{01},\theta'_{02})'$ satisfies (1.1) and all the assumptions. Hence, using Bonferroni arguments, it follows that the asymptotic size of the efficient projection-based K test cannot exceed $1-(1-\zeta)(1-\epsilon) \le \zeta+\epsilon$.

(ii) The confidence region $\mathcal{C}^S_{2n}(1-\zeta,\theta_{*1})$ is assumed to be nonempty and hence $\inf_{\theta_{*2}\in\mathcal{C}^S_{2n}(1-\zeta,\theta_{01})} K_{n1}(\theta_{01},\theta_{*2})$ exists. Now from Lemma A.3(i.a) and (ii) (and Remark (ii) following the proof) and using the same arguments as in the proof of Theorem 2.3(ii), the n-th element of the sequence $\{\theta_2^{\inf}(\theta_{*1}) \in \mathcal{C}^S_{2n}(1-\zeta,\theta_{*1})\}_n$, where the infimum $\inf_{\theta_{*2}\in\mathcal{C}^S_{2n}(1-\zeta,\theta_{*1})} K_{n1}(\theta_{*1},\theta_{*2})$ is attained, can also be expressed as $\theta_{02}+d_2^{\inf}/\sqrt{n}$ for some $d_2^{\inf}=O_p(1)$. Hence, as before, the result follows directly from Lemma 2.2(ii)-(iv). ■

Proof of Corollary 2.5: The result follows using the asymptotic equivalence between the efficient projection-based K test and the subset-K test after we note that

$$\begin{aligned}
\mathcal{P}\left[\inf_{\theta_{*2}\in\Theta_2} K_n(\theta_{*1},\theta_{*2}) > \chi^2_\nu(1-\epsilon)\right] &\le \mathcal{P}\left[K_n\left(\theta_{*1},\tilde\theta_{n2}(\theta_{*1})\right) > \chi^2_\nu(1-\epsilon)\right] \\
&= \mathcal{P}\left[K_{n1}\left(\theta_{*1},\tilde\theta_{n2}(\theta_{*1})\right) > \chi^2_\nu(1-\epsilon)\right] \text{ (from Assumption O1)} \\
&\le \mathcal{P}\left[K_{n1}\left(\theta_{*1},\tilde\theta_{n2}(\theta_{*1})\right) > \chi^2_{\nu_1}(1-\epsilon)\right]
\end{aligned}$$

where $\mathcal{P}$ is the probability measure that considers θ_0 as the true value of θ (see paragraph one in the introduction). ■

Acknowledgements

We thank Ron Gallant (co-editor), an anonymous associate editor, two anonymous referees, Byron Tsang, Eric Renault, Dick Startz, Donald Andrews, Frank Kleibergen, Jean-Marie Dufour, Jamie Robins, Jonathan Hill, Sophocles Mavroeidis, Thomas Richardson, the seminar participants at University of Montreal, NCSU, UNC Chapel Hill, Queen's and Waterloo and the participants in the ESRC Study Group Conference (2007), Midwest Econometrics Conference (2007) and the Summer (2009) and Winter(2008) meetings of the Econometric Society for helpful suggestions. Zivot acknowledges the financial support from the Gary Waterman Distinguished Scholar fund.

References

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.

Conniffe, D. (2001). Score tests when a nuisance parameter is unidentified under the null hypothesis. *Journal of Statistical Planning and Inference*, 97: 67–83.

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. (2008). Comments on "Weak instrument robust tests in GMM and the new Keynesian Phillips curve" by F. Kleibergen and S. Mavroeidis. Discussion Paper: McGill University (Department of Economics), CIREQ and CIRANO.

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

Kleibergen, F. and Mavroeidis, S. (2009). Inference on Subsets of Parameters in GMM without Assuming Identification. Technical report, Brown University Working Paper.

Moon, H. R. and Schorfheide, F. (2009). Estimation with Overidentifying Inequality Moment Conditions. Technical report, University of Southern California and University of Pennsylvania.

Newey, W. K. and McFadden, D. L. (1994). Large Sample Estimation and Hypothesis Testing. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume IV, chapter 36, pages 2212–2245. Elsevier Science Publisher.

Neyman, J. (1959). Optimal Asymptotic Test of Composite Statistical Hypothesis. In Grenander, U., editor, *Probability and Statistics, the Harald Cramer Volume*, pages 313–334. Almqvist and Wiksell, Uppsala.

Phillips, P. C. B. (1989). Partially Identified Econometric Models. *Econometric Theory*, 5: 181–240.

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.

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.

Wang, J. and Zivot, E. (1998). Inference on a Structural Parameter in Instrumental Variables Regression with Weak Instruments. *Econometrica*, 66: 1389–1404.

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

The covariance matrix for the structural errors is assumed to be of the form:

$$\Sigma = \begin{bmatrix} 1 & \rho_{u1} & \rho_{u2} \\ \rho_{u1} & 1 & 0 \\ \rho_{u2} & 0 & 1 \end{bmatrix}$$

In Table 2 Σ_1, Σ_2 and Σ_3 are characterized by:

1. Σ_1: $\rho_{u1} = 0.5$ and $\rho_{u2} = 0.5$
2. Σ_2: $\rho_{u1} = 0.1$ and $\rho_{u2} = 0.99$
3. Σ_3: $\rho_{u1} = 0.99$ and $\rho_{u2} = 0.1$.

% of times $\mathcal{C}_{2n}^S(1-\zeta,\theta_{01})$ is Empty!

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
10^3	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
10^4	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

% of times $\mathcal{C}_{2n}^S(1-\zeta,\theta_{01}) = [a,b]$ for some $-\infty < a < b < +\infty$ (Θ_2 allowed to be unbounded)

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	7.53	19.91	88.21	95.05	8.05	20.16	88.46	95.30
10^2	2	Σ_2	7.51	18.46	87.80	94.75	7.30	18.93	88.59	94.69
10^2	2	Σ_3	7.57	20.21	88.23	95.28	8.05	20.22	88.13	95.29
10^2	4	Σ_1	11.77	26.53	98.46	97.48	11.59	26.63	98.48	97.09
10^2	4	Σ_2	10.72	24.18	98.15	97.06	11.12	24.54	98.20	96.74
10^2	4	Σ_3	11.62	26.72	98.57	97.59	11.39	27.32	98.51	97.25
10^3	2	Σ_1	8.91	22.89	94.91	97.56	9.03	23.28	95.02	97.43
10^3	2	Σ_2	8.68	21.97	94.65	97.34	8.56	21.14	94.85	97.33
10^3	2	Σ_3	8.91	23.44	94.87	97.51	8.80	22.35	95.19	97.79
10^3	4	Σ_1	12.72	29.04	99.26	97.60	12.86	29.79	99.55	97.84
10^3	4	Σ_2	12.84	27.54	99.17	97.53	12.05	27.10	99.41	97.72
10^3	4	Σ_3	12.02	29.34	99.17	97.74	12.84	29.62	99.42	97.97
10^4	2	Σ_1	8.88	22.23	93.86	97.17	8.36	21.51	94.00	97.31
10^4	2	Σ_2	8.26	21.02	93.89	96.97	7.96	21.19	93.90	96.90
10^4	2	Σ_3	8.37	22.06	93.87	97.07	8.40	22.05	94.29	97.16
10^4	4	Σ_1	13.66	31.25	99.52	97.90	15.23	32.46	99.64	97.98
10^4	4	Σ_2	12.89	29.02	99.53	97.57	14.45	30.67	99.63	97.82
10^4	4	Σ_3	13.82	31.77	99.60	97.95	14.77	32.18	99.70	98.11

Table 2: Two important types of 1st-step confidence interval for θ_2 obtained by inverting the S test under the restriction $\theta_1 = \theta_{01}$.

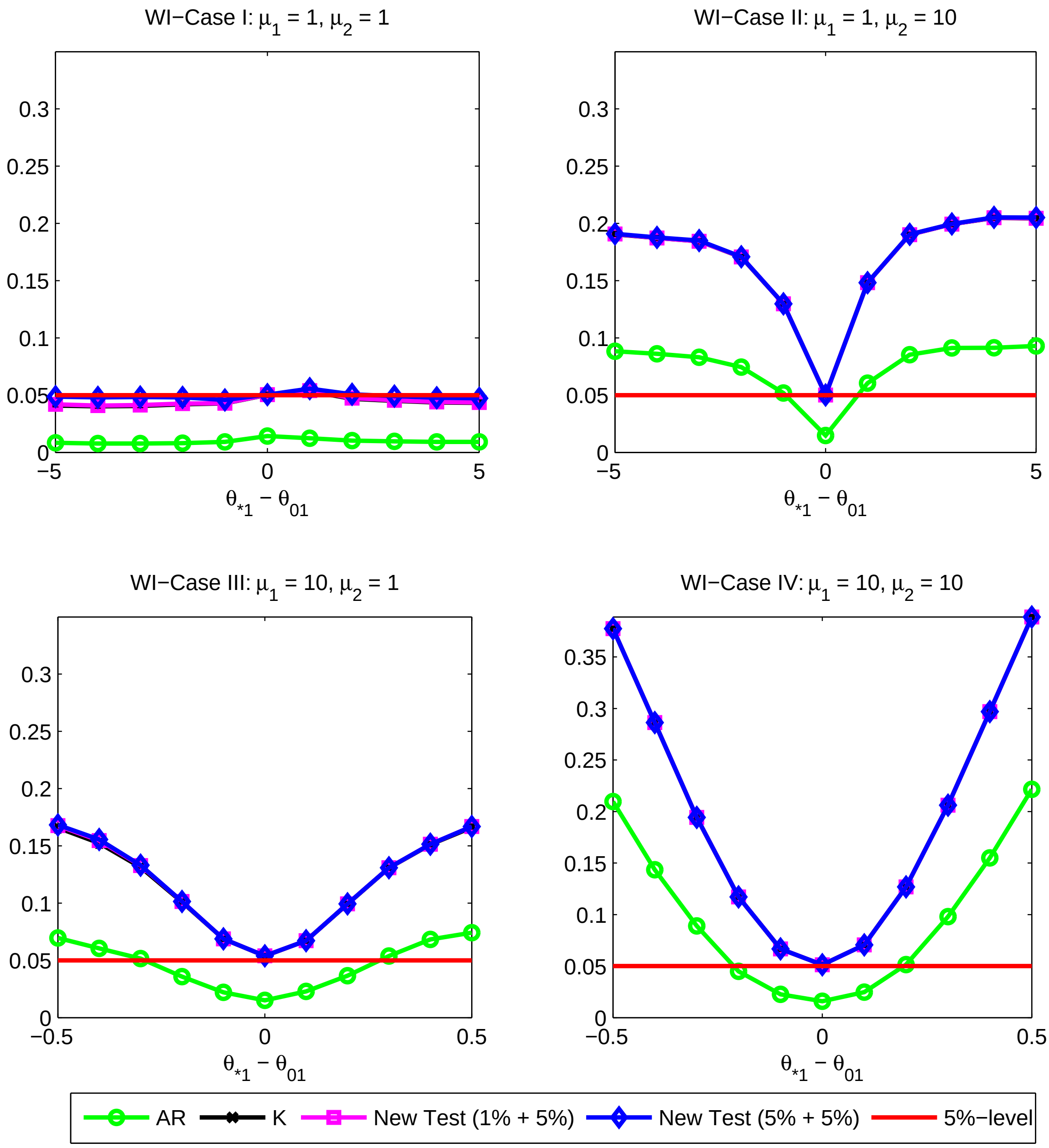

Figure 1: Rejection rate of $H_1 : \theta_1 = \theta_{*1}$ with Sample Size = 100, 2 Instruments, $\rho_{u1} = 0.1$, $\rho_{u2} = 0.99$ and $\rho_{12} = 0$. Weak identification characterized by $\mu = 1$ and strong identification by $\mu = 10$.

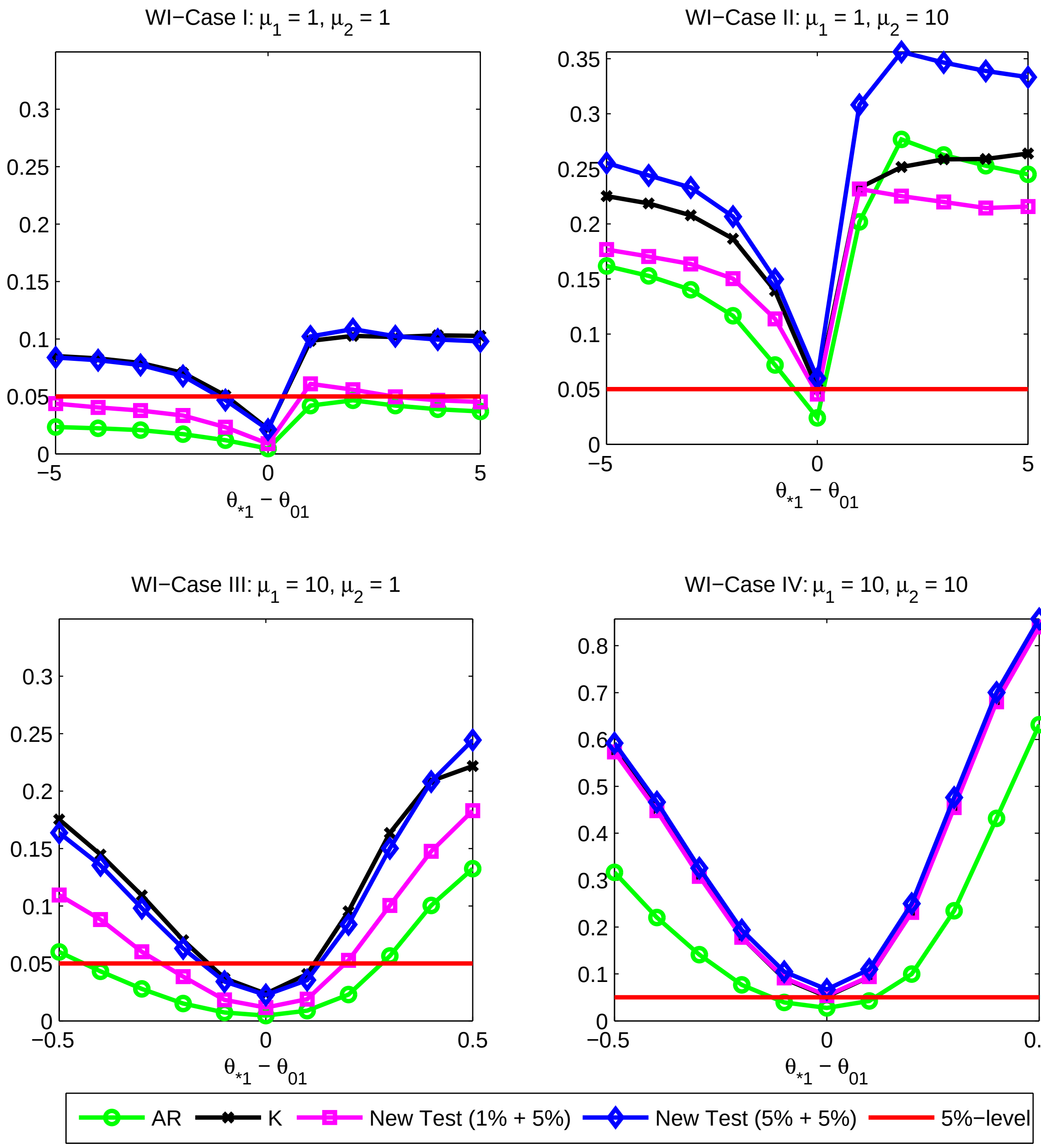

Figure 2: Rejection rate of $H_1 : \theta_1 = \theta_{*1}$ with Sample Size $= 100$, 4 Instruments, $\rho_{u1} = 0.5$, $\rho_{u2} = 0.5$ and $\rho_{12} = 0$. Weak identification characterized by $\mu = 1$ and strong identification by $\mu = 10$.

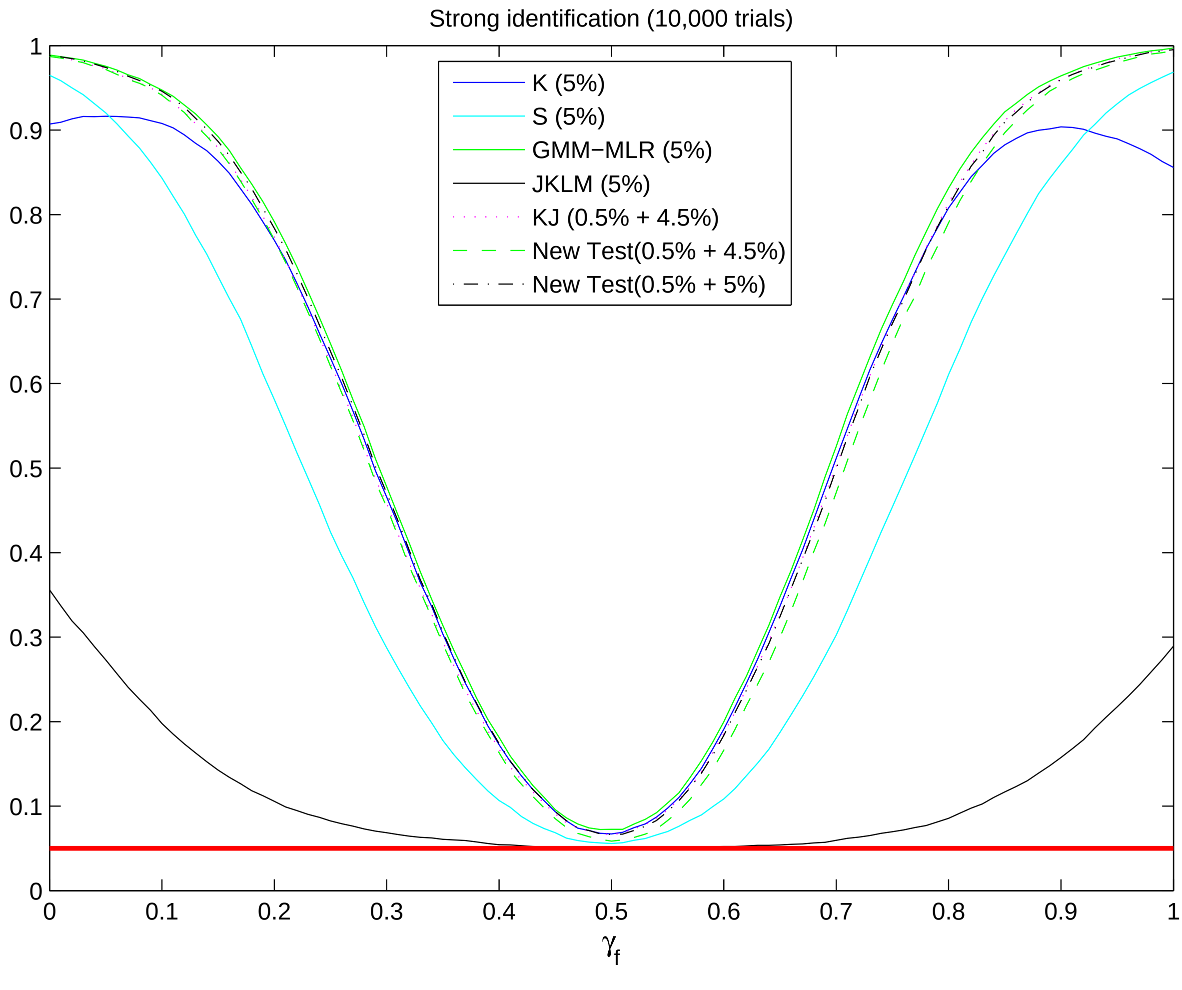

Figure 3: Rejection rate of weak identification robust tests for testing γ_f (true value $= 0.5$) with sample size $= 1000$ and 6 Instruments. Strong identification characterized by $\mu = 36$ [Zivot and Chaudhuri (2008)].