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

Saraswata Chaudhuri[*] and Eric Zivot[†]

Abstract

Projection-based tests for subsets of parameters are useful because they do not over-reject the true parameter values when either it is difficult to estimate the nuisance parameters or their identification status is questionable. However, they are also often criticized for being overly conservative. We overcome this conservativeness by introducing a new projection-based test that is more powerful than the traditional projection-based tests. The new test is even asymptotically equivalent to the related plug-in-based tests when all the parameters are identified. Extension to models with weakly identified parameters shows that the new test is not dominated by the related plug-in-based tests.

JEL Classification: C12; C13; C30

Keywords: Projection test; Nuisance parameters; $C(\alpha)$ statistic; GMM; Weak identification

[*]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

The usefulness of the traditional projection principle in designing tests that are not over-sized has 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 (2005b, 2007)]. However, there are two reasons why these tests are also often criticized for being overly conservative. First, they use conventional critical values that are conservative. Second, the test statistics used are typically smaller than the corresponding plug-in-based test statistics rendering the conventional critical values even more conservative.

The purpose of this paper is to introduce a new method of projection-based inference that overcomes these two problems and reduces the conservativeness generally associated with the traditional projection principle. In addition, when the parameters in the model are identified, the new method can be made asymptotically equivalent to the plug-in-based methods; a feature that can be quite useful when the plug-ins required for the latter are computationally difficult to estimate.

We describe the new method in the context of testing for subsets of parameters in a moment conditions model. The setup is described below. Consider a set of parameters θ whose unknown "true value" θ_0 is defined by the moment restrictions

$$E\left[g_n(w_t, \theta)\right] = 0 \Longleftrightarrow \theta = \theta_0 \tag{1.1}$$

where $g_n : \mathfrak{S} \times \Theta \mapsto \mathbb{R}^k$ is a known measurable function possibly dependent on sample size n, $\Theta \subset \mathbb{R}^\nu$ $(\nu \leq k)$ is the parameter space, $\{w_t \in \mathfrak{S} : t = 1, \ldots, n\}$ is the sample of observations from the sample space $\mathfrak{S}$, and $E[.]$ is the expectation with respect to a probability measure P_0 that considers θ_0 as the true value of θ.

Consider the partition $\theta = (\theta_1', \theta_2')'$, $\theta_0 = (\theta_{01}', \theta_{02}')'$ and $\theta_1 \in \Theta_1, \theta_2 \in \Theta_2$ where $\Theta_1 \times \Theta_2 = \Theta$. We are interested in the projection-based inference on the subsets of parameters θ_1, i.e., in testing

$$H^0 : \theta_1 = \theta_1^0 \text{ versus } H^1 : \theta_1 \neq \theta_1^0,$$

by treating θ_2 as the unknown nuisance parameters. We restrict attention to $\theta_1^0 \in \Theta_1$ in the

$\sqrt{n}$-neighborhood of θ_{01} when considering the properties of the testing procedures.

In Section 2 we briefly describe the traditional projection principle for such testing and explain the reasons for its conservativeness. We refer to a test as "conservative" if its asymptotic size is less than the allowable rate of Type-I error pre-specified by the user. Conservativeness (in terms of size) of a test leads to low power. In Section 3 we introduce our new method of projection and show how it overcomes the conservativeness of the traditional projection methods. We also show that the new method achieves ($\sqrt{n}$-local) asymptotic equivalence with the plug-in-based methods of inference when θ is identified (to be defined precisely in assumption W). This is done by the use of Neyman (1959)'s $C(\alpha)$ principle and a form of restricted projection that is motivated by Dufour (1990) and Robins (2004). Section 4 extends the new method to models where some or all the parameters are weakly identified in the sense of Stock and Wright (2000). The extension is non trivial and non unique because the feasible subset tests (both projection and plug-in-based), that are generally recognized to work best, and our new method are only boundedly asymptotically pivotal under the generality of our weak identification assumptions. This hinders a neat analytical comparison of power. However, an extensive Monte-Carlo experiment shows that the performance of the new method is always better than the traditional projection methods and is not always dominated by the related plug-in methods proposed by Kleibergen and Mavroeidis (2009), even in relatively small samples. A subset of our simulation results is reported in Section 4. We conclude in Section 5. Proofs of all results are collected in the Appendix.

We use the following notations. For an $a \times b$ matrix A with full column rank, $P(A) := A(A'A)^{-1}A'$ 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 $A = A^{\frac{1}{2}} A^{\frac{1}{2}'}$. χ^2_v and $\chi^2_v(1-\epsilon)$, for $\epsilon \in (0,1)$, are used to denote the distribution and the $1-\epsilon$ quantile, respectively, of a central chi-square distribution with v degrees of freedom. Unless confusing, we suppress the explicit dependence of all the functions on the sample size n and the data and, for e.g., write $g_n(w_t, \theta)$ as $g_t(\theta)$. For any $X_1, X_2, \ldots, X_n$, we use $\bar{X}_n$ to denote the sample average $\sum_{t=1}^{n} X_t/n$.

2 The traditional projection principle

In this section we briefly describe the methods of inference based on the traditional projection principle and discuss the reason for their conservativeness with respect to the plug-in-based methods. The conservativeness in terms of size arises because these projection methods are only boundedly asymptotically pivotal. The conservativeness leads to the low power of the traditional projection-based methods relative to the plug-in-based methods.

Our maintained assumption for the next two sections is that there is no problem with the identification of θ. This serves two purposes. First, it provides a simple setup for describing these methods. Second, while the projection-based methods are practically useful in the presence of identification problems, their low power (due to conservativeness) relative to the plug-in-based methods is difficult to justify when there is no such problems. Therefore, this setup also allows us to emphasize the benefit of using our new method of projection when we introduce it in the next section, because we show that in the absence of identification problems our new method is asymptotically equivalent to the plug-in-based methods of inference.

Under the standard assumptions (to be stated clearly in Section 3) that – (i) $\sqrt{n}\bar{g}_n(\theta_0) \xrightarrow{d} N(0, V_{gg})$ and (ii) $\widehat{V}_{gg}(\theta) \xrightarrow{P} V_{gg}(\theta)$ (both are continuous and positive definite, and $V_{gg}(\theta_0) = V_{gg}$) – the Dufour and Taamouti (2005b)-type projection-based test rejects $H^0 : \theta_1 = \theta_1^0$ if

$$\inf_{\theta_2^0 \in \Theta_2} S_n(\theta_1^0, \theta_2^0) > \chi_k^2(1 - \epsilon), \tag{2.1}$$

$$\text{where } S_n(\theta) := n\bar{g}_n'(\theta)\widehat{V}_{gg}^{-1}(\theta)\bar{g}_n(\theta). \tag{2.2}$$

In this paper we will consider $\epsilon \in (0, 1)$ as the allowable rate of Type-I error pre-specified by the user. The test in (2.1) is the projection-based test using the S-statistic proposed by Stock and Wright (2000) and defined in (2.2). In the context of linear instrumental variables models with serially uncorrelated and conditionally homoskedastic errors, this is the projection-based Anderson-Rubin test [see Staiger and Stock (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005b, 2007)]. The asymptotic size of the test cannot exceed ϵ. This test, however, has low power in over-identified models (i.e., with $k > \nu$) if some, for e.g., $k - \nu$, moment restrictions are not (very) informative of the unknown parameters θ.

One way to reduce the conservativeness is to choose the ν restrictions that are most informative:

$$E\left[G'V_{gg}^{-1}g_t(\theta)\right] = 0 \Longleftrightarrow \theta = \theta_0 \tag{2.3}$$

where $G := G(\theta_0)$, $G(\theta) := \lim_{n\to\infty} E[\partial \bar{g}_n(\theta)/\partial\theta']$ (assuming differentiability and existence). Then, under the assumptions (i), (ii) and, additionally, (iii) $\widehat{G}_n(\theta) \xrightarrow{P} G(\theta)$ (both full column rank at θ_0 and continuous) the Dufour and Taamouti (2005b)-type projection-based test rejects $H^0 : \theta_1 = \theta_1^0$ if

$$\inf_{\theta_2^0\in\Theta_2} \mathfrak{LM}_n(\theta_1^0, \theta_2^0) > \chi^2_\nu(1-\epsilon) \tag{2.4}$$

$$\text{where } \mathfrak{LM}_n(\theta) := n\bar{g}_n'(\theta)\widehat{V}_{gg}^{-1/2}(\theta)P(\widehat{V}_{gg}^{-1/2'}(\theta)\widehat{G}_n(\theta))\widehat{V}_{gg}^{-1/2'}(\theta)\bar{g}_n(\theta). \tag{2.5}$$

Here, $\mathfrak{LM}_n(\theta)$ is the S-statistic based on the efficient choice of ν linear combinations of k (estimated) moment restrictions given in (2.3). In other words, (2.5) gives a general form of the score statistic that, depending on the hitherto unspecified estimators $\widehat{G}_n(\theta)$ and $\widehat{V}_{gg}(\theta)$, includes the GMM score statistic of Newey and West (1987), the Continuous Updating GMM (CU-GMM) score statistic (i.e., the K statistic of Kleibergen (2005)), the Generalized Empirical Likelihood (GEL) score statistic of Guggenberger and Smith (2005) and other versions of score statistics (based on the GEL-implied probabilities and the naive empirical probabilities) described in Chaudhuri and Renault (2011). We call the test based on (2.4) - (2.5) the usual projection-based score test. This test is still conservative because the critical value comes from the χ^2_ν distribution whereas the number of parameters to be tested is ν_1. To see this, suppose there exists $\widehat{\theta}_{n2}(\theta_1) \in \Theta_2$ (indexed by θ_1) such that

$$\widehat{G}_{n2}'(\theta_1, \widehat{\theta}_{n2}(\theta_1))\widehat{V}_{gg}^{-1}(\theta_1, \widehat{\theta}_{n2}(\theta_1))\bar{g}_n(\theta_1, \widehat{\theta}_{n2}(\theta_1)) = 0 \tag{2.6}$$

$$\text{and } \sqrt{n}\left(\widehat{\theta}_{n2}(\theta_{01}) - \theta_{02}\right) = \mathcal{O}_p(1), \tag{2.7}$$

where $\widehat{G}_{n2}(\theta)$ denotes the last ν_2 columns of $\widehat{G}_n(\theta)$. If $\theta_1^0 = \theta_{01}$, it follows under standard assumptions that

$$\inf_{\theta_2\in\Theta_2} \mathfrak{LM}_n(\theta_{01}, \theta_2) \leq \mathfrak{LM}_n(\theta_{01}, \widehat{\theta}_{n2}(\theta_{01})) \xrightarrow{d} \chi^2_{\nu_1}, \tag{2.8}$$

and, therefore, the critical value $\chi^2_\nu(1-\epsilon)$ for the usual projection-based score test is conservative.

A strategy to avoid this problem could be to define an alternative projection-based score test that rejects $H^0 : \theta_1 = \theta_1^0$ if

$$\inf_{\theta_2^0 \in \Theta_2} \mathfrak{LM}_{n1}(\theta_1^0, \theta_2^0) > \chi^2_{\nu_1}(1-\epsilon) \tag{2.9}$$

$$\text{where } \mathfrak{LM}_{n1}(\theta) := n\bar{g}_n'(\theta)\widehat{V}_{gg}^{-1/2}(\theta)P(\widehat{V}_{gg}^{-1/2'}(\theta)\widehat{G}_{n1}(\theta))\widehat{V}_{gg}^{-1/2'}(\theta)\bar{g}_n(\theta), \tag{2.10}$$

and $\widehat{G}_{n1}(\theta)$ denotes the first ν_1 columns of $\widehat{G}_n(\theta)$. Here, $\mathfrak{LM}_{n1}(\theta)$ is the S-statistic based on the first ν_1 of the efficient choice of ν linear combinations of k (estimated) moment restrictions given in (2.3). When θ_{02} is known, these ν_1 moment restrictions are the most informative about the $\nu_1 \times 1$ vector θ_{01} and are sufficient to uniquely identify it. The alternative projection-based test in (2.9), however, is still conservative. To see this, note that for any θ,

$$\mathfrak{LM}_n(\theta) = \mathfrak{LM}_{n1}(\theta) + n\bar{g}_n'(\theta)\widehat{V}_{gg}^{-1/2}(\theta)P\left(N\left(\widehat{V}_{gg}^{-1/2'}(\theta)\widehat{G}_{n1}(\theta)\right)\widehat{V}_{gg}^{-1/2'}(\theta)\widehat{G}_{n2}(\theta)\right)\widehat{V}_{gg}^{-1/2'}(\theta)\widehat{g}_n(\theta).$$

The right hand side is the sum of two (almost surely) non-negative variables. When $\theta_1^0 = \theta_{01}$, without the (asymptotic) column rank deficiency of $N\left(\widehat{V}_{gg}^{-1/2'}(\theta)\widehat{G}_{n1}(\theta)\right)\widehat{V}_{gg}^{-1/2'}(\theta)\widehat{G}_{n2}(\theta)$ at $\theta = (\theta_{01}', \widehat{\theta}_{n2}'(\theta_{01}))'$, it follows under standard assumptions that

$$\inf_{\theta_2 \in \Theta_2} \mathfrak{LM}_{n1}(\theta_{01}, \theta_2) \leq \mathfrak{LM}_{n1}(\theta_{01}, \widehat{\theta}_{n2}(\theta_{01})) < \mathfrak{LM}_n(\theta_{01}, \widehat{\theta}_{n2}(\theta_{01})) \xrightarrow{d} \chi^2_{\nu_1}, \tag{2.11}$$

and, therefore, the critical value $\chi^2_{\nu_1}(1-\epsilon)$ is conservative. Our simulations show that the alternative projection-based score test can even be less powerful than the usual projection-based score test.

Neither the usual projection-based score test nor the alternative projection-based score test achieves asymptotic equivalence with the plug-in-based score test that rejects $H^0 : \theta_1 = \theta_1^0$ if

$$\mathfrak{LM}_n(\theta_1^0, \widehat{\theta}_{n2}(\theta_1^0)) > \chi^2_{\nu_1}(1-\epsilon) \tag{2.12}$$

where $\widehat{\theta}_{n2}(\theta_1^0)$ satisfies (2.6) for $\theta_1 = \theta_1^0$. This is a serious drawback of the projection-based methods because the plug-in-based score test, under standard regularity conditions (i.e., when it works), is known to have certain local optimality properties.

3 The new projection-based method of inference

In this section we describe how to overcome the aforementioned drawbacks of the usual and alternative projection-based score tests by using a $C(\alpha)$ form of the score statistic accompanied by a restricted but feasible projection.[1] We will define an infeasible score test for $H^0 : \theta_1 = \theta_1^0$ that uses the unknown true value θ_{02} of the nuisance parameters θ_2 as plug-in and show that both the plug-in-based score test defined in (2.12) and the new projection-based score test to be defined in (3.4) are asymptotically equivalent to this infeasible score test. However, first let us list a set of standard assumptions for the results discussed so far and the new ones.

Assumption Θ: [*assumptions on* θ_0 *and the parameter space*]

The $k \geq \nu$ moment restrictions in (1.1) are satisfied for $\theta_0 \in \text{interior}(\Theta)$ where Θ is a ν-dimensional compact subset of $\mathbb{R}^\nu$. The partition $\theta_0 = (\theta_{01}', \theta_{02}')'$ is such that $\theta_{0i} \in \text{interior}(\Theta_i)$ where Θ_i is a ν_i-dimensional compact subset of $\mathbb{R}^{\nu_i}$ for $i = 1, 2$ and $\Theta = \Theta_1 \times \Theta_2$.

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

The following hold for $\theta \in \mathcal{N} \in \Theta$ where $\mathcal{N}$ is a non-shrinking open neighborhood of θ_0:

D1. $\sqrt{n}\bar{g}_n(\theta_0) \xrightarrow{d} \Psi_g \sim N(0, V_{gg})$, and $\sqrt{n}(\theta - \theta_0) \neq o_p(1) \Rightarrow \lim_{n\to\infty} E[\sqrt{n}\bar{g}_n(\theta)] \neq 0$.

D2. $\widehat{V}_{gg}(\theta) \xrightarrow{P} V_{gg}(\theta)$ uniformly, where $V_{gg}(\theta)$ is continuous at θ_0 and $V_{gg}(\theta_0) = V_{gg}$ is positive definite. $\widehat{G}_n(\theta) \xrightarrow{P} G(\theta) := \lim_{n\to\infty} E[\partial\bar{g}_n(\theta)/\partial\theta']$ uniformly, $G(\theta)$ is continuous at θ_0 and $G := G(\theta_0)$ is full column rank.

D3. $\widehat{G}_n(\theta)'\widehat{V}_{gg}^{-1}(\theta)\widehat{G}_n(\theta)$ is positive definite. (Assumed for convenience.)

Remark 1: Assumption Θ and the first part of assumption $D1$ rule out problems of asymptotic size distortion due to failure of the moment restrictions (for e.g., endogeneity or near exogeneity of instruments). The second part of the assumption D rules out problems due to weak identification. Both these are explicitly incorporated in assumption W in the next section.

[1]See Bera and Bilias (2001) for a survey of the use of Neyman (1959)'s $C(\alpha)$ statistic in econometrics.

The infeasible score test is defined as the test that rejects $H^0 : \theta_1 = \theta_1^0$ if

$$\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0, \theta_{02})) > \chi^2_{\nu_1}(1-\epsilon), \text{ where} \tag{3.1}$$

$$\mathfrak{LM}_{n1}^{\text{eff}}(\theta) := n\bar{g}_n'(\theta)\widehat{V}_{gg}^{-1/2}(\theta)P\left(N\left(\widehat{V}_{gg}^{-1/2'}(\theta)\widehat{G}_{n2}(\theta)\right)\widehat{V}_{gg}^{-1/2'}(\theta)\widehat{G}_{n1}(\theta)\right)\widehat{V}_{gg}^{-1/2'}(\theta)\widehat{g}_n(\theta). \tag{3.2}$$

The test in (3.1) is infeasible because it uses the unknown true value of the nuisance parameters θ_2. Now a word on the superscript "eff" that is used to mean efficient. Note that the estimator $\widehat{\theta}_1^{\text{eff}}$ obtained by solving for θ_1 from

$$G_1'V_{gg}^{-1/2}N\left(V_{gg}^{-1/2'}G_2\right)\bar{g}_n(\theta_1, \theta_{02}) = 0 \tag{3.3}$$

has asymptotic variance $\left(G_1'V_{gg}^{-1/2}N\left(V_{gg}^{-1/2'}G_2\right)V_{gg}^{-1/2'}G_1\right)^{-1}$ where G_1 and G_2 are, respectively, the first ν_1 and the remaining ν_2 columns of G. Under standard assumptions, this asymptotic variance attains the semi-parametric efficiency bound for estimators of θ_1 in moment conditions models like (1.1) when θ_{02} is unknown. Therefore, in some sense, the left hand side of (3.3) is an efficient score function for θ_1. The statistic $\mathfrak{LM}_{n1}^{\text{eff}}(\theta)$ is a quadratic form of an estimator of this efficient score with respect to an estimator of the inverse of its asymptotic variance (or the asymptotic variance of $\widehat{\theta}_1^{\text{eff}}$). In other words, $\mathfrak{LM}_{n1}^{\text{eff}}(\theta)$ is the S-statistic based on ν_1 restrictions obtained by taking the ortho-complement of the projection of the first ν_1 rows of an estimator of the efficient moment restrictions in (2.3) on its last ν_2 rows. $\mathfrak{LM}_{n1}^{\text{eff}}(\theta)$ can also be interpreted as Neyman (1959)'s $C(\alpha)$ statistic [see Bera and Bilias (2001)]. The local optimality properties of the plug-in score test in (2.12), when they hold, are due to its asymptotic equivalence with the infeasible score test in (3.1) when θ_1^0 is $\sqrt{n}$-local to θ_{01}.

Now we define the new projection method using the statistic $\mathfrak{LM}_{n1}^{\text{eff}}(\theta)$. The new projection-based score test is defined as a test that rejects $H^0 : \theta_1 = \theta_1^0$

$$\left\{\text{if } \mathcal{C}_{2n}(1-\zeta, \theta_1^0) = \text{ empty, or if } \inf_{\theta_2^0 \in \mathcal{C}_{2n}(1-\zeta,\theta_1^0)} \mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0, \theta_2^0) > \chi^2_{\nu_1}(1-\tau)\right\}, \tag{3.4}$$

where $\mathcal{C}_{2n}(1-\zeta, \theta_1^0) \subset \Theta_2$ is any asymptotic $(1-\zeta)\times 100\%$ confidence region for θ_2, and $\zeta, \tau \in (0,1)$ are specified by the user. This is a two-step procedure: In the first step one constructs any asymptotic confidence region $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$ for the nuisance parameters θ_2 with or without imposing the null hypothesis of interest. In the second step, one rejects $H^0:\theta_1=\theta_1^0$ if either $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$ is empty or if the infimum of $\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0,\theta_2^0)$ for all $\theta_2^0\in\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$ is larger than $\chi^2_{\nu_1}(1-\tau)$.[2]

The following lemma is key to explaining the local asymptotic properties of the new projection-based score test for testing $H^0:\theta_1=\theta_1^0$. These properties are described in Theorem 3.2.

Lemma 3.1 *Let $\theta_1^0=\theta_{01}+d_1/\sqrt{n}\in\Theta_1$ for some fixed d_1 and consider any $\theta_2^0=\theta_{02}+d_{n2}/\sqrt{n}\in\Theta_2$ for some $d_{n2}=\mathcal{O}_p(1)$. Then, under assumptions Θ and D, $\mathfrak{LM}_{n1}^{eff}(\theta_1^0,\theta_2^0)=\mathfrak{LM}_{n1}^{eff}(\theta_1^0,\theta_{02})+o_p(1)$ and converges in distribution to a non-central $\chi^2_{\nu_1}$ with non-centrality parameter $d_1'G_1'V_{gg}^{-1/2}N(V_{gg}^{-1/2'}G_2)V_{gg}^{-1/2'}G_1d_1$.*

Remark 2:

(a) Lemma 3.1 states that as long as the unknown nuisance parameters θ_2 are replaced by any $\sqrt{n}$-consistent estimator, the statistic $\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0,\theta_2)$ is asymptotically equivalent to its "infeasible form" that replaces θ_2 by its unknown true value θ_{02}. In other words, small deviations of $\mathcal{O}_p(1/\sqrt{n})$ from the unknown true value θ_{02} of the nuisance parameters θ_2 do not affect the asymptotic behavior of $\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0,\theta_2)$. This property does not hold for the statistics $\mathfrak{LM}_n(\theta_1^0,\theta_2)$ and $\mathfrak{LM}_{n1}(\theta_1^0,\theta_2)$.

(b) Additionally when (2.6) holds for $\theta_1=\theta_1^0$, in which case $\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0,\widehat{\theta}_{n2}(\theta_1^0))=\mathfrak{LM}_n(\theta_1^0,\widehat{\theta}_{n2}(\theta_1^0))$, and when (2.7) holds, Lemma 3.1 justifies the use of the plug-in-based score test defined in (2.12).

Theorem 3.2 *Let assumptions Θ and D hold. Then the following results hold for any given $\zeta,\tau\in(0,1)$ as $n\to\infty$:*

(i) If the asymptotic coverage of $\mathcal{C}_{2n}(1-\zeta,\theta_{01})$ is $1-\zeta$ uniformly in $\theta_{02}\in\Theta_2$ such that $\theta_0:=(\theta_{01}',\theta_{02}')'$ satisfies assumptions Θ and D, then the asymptotic size of the new projection-based score test defined in (3.4) cannot exceed $\min(\zeta+\tau,1)$.

(ii) If $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$ is non-empty almost surely and if $\sup_{\theta_2^0\in\mathcal{C}_{2n}(1-\zeta,\theta_1^0)}\sqrt{n}\|\theta_2^0-\theta_{02}\|=\mathcal{O}_p(1)$, then the new projection-based score test defined in (3.4) is asymptotically equivalent to the

[2] If $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)\cap\text{boundary}(\Theta_2)\neq\varnothing$, then one can take $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)=\mathcal{C}_{2n}(1-\zeta,\theta_1^0)\cap\text{interior}(\Theta_2)$ to avoid the occurrence of $\inf_{\theta_2^0\in\mathcal{C}_{2n}(1-\zeta,\theta_1^0)}\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0,\theta_2^0)$ on the boundary of Θ_2. This should not create any problem asymptotically because $\theta_{02}\in\text{interior}(\Theta_2)$ by assumption Θ.

infeasible score test defined in (3.1) *for any hypothesized value* $\theta_1^0 \in \Theta_1$ *such that* $\theta_1^0 = \theta_{01} + d_1/\sqrt{n}$ *(where* d_1 *is fixed).*

Remark 3:

(a) Part (i) of the theorem allows for empty confidence regions $\mathcal{C}_{2n}(1-\zeta, \theta_{01})$, for the sake of the discussion of weak identification in the next section, and states that the size of the new projection test is always bounded from above by $\min(\zeta+\tau, 1)$. For the previously stated allowable rate of Type-I error ϵ, the user can specify ζ and τ such that $\zeta + \tau = \epsilon$. Our experience suggests that the upper bound is not sharp and is mainly affected by τ whenever $\mathcal{C}_{2n}(1-\zeta, \theta_{01})$ is non-empty.

(b) Part (ii) of the Theorem is remarkable. As long as $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$ is non-empty and belongs in the $\sqrt{n}$-neighborhood of θ_{02} (as will be the case when θ_{02} is identified), the choice of ζ does not matter asymptotically and one can safely choose $\tau = \epsilon$, i.e., the allowable rate of Type-I error.[3] In this case, the new projection-based score test is asymptotically equivalent to the infeasible score test (defined in (3.1)) that uses the unknown true value θ_{02} of the nuisance parameters θ_2. The plug-in-based test defined in (2.12), when it works, is also asymptotically equivalent to the infeasible score test and thus equivalent to the new projection-based test. We note that the new projection-based test may have a computational advantage over the plug-in-based test whenever it is difficult to find the restricted estimator $\widehat{\theta}_{n2}(\theta_1^0)$ of the nuisance parameters (to be plugged-in) using certain methods like Empirical Likelihood (EL).[4]

(c) The new test can also be seen as a Bonferroni-type test. However, it has an important difference with the usual Bonferroni-type tests [see Moon and Schorfheide (2009) and the references therein]. Unlike the latter tests where standard Bonferroni arguments give an upper bound $\zeta + \tau$ for the asymptotic size, the asymptotic size for the new projection-based score test is τ if the conditions in (ii) are satisfied by $\mathcal{C}_{2n}(1-\zeta, \theta_{01})$. This is achieved by the use of the $C(\alpha)$ form

[3] While we simply assume the aforementioned properties of $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$, extending assumption D to $\theta \in \mathcal{N}_1 \times \Theta_2$ where $\mathcal{N}_1 \in \Theta_1$ is a non-shrinking open neighborhood of θ_{01} is typically sufficient for these to hold.

[4] It might be tempting for the user to plug-in any easy to obtain $\sqrt{n}$-consistent estimator (for e.g., two-step GMM) of the nuisance parameter θ_2 in an EL score test and avoid the difficulty of solving a saddle-point problem while hoping to exploit the desirable higher order properties of EL established by Newey and Smith (2004) (although in a different context). Lemma 3.1 assures that this should not affect the first order asymptotic properties of the score test. However, in simulations by Chaudhuri and Renault (2011) this often caused some upward size distortion in (very small) finite samples. Their simulations also show that projection from a (easy to obtain) confidence region of θ_2, as done in Theorem 3.2(ii), mitigates this issue without adversely affecting the finite-sample power much and still reduces the computational burden of solving the saddle-point problem of EL by restricting the search of the nuisance parameters to a confidence region instead of the entire parameter space. This also applies to the other computationally difficult GEL score tests.

of the score statistic.

4 Models with weakly identified parameters

In this section we allow some or all elements of θ_{01} and θ_{02} to be weakly identified.[5] The generality of the results and also the exposition in the last section do not carry through completely. The problems come from two distinct sources which we briefly discuss below.

4.1 Problems and the recent developments related to weak identification

First, if some elements of θ_{01} (or θ_{02}) are weakly identified, the score statistic of Newey and West (1987) is usually not asymptotically pivotal (under H^0) and can lead to severe upward size distortion [see Kleibergen (2005)]. The problem is due to the estimator of the Jacobian G. In this case it is important to use an estimator of G that is independent of the average moment vector $\bar{g}_n(\theta)$ (both appropriately scaled to avoid degeneracy), at least asymptotically. Kleibergen (2005) showed that the Jacobian estimator in the CU-GMM score statistic satisfies this property. Guggenberger and Smith (2005) extended this result by establishing the first order equivalence of the Jacobian estimator for the entire GEL class that also includes CU-GMM [see also Newey and Smith (2004)].

Second, if some elements of θ_{02} (i.e., the nuisance parameters) are weakly identified, Stock and Wright (2000) showed that it is no longer possible to estimate these unknown (and unspecified by H^0) elements ($\sqrt{n}$-)consistently. As a result, the entire GEL class of plug-in score statistics are also no longer asymptotically pivotal (under H^0). Projection-based methods have been traditionally found to be theoretically and practically useful in this case because they enable the user to impose any pre-specified allowable rate of Type-I error (ϵ) as the upper bound to the asymptotic size of the test.

In a major development to the weak identification literature, Kleibergen and Mavroeidis (2009) [henceforth, KM09] established an interesting result where they showed that under certain conditions the CU-GMM plug-in score statistic (i.e., the plug-in or subset-K statistic) is boundedly

[5] In this paper, identification of $\theta, \theta_1, \theta_2$ and $\theta_0, \theta_{01}, \theta_{02}$ respectively are used interchangeably to refer to the same thing. In particular, we say that there is identification problem with $\theta/\theta_1/\theta_2$ if some elements of $\theta_0/\theta_{01}/\theta_{02}$ are weakly identified. We hope that this is not unduly confusing.

pivotal asymptotically – the asymptotic distribution function under the H^0 is bounded from the right by that of a central $\chi^2_{\nu_1}$ where ν_1 is the dimension of θ_1. Hence the asymptotic size of the plug-in CU-GMM score test as defined in (2.12) cannot exceed the allowable rate of Type-I error ϵ. KM09 further argued that a plug-in-based method should be preferred over the corresponding projection-based method because while the use of the standard fixed χ^2 critical values does not lead to over-rejection by either, the former has better power properties. Therefore, in addition to showing that our new projection-based method outperforms the usual and alternative projection methods, it will also be worthwhile to explore how effective the use of the restricted projection and the $C(\alpha)$ statistic is in reducing the difference in power from the plug-in methods (when the latter work).

While our new method and all the projection methods are applicable to the GEL class of score statistics of Guggenberger and Smith (2005) we focus our attention to CU-GMM because the results of KM09 have, as of now, only been proved for it. Use of the CU-GMM score necessitates specifying the form of the estimator of the expected Jacobian, i.e., $\widehat{G}_n(\theta)$. From Kleibergen (2005) we know that this is given by

$$\widehat{G}_n(\theta) := \bar{G}_n(\theta) - \left[\widehat{V}_{1g}(\theta)\widehat{V}_{gg}^{-1}(\theta)\bar{g}_n(\theta), \widehat{V}_{2g}(\theta)\widehat{V}_{gg}^{-1}(\theta)\bar{g}_n(\theta), \ldots, \widehat{V}_{\nu g}(\theta)\widehat{V}_{gg}^{-1}(\theta)\bar{g}_n(\theta)\right], \tag{4.1}$$

where $\bar{G}_n(\theta) := \partial\bar{g}_n(\theta)/\partial\theta'$. For $l = 1, \ldots, \nu$ the $k \times k$ matrices $\widehat{V}_{lg}(\theta)$ are obtained as a byproduct of differentiating with respect to θ_l the CU-GMM objective function $S_n(\theta)$ defined in (2.2).

Since tests involving the CU-GMM score statistic suffer from an undesired decline in power at irrelevant parameter values, Kleibergen (2005) proposed a hybrid test, called the subset-JKLM test, that rejects $H^0 : \theta_1 = \theta_1^0$

$$\left\{\text{if } S_n(\theta_1^0, \widehat{\theta}_{n2}(\theta_1^0)) - \mathfrak{LM}_n(\theta_1^0, \widehat{\theta}_{n2}(\theta_1^0)) > \chi^2_{k-\nu}(1-\zeta), \text{ or if } \mathfrak{LM}_n(\theta_1^0, \widehat{\theta}_{n2}(\theta_1^0)) > \chi^2_{\nu_1}(1-\tau)\right\}, \tag{4.2}$$

where $S_n(\theta)$ and $\mathfrak{LM}_n(\theta)$ are as defined in (2.2) and (2.5) respectively. Kleibergen (2005) provided simulation evidence of the undesired decline in power of the K test and its removal by the JKLM test. KM09 showed that the asymptotic size of the subset-JKLM test cannot exceed $\zeta + \tau$ and recommended choosing ζ, τ such that $\zeta + \tau = \epsilon$, i.e., the allowable rate of Type-I error.

4.2 The new projection-based score test

The new projection-based score test in the presence of weakly identified parameters is the same as that in (3.4) where $\widehat{G}_n(\theta)$ is as defined in (4.1) and

$$\mathcal{C}_{2n}(1-\zeta,\theta_1^0) := \left\{\theta_2^0 \in \Theta_2 : S_n(\theta_1^0,\theta_2^0) \le \chi_k^2(1-\zeta)\right\}. \tag{4.3}$$

The choice of $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$, which can be empty in practice, helps to eliminate the undesired decline in power at certain irrelevant parameter values [see Chaudhuri (2008) for details]. This is done in the same spirit of the subset-JKLM test. Simulations at the end of this section show that the performance of the new test (with $\zeta+\tau=\epsilon$) is better than that of the usual and alternative projection-based tests and is not necessarily dominated by that of the plug-in-based tests like the subset-K test and the subset-JKLM test (with $\zeta+\tau=\epsilon$).

Our results are based on assumptions that are relatively mild as compared to those in KM09. For e.g., the existence of a Central Limit Theorem is assumed only at the truth (see assumption D$'$2 below), and we do not need to make the difficult to verify assumption that the plug-in estimator $\widehat{\theta}_{n2}(\theta_{01}) := \arg\min_{\theta_2^0\in\Theta_2} S_n(\theta_{01},\theta_2^0)$ is the only $\theta_2 \in \Theta_2$ such that (2.6) (for $\theta_1 = \theta_{01}$) holds. We also note that the method proposed here is more general than that in Chaudhuri et al. (2010) which only applies to split-sample linear Instrumental Variables regressions with weak instruments.

Let us now state explicitly the assumptions before we describe the properties of the new projection-based score test.

First we define the identification properties of the parameters. While the characterization of identification is a straightforward extension of the framework of Stock and Wright (2000), to the best of our knowledge this is the first paper that allows for both weakly and strongly identified parameters of interest and nuisance parameters. In what follows we use the following notation to distinguish weakly and strongly identified parameters. For $j = w, s$, let $\nu_j = \nu_{1j} + \nu_{2j}$, $\theta_j = (\theta'_{1j}, \theta'_{2j})'$ and $\Theta_j = \Theta_{1j} \times \Theta_{2j}$. This notation regroups the weakly identified parameters as θ_w and the (strongly) identified parameters as θ_s (as defined in assumption W). The true values are, when convenient, regrouped as $\theta_{0w} = (\theta'_{01w}, \theta'_{02w})'$ and $\theta_{0s} = (\theta'_{01s}, \theta'_{02s})'$ respectively. When necessary, $\mathcal{N} \subset \Theta$ and $\mathcal{N}_r \subset \Theta_r$ are generically used to denote non-shrinking open neighborhoods of θ_0 and θ_{0r} for $r = 1w, 1s, 2w, 2s, w, s, 1, 2$ respectively. Define $\widetilde{\mathcal{N}} := \mathcal{N}_w \times \mathcal{N}_{1s} \times \Theta_{2s}$.

Assumption Θ: [continued]

For $l = 1, 2$, let $\Theta_l = \Theta_{lw} \times \Theta_{ls}$ and for $j = w, s$, let $\theta_{0lj} \in \text{interior}(\Theta_{lj})$ where $\Theta_{lj} \subset \mathbb{R}^{\nu_{lj}}$ is compact.

Assumption W: [*characterization of weak identification*]

$E[\bar{g}_n(\theta)] = \widetilde{m}_n(\theta)/\sqrt{n} + m(\theta_s)$ where

(a) $\widetilde{m}_n(\theta) : \Theta \mapsto \mathbb{R}^k$ is such that $\widetilde{m}_n(\theta) \to \widetilde{m}(\theta)$ uniformly for $\theta \in \widetilde{\mathcal{N}}$ where $\widetilde{m}(\theta)$ is bounded and continuous and $\widetilde{m}(\theta_0) = 0$. For $\theta \in \widetilde{\mathcal{N}}$, $\widetilde{M}_n(\theta) := \partial \widetilde{m}_n(\theta)/\partial\theta'$, $\widetilde{M}_n(\theta) \to \widetilde{M}(\theta)$ uniformly. $\widetilde{M}(\theta) = [\widetilde{M}_{1w}(\theta), \widetilde{M}_{1s}(\theta), \widetilde{M}_{2w}(\theta), \widetilde{M}_{2s}(\theta)]$ where, for $l = 1, 2$ and $j = w, s$, the $k \times \nu_{lj}$ matrix $\widetilde{M}_{lj}(\theta)$ is bounded and continuous.

(b) $m(\theta_s) : \Theta_s \mapsto \mathbb{R}^k$ is a continuous function and $m(\theta_s) = 0$ if and only if $\theta_s = \theta_s^0$. For $\theta_s \in \mathcal{N}_{1s} \times \Theta_{2s}$, $M(\theta_s) := \partial m(\theta_s)/\partial\theta_s'$ is bounded and continuous. $M(\theta_{0s})$ has full column rank. Here, $M(\theta_s) = [M_1(\theta_s), M_2(\theta_s)]$ where $M_l(\theta_s) := \partial m(\theta_s)/\partial\theta_{ls}'$ for $l = 1, 2$.

To establish the desirable asymptotic properties of the new projection-based score test in the presence of weakly identified parameters, assumption D from the last section needs to be augmented with further assumptions following Guggenberger and Smith (2005) and Kleibergen (2005). These assumptions are listed under assumption D' which, hereafter, replaces assumption D.

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

D′1. $\bar{G}_n(\theta) := \partial \bar{g}_n(\theta)/\partial\theta' = [G_{1wn}(\theta), G_{1sn}(\theta), G_{2wn}(\theta), G_{2sn}(\theta)] = E[\bar{G}_n(\theta)] + o_p(1)$ uniformly for $\theta \in \widetilde{\mathcal{N}}$ where $E[\bar{G}_n(\theta)] = \partial E[\bar{g}_n(\theta)]/\partial\theta' = \widetilde{M}_n(\theta)/\sqrt{n} + [0, M_1(\theta_s), 0, M_2(\theta_s)]$ by imposing interchangeability of the order of differentiation and integration (and from assumption W).

D′2. $\sqrt{n}\left[\bar{g}_n'(\theta_0), vec'(\bar{G}_{wn}(\theta_0) - E[\bar{G}_{wn}(\theta_0)])\right] \xrightarrow{d} [\Psi_g', \Psi_w']$ where [6]

$$\begin{bmatrix} \Psi_g \\ \Psi_w \end{bmatrix} \sim \mathcal{N}\left(0, V(\theta_0) := \begin{bmatrix} \underset{k\times k}{V_{gg}(\theta_0)} & \underset{k\times k\nu_w}{V_{gw}(\theta_0)} \\ \underset{k\nu_w\times k}{V_{wg}(\theta_0)} & \underset{k\nu_w\times k\nu_w}{V_{ww}(\theta_0)} \end{bmatrix}\right).$$

[6] The partition of $\Psi_w(\theta) = [\Psi_{1w}'(\theta), \Psi_{2w}'(\theta)]'$, $V_{gw}(\theta) = [V_{g1}(\theta), V_{g2}(\theta)] = V_{wg}'(\theta)$ and $V_{ww}(\theta) = (V_{ll'}(\theta))_{l,l'=1,2}$ follows the partition of $\theta_w = (\theta_{1w}', \theta_{2w}')'$, i.e., the partition of the weakly identified elements of θ into those from θ_1 and θ_2 respectively.

$V_{gg}(\theta)$ is bounded, continuous and positive definite, and $\widehat{V}_{gg}(\theta) \xrightarrow{P} V_{gg}(\theta)$ uniformly for $\theta \in \widetilde{\mathcal{N}}$. $V_{wg}(\theta)$ is bounded and continuous. $\widehat{V}_{wg}(\theta) := [\widehat{V}'_{1g}(\theta), \ldots, \widehat{V}'_{\nu_{1w},g}(\theta), \widehat{V}'_{\nu_1+1,g}(\theta), \ldots, \widehat{V}'_{\nu_1+\nu_{2w},g}(\theta)]' \xrightarrow{P} V_{wg}(\theta)$ uniformly for $\theta \in \mathcal{N}$.[7] For $l = \nu_{1w}+1, \ldots, \nu_1, \nu_1+\nu_{2w}+1, \ldots, \nu$ the $k \times k$ matrices $\widehat{V}_{lg}(\theta)$ are such that $\widehat{V}_{lg}(\theta)\widehat{V}_{gg}^{-1}(\theta) = \mathcal{O}_p(1)$ uniformly for $\theta \in \mathcal{N}$. [See (4.1) for the notations.]

Remark 4:

(a) Assumption $D'1$ implies that $G(\theta) := \lim_{n\to\infty} E[\bar{G}_n(\theta)] = [G_{1w}(\theta), G_{1s}(\theta), G_{2w}(\theta), G_{2s}(\theta)] = [0, M_1(\theta_s), 0, M_2(\theta_s)]$. Hence $G(\theta_0) = [G_{1w}(\theta_0), G_{1s}(\theta_0), G_{2w}(\theta_0), G_{2s}(\theta_0)] = [0, M_1(\theta_{0s}), 0, M_2(\theta_{0s})]$ has rank ν_s under assumption W(b) and satisfies the local identification condition for $\theta_s = (\theta'_{1s}, \theta'_{2s})'$.

(b) Some results are assumed to hold in $\widetilde{\mathcal{N}}$ instead of $\mathcal{N}$ to ensure that for $\theta_1^0 = \theta_{01} + d_1/\sqrt{n}$ (for some fixed d_1), the region $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$ defined in (4.3) belongs in the $\sqrt{n}$-neighborhood of θ_{02} almost surely whenever $\nu_{2w} = 0$, i.e., whenever there are no weakly identified nuisance parameters.

The following results assume Θ, W and D'. In Lemma 4.1 we discuss the asymptotic properties of the $C(\alpha)$ form of the score statistic $\mathfrak{LM}^{\text{eff}}_{n1}(\theta)$ defined in (3.2) and using $\widehat{G}_n(\theta)$ from (4.1). In Theorem 4.2 we discuss the properties of the new projection-based score test defined in (3.4) and using $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$ from (4.3).

Lemma 4.1 *Let assumptions Θ, W and D' hold. Then the following results hold for $\mathfrak{LM}^{\text{eff}}_{n1}(\theta_1, \theta_2)$ defined in* (3.2) *using $\widehat{G}_n(\theta)$ from* (4.1)*, as $n \to \infty$:*

(i) $\mathfrak{LM}^{\text{eff}}_{n1}(\theta_{01}, \theta_{02}) \xrightarrow{d} \chi^2_{\nu_1}$.

(ii) Let $\nu_w = 0$, $\theta_1^0 = \theta_{01} + d_1/\sqrt{n} \in \Theta_1$ for some fixed d_1 and consider any $\theta_2^0 = \theta_{02} + d_{n2}/\sqrt{n} \in \Theta_2$ for some $d_{n2} = \mathcal{O}_p(1)$. Then, $\mathfrak{LM}^{\text{eff}}_{n1}(\theta_1^0, \theta_2^0) = \mathfrak{LM}^{\text{eff}}_{n1}(\theta_1^0, \theta_{02}) + o_p(1)$ and converges in distribution to a non-central $\chi^2_{\nu_1}$ with non-centrality parameter

$$d_1' M_1'(\theta_{0s}) V_{gg}^{-1/2}(\theta_0) N\left(V_{gg}^{-1/2'}(\theta_0) M_2(\theta_{0s})\right) V_{gg}^{-1/2'}(\theta_0) M_1(\theta_{0s}) d_1.$$

[7] For convenience of reading the messy notation let us point out that the $k \times k$ matrix $\widehat{V}_{lg}(\theta_0) \xrightarrow{P} V_{lg}(\theta_0) = \text{Asym.Cov}(n^{-1/2}\partial\bar{g}_n(\theta_0)/\partial\theta_l, n^{-1/2}\bar{g}_n(\theta_0))$ where θ_l is the l-th element $(l = 1, \ldots, \nu_{1w}, \nu_1+1, \ldots, \nu_1+\nu_{2w})$ of the vector θ.

Theorem 4.2 *Let assumptions Θ, W and D' hold. Then the following properties of the new projection-based score test defined in* (3.4)*, using* $\widehat{G}_n(\theta)$ *from* (4.1) *and* $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$ *from* (4.3)*, hold for any given* $\zeta, \tau \in (0,1)$ *as* $n \to \infty$*:*

(i) The asymptotic size of the new projection-based score test cannot exceed $\min(\zeta + \tau, 1)$*.*

(ii) If $\nu_w = 0$*, then the new projection-based score test is asymptotically more powerful than the infeasible score test defined in* (3.1) *using* $\widehat{G}_n(\theta)$ *from* (4.1) *and* $\epsilon = \tau$ *and for any hypothesized value* $\theta_1^0 \in \Theta_1$ *such that* $\theta_1^0 = \theta_{01} + d_1/\sqrt{n}$ *(where* d_1 *is fixed).*

Remark 5:

(a) *Asymptotic size:* Theorem 4.2(i) gives an upper bound to the asymptotic size of the new projection-based score test and extends Theorem 3.2(i) to cases where some or all elements of the parameters of interest (θ_1) and/or nuisance parameters (θ_2) can be weakly identified. In the context of the plug-in-based tests such upper bounds were recently provided by KM09. Under their assumptions, plugging in the restricted CU-GMM estimator of θ_2 (which, additionally, needs to be the only $\theta_2 \in \Theta_2$ satisfying (2.6) for $\theta_1 = \theta_{01}$) results in an upper bound ϵ for the asymptotic size of the plug-in test in (2.12).[8] Choosing ζ, τ such that $\zeta + \tau = \epsilon$ matches both the upper bounds.

(b) *Analytical comparison of power:* We emphasize that under the generality of our assumption W the feasible plug-in-based tests described in KM09 and our new test are only boundedly asymptotically pivotal. While the plug-in-based tests (when they work) are analytically shown by KM09 to be more powerful than the traditional projection-based tests, their argument does not hold when the plug-in tests are compared with the new projection test. Our only analytical result related to power is given in Theorem 4.2(ii) and is limited in scope [also see Remark 5(c)].[9] Although under the full generality of assumption W we were unable to provide a comprehensive analytical comparison of the plug-in and the new projection methods, our Monte-Carlo simulations

[8]Discussion of the plug-in-based tests are qualified by the phrase "when they work" because, following KM09, additional assumptions beyond assumptions Θ, W, D are required for the tests to be boundedly asymptotically pivotal and hence for the upper bound to be valid.

[9]As in Remark 3(b), we note here that the plug-in-based score test defined in (2.12) with $\epsilon = \tau$, when it works, is also asymptotically equivalent to the infeasible score test and hence is covered by this result. However, as will be seen in the simulations under the conditions of Theorem 4.2(ii), even with $\epsilon = \zeta + \tau$ for the plug-in-based test, there is no observable difference between its rejection frequencies with that of the new projection-based score test when ζ is small relative to τ and the conditions of Theorem 4.2(ii) are satisfied.

show that when the upper bound ϵ for the size of the plug-in-based tests is matched with that of the new test, no test dominates uniformly in terms of finite-sample power.

(c) *Undesired decline in power and the choice of 1st step confidence region:* Theorem 4.2(ii) is weaker than Theorem 3.2(ii) because even under the null hypothesis the particular choice of $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$ can lead to empty confidence regions in the first step of the new projection-based score test. While allowing $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$ to be empty has a positive effect on power and, in particular, some undesired decline in power due to the use of the CU-GMM score is eliminated, it also makes the size of the test dependent on ζ and breaks the asymptotic equivalence with a size τ infeasible score test.[10] On the other hand, with this particular choice of $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$ computation is easier (empty $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$, see e.g. Table 3, eliminates the need for the second step test) and at the same time the new test is not dominated in terms of power by either the subset-K or the subset-JKLM test.

(d) *Choice of ζ and τ:* Similar to the subset-JKLM test defined in (4.2), the performance of the new projection-based score test now depends on the choice of ζ. There are many possible choices of ζ and τ such that $\zeta + \tau = \epsilon$, the allowable rate of Type-I error. While Kleibergen (2005) recommended choosing τ close to ϵ (for e.g., $\zeta = 4.5\%, \tau = .5\%$ when $\epsilon = 5\%$), our simulations indicate that other choices can sometimes yield better power.

(e) *Restricted projection and other score statistics:* The restricted projection from the first step confidence region instead of the entire nuisance parameter space Θ_2, when applied to the usual projection-based score test (defined in (2.4)) or the alternative projection-based score test (defined in (2.9)) does not reduce their conservativeness much. As a result, while the upper bound from Theorem 4.2(i) continues to hold in such cases, the power property described in Theorem 4.2(ii) does not. Intuitively, whenever $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$ defined in (4.3) is non-empty, it contains the CU-GMM estimator $\widehat{\theta}_{n2}(\theta_1^0) := \arg\min_{\theta_2 \in \Theta_2} S_n(\theta_1^0, \theta_2)$. Therefore, results in (2.8) and (2.11) regarding the conservativeness of these two tests still apply and any increase in the

[10] To avoid this, other confidence regions can be used; e.g., those obtained by imposing H^0 and then subsequently inverting the K test or the modified quasi-likelihood ratio test or even the Wald test for θ_2 treating $\theta_1 = \theta_1^0$ as given. While the former two regions are known to have correct coverage probabilities (under H^0), they are also computationally extremely demanding [see Mikusheva (2010)]. The Wald confidence region does not have correct coverage probability when $\nu_{2w} > 0$ i.e., when there are weakly identified nuisance parameters, but this problem can be avoided by using KM09's results. We do not, however, recommend the use of these confidence regions because the resulting projection-based test is always less powerful than the subset-K test with $\epsilon = \tau$, and does not have any theoretical or practical advantage over the subset-K test. KM09's criticism of projection methods apply here.

rejection frequency due to the restricted projection happens only if the first step confidence region is empty. On the other hand, the restricted projection is useful for the new test only because it uses the $C(\alpha)$ form of the score statistic.

4.3 Monte-Carlo Study

In this subsection we perform a Monte-Carlo experiment to compare the finite-sample rejection rates of various projection-based score tests and also the subset-K and subset-JKLM tests of Kleibergen (2005). We illustrate the observations in Remark 5 with the help of these simulations. Following Dufour and Taamouti (2005a), we draw $w_t = (y_t, X_{1t}, X_{2t}, Z_t')'$ for $t = 1, \ldots, n(= 100)$ such that

$$\left\{ \begin{array}{c} y_t = X_{1t}\theta_{01} + X_{2t}\theta_{02} + u_t, \\ X_{1t} = Z_t'\Pi_1 + U_{1t}, \\ X_{2t} = Z_t'\Pi_2 + U_{2t} \end{array} \right\} \text{ where } (u_t, U_{1t}, U_{2t}) \overset{\text{i.i.d}}{\sim} \mathcal{N}\left(0, \Sigma = \begin{bmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.3 \\ 0.8 & 0.3 & 1 \end{bmatrix}\right).$$

The individual instruments in Z_t are generated as i.i.d. $\mathcal{N}(0, 1)$ variables but are kept fixed over simulations. We report the results for $k = 2$, 4 and 8 instruments. The matrix $\Pi = [\Pi_1, \Pi_2]$ is constructed such that $\Pi = \mathbb{C}/\sqrt{n}$ where $\mathbb{C} = [\mathbb{C}_1, \mathbb{C}_2]$ and the elements of $\mathbb{C}_j$ are set at 1.1547 and 20 for $j = 1, 2$ to represent "weak identification" and "strong identification" of θ_1 and θ_2 respectively.[11] From the general setup of weak identification described in assumption W, we consider the following four special cases listed in Table 1 for the simulations.[12]

The true values of θ_1 and θ_2 are set at $\theta_{01} = 0.5$ and $\theta_{02} = 1$. We vary the hypothesized value θ_1^0 and report the empirical rejection rates of various nominal 5% tests when $\theta_1^0 = \theta_{01}$ in Table 2 to show finite-sample size. We report the same for a grid of values around $\theta_1^0 - \theta_{01} = 0$ in Figures 1-3 to show finite-sample power. Results are based on 5000 simulations.

We consider a variety of feasible tests that share the following two characteristics – (i) all tests use fixed critical values that are quantiles of χ^2 distributions, and (ii) the only theoretical result known to be valid under WI-Cases I-IV is that the asymptotic size of these tests is bounded from above by the same number, 5%. These tests are: (1) the CU-GMM score/subset-K test defined

[11] For all the cases considered, the minimum eigen value of the concentration matrix varies from 0.4762 to 1.3329 whenever θ_1 and/or θ_2 is weakly identified.

[12] Additional simulation results based on related specifications are available upon request.

in (2.12) with $\epsilon = 5\%$; (2) the usual projection-based score test defined in (2.4) with $\epsilon = 5\%$, (3) the new projection-based score test defined in (3.4), and (4) the subset-JKLM test defined in (4.2). For the last two tests we consider a series of choices of ζ and τ such that $\zeta + \tau = 5\%$. For brevity, we only report a small subset of the results for these tests.[13]

From Table 2 it is clear that the finite-sample size of these tests is, in general, less than 5% with the exception of the plug-in-based tests – the subset-K and subset-JKLM tests – which tend to slightly over-reject the truth in over-identified models. In terms of finite-sample power, the new projection-based score test is vastly superior to the usual (and the alternative) projection-based score test. In fact, its finite-sample power is often very similar to that of the plug-in-based tests. While unlike the simulations in Kleibergen (2005), ours do not show the dramatic undesired decline in finite-sample power of the subset-K test, there seems to be a small dip in WI-Case II (in Figures 2-3 to the right of the point $\theta_1^0 - \theta_{01} = 0$) that is, as expected, corrected by the subset-JKLM test and the new projection-based score test. Our simulations show that the new projection-based score test can also be more powerful than the plug-in-based tests.[14] This is illustrated for WI-Case III in Figure 3 (and less clearly in Figure 2) for alternatives to the left of the point $\theta_1^0 - \theta_{01} = 0$. This remarkable result counters KM09's claim that the projection-based tests cannot be more powerful than the plug-in-based tests. Their claim is of course correct for the traditional projection-based score tests.

It may appear that the restricted projection is the main driving force behind the good power performance of the new projection-based score test and hence could be applied to the usual projection-based score test and the alternative projection-based score test to make their performance comparable to the new test. This intuition, however, is not correct. While, as mentioned in Remark 5(e), such restricted projection indeed improves the power performance of the usual and the alternative projection-based score tests, the improvement is not sufficient to make them comparable to the new test. To see this, first note that the restricted projection applied to the

[13]Simulation results for the alternative projection-based score test defined in (2.9) with $\epsilon = 5\%$ is not reported because this test seems to have unusually low finite-sample power. Simulation results for the projection-based S test defined in (2.1) with $\epsilon = 5\%$ are not reported because Chaudhuri (2008) already documented the relatively poor finite-sample power properties for this test.

[14]A similar point was made, albeit less convincingly, in Zivot and Chaudhuri (2009).

usual and alternative methods results in tests that reject $H^0 : \theta_1 = \theta_1^0$ if

$$\text{Restricted Usual: } \mathcal{C}_{2n}(1-\zeta, \theta_1^0) = \text{empty, or } \inf_{\theta_2^0 \in \mathcal{C}_{2n}(1-\zeta,\theta_1^0)} \mathfrak{LM}_n(\theta_1^0, \theta_2^0) > \chi^2_\nu(1-\tau), \tag{4.4}$$

$$\text{Restricted Alternative: } \mathcal{C}_{2n}(1-\zeta, \theta_1^0) = \text{empty, or } \inf_{\theta_2^0 \in \mathcal{C}_{2n}(1-\zeta,\theta_1^0)} \mathfrak{LM}_{n1}(\theta_1^0, \theta_2^0) > \chi^2_{\nu_1}(1-\tau). \tag{4.5}$$

Now note that whenever $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$ is non-empty it will, by construction, contain $\widehat{\theta}_{n2}(\theta_1^0)$, i.e., the CU-GMM estimator of θ_2 restricted by H^0. Therefore, the tests in (4.4) and (4.5) are less powerful than their hybrid (of projection and plug-in) versions that reject $H^0 : \theta_1 = \theta_1^0$ if

$$\text{Hybrid Usual: } \mathcal{C}_{2n}(1-\zeta, \theta_1^0) = \text{empty, or } \mathfrak{LM}_n(\theta_1^0, \widehat{\theta}_{n2}(\theta_1^0)) > \chi^2_\nu(1-\tau), \tag{4.6}$$

$$\text{Hybrid Alternative: } \mathcal{C}_{2n}(1-\zeta, \theta_1^0) = \text{empty, or } \mathfrak{LM}_{n1}(\theta_1^0, \widehat{\theta}_{n2}(\theta_1^0)) > \chi^2_{\nu_1}(1-\tau). \tag{4.7}$$

In Figure 4 we plot the finite-sample power of these two tests along with that of the new projection-based score test (all with $\zeta = .5\%$ and $\tau = 4.5\%$). The new projection-based score test is still more powerful than the hybrid tests and hence the restricted usual and restricted alternative projection-based score tests. As mentioned before in Remark 5(e), although the contribution of the empty first step confidence region (listed in Table 3) to the finite-sample power is same for all three tests in Figure 4, it is the extra power due to the use of the $C(\alpha)$ form that leads to the superior performance of the new projection-based score test.

5 Conclusion

Projection-based methods of inference on subsets of parameters have been traditionally considered useful for obtaining tests that do not over-reject the true parameter values when either it is difficult to estimate the nuisance parameters or their identification status is questionable. However, these tests are also often criticized for being overly conservative. In this paper we tried to address the problem of conservativeness by introducing a new method of projection-based inference for subsets of parameters. The new projection-based test substantially outperforms the traditional projection-based tests in terms of power. The new test relies on a $C(\alpha)$ form of the score statistic for the parameters of interest and a restricted projection for the nuisance parameters. In the context of moment conditions models without any identification problem, this new projection-based test is even asymptotically equivalent to an infeasible test that uses the unknown true value of the nuisance parameters as plug-in. This also leads to the asymptotic equivalence with the feasible plug-in-based tests (when they work) that plug in an estimator of the nuisance parameters obtained by, say, solving a set of first conditions after imposing the null hypothesis.

The result is remarkable because, to our knowledge, the existing projection-based methods in the literature do not possess this desirable asymptotic property. The result also has practical relevance especially when it is difficult or impossible to obtain a point estimator of the nuisance parameters to be plugged-in to implement the feasible plug-in-based tests [see Chaudhuri and Renault (2011)]. Kim (2009) makes a promising advance in this direction by applying the idea of restricted projection for inference in models with partially identified parameters.

In the context of inference using CU-GMM in models with weakly identified parameters, the new method of projection also conclusively serves the main purpose of our paper, that is, it reduces the conservativeness of the traditional projection-based methods of inference substantially. Interestingly, the simulations also indicate that with our preferred choice of restricted projection, the finite-sample power of the new projection-based test can be more than that of the plug-in-based methods (with matched upper bounds for asymptotic size) against certain alternatives.

A Appendix

Proof of Lemma 3.1: First note that for n large enough, $\theta^0 := (\theta_1^0, \theta_2^0) \in \mathcal{N}$ and hence applying assumption D2 after a mean-value expansion of $\sqrt{n}\bar{g}_n(\theta^0)$ around $\sqrt{n}\bar{g}_n(\theta_0)$, it follows that

$$\begin{aligned}
\mathfrak{LM}_{n1}^{\text{eff}}(\theta^0) &= \left(\sqrt{n}\bar{g}_n(\theta_0) + G_1 d_1 + G_2 d_{n2}\right)' V_{gg}^{-1/2} P\left(N\left(V_{gg}^{-1/2'} G_2\right) V_{gg}^{-1/2'} G_1\right) \\
&\quad \times V_{gg}^{-1/2'}\left(\sqrt{n}\bar{g}_n(\theta_0) + G_1 d_1 + G_2 d_{n2}\right) + o_p(1) \\
&= \left(\sqrt{n}\bar{g}_n(\theta_0) + G_1 d_1\right)' V_{gg}^{-1/2} P\left(N\left(V_{gg}^{-1/2'} G_2\right) V_{gg}^{-1/2'} G_1\right) V_{gg}^{-1/2'}\left(\sqrt{n}\bar{g}_n(\theta_0) + G_1 d_1\right) + o_p(1) \\
&= \mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0, \theta_{02}) + o_p(1).
\end{aligned}$$

The second equality follows by noting that $N\left(V_{gg}^{-1/2'} G_2\right) V_{gg}^{-1/2'} G_2 = 0$. Convergence in distribution to the non-central $\chi^2_{\nu_1}$ distribution follows from assumption D by noting that $V_{gg}^{-1/2'}\sqrt{n}\bar{g}_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, I_k)$ and that rank of $P\left(N\left(V_{gg}^{-1/2'} G_2\right) V_{gg}^{-1/2'} G_1\right)$ is ν_1. ■

Remark 6: In the rest of the proofs when we discuss the size of the tests for $H^0 : \theta_1 = \theta_1^0$ or the asymptotic coverage of the confidence regions for θ_{02} under $H^0 : \theta_1 = \theta_1^0$, it is worthwhile to note that the results hold for any true value θ_{02} that (along with θ_{01}) characterize the model by satisfying assumptions Θ and D for results in Section 3, and Θ, W and D' for results in Section 4. In this regard the result of Lemma 3.1 is important because it shows by considering the sequence of the nuisance parameters $\theta_2^0 = \theta_{02} + d_{n2}/\sqrt{n}$, there is no discontinuity in the asymptotic distribution of the statistic $\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0, \theta_2)$ at θ_{02}. For all other statistics considered in this paper, the discontinuity due to such a sequence of nuisance parameters manifest in the non-centrality parameter of the asymptotic distribution. A formal characterization of the model and the appropriate sequence of nuisance parameters (required for the asymptotic approximation of the exact size) is possible by extending, for e.g., equations (2.5) and (3.16) of Guggenberger (2010) and imposing $h_{11} = 0$. However, this is not done here explicitly because our main results do not establish the asymptotic size of the sub-vector tests; rather they establish an upper bound to it by using the asymptotic size of the full-vector weak-identification-robust tests that are already well documented in Kleibergen (2005).[15]

[15] Noting that $\mathfrak{LM}_n(\theta) = \mathfrak{LM}_{n1}^{\text{eff}}(\theta) + \mathfrak{LM}_{n2}(\theta)$ where the last statistic is the same as (2.10) with the subscripts 1 replaced by 2, the remark is also applicable to the infeasible test by appealing to Kleibergen's results.

Proof of Theorem 3.2: (i) First note that given the correct asymptotic coverage, θ_{02} is contained in $\mathcal{C}_{2n}(1-\zeta,\theta_{01})$ with probability approaching $1-\zeta$. Now, conditional on $\theta_{02}\in\mathcal{C}_{2n}(1-\zeta,\theta_{01})$, Lemma 3.1 implies that

$$\inf_{\theta_2^0\in\mathcal{C}_{2n}(1-\zeta,\theta_{01})}\mathfrak{LM}_{n1}^{\text{eff}}(\theta_{01},\theta_2^0)\leq\mathfrak{LM}_{n1}^{\text{eff}}(\theta_{01},\theta_{02})\leq\chi^2_{\nu_1}(1-\tau)$$

with probability approaching $1-\tau$ (for the last inequality). Therefore, from the definition of the new projection-based score test in (3.4), it follows by Bonferroni arguments that the (unconditional) asymptotic size of the test cannot exceed $1-(1-\zeta)(1-\tau)\leq\zeta+\tau$.
(ii) $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$ is assumed to be nonempty almost surely and hence $\inf_{\theta_2^0\in\mathcal{C}_{2n}(1-\zeta,\theta_1^0)}\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0,\theta_2^0)$ exists except for a negligible set that should not affect the rest of the arguments required for the result. Now since $\sup_{\theta_2^0\in\mathcal{C}_{2n}(1-\zeta,\theta_1^0)}\sqrt{n}\|\theta_2^0-\theta_{02}\|=\mathcal{O}_p(1)$, the n-th element of the sequence

$$\{\theta_{n2}^{\text{inf}}(\theta_1^0)\in\mathcal{C}_{2n}(1-\zeta,\theta_1^0),\text{ where the infimum }\inf_{\theta_2^0\in\mathcal{C}_{2n}(1-\zeta,\theta_1^0)}\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0,\theta_2^0)\text{ is attained }\},$$

can also be expressed as $\theta_{n2}^{\text{inf}}(\theta_1^0)=\theta_{02}+d_{n2}^{\text{inf}}(\theta_1^0)/\sqrt{n}$ for some $d_{n2}^{\text{inf}}(\theta_1^0)=O_p(1)$, and, therefore, for n large enough $(\theta_1^{0'},\theta_{n2}^{\text{inf}'}(\theta_1^0))\in\mathcal{N}$. Hence, the result follows directly from Lemma 3.1 which states that $\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0,\theta_2)=\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0,\theta_{02})+o_p(1)$ for any θ_2 that is $\sqrt{n}$-local to θ_{02}. ■

Proof of Lemma 4.1: (i) For $l=1,2$, define Λ_l as a diagonal matrix of order ν_l such that the first ν_{lw} diagonal elements are $\sqrt{n}$ and the rest of the diagonal elements are 1. Note that for any θ, the statistic $\mathfrak{LM}_{n1}^{\text{eff}}(\theta)$ is numerically invariant to post multiplication of $\widehat{G}_{n1}(\theta)$ by Λ_1 and of $\widehat{G}_{n2}(\theta)$ by Λ_2 respectively.

From assumptions W and $D'1$ we know that $\bar{g}_n(\theta_0)=\mathcal{O}_p(1/\sqrt{n})$. Hence for $l=1,2$ and $q=(l-1)\nu_1+\nu_{lw}$, further application of assumption D' gives

$$\begin{aligned}\widehat{G}_{lsn}(\theta_0)&=\bar{G}_{lsn}(\theta_0)-\left[\widehat{V}_{q+1,g}(\theta_0)\widehat{V}_{gg}^{-1}(\theta_0)\bar{g}_n(\theta_0),\ldots,\widehat{V}_{q+\nu_{ls},g}(\theta_0)\widehat{V}_{gg}^{-1}(\theta_0)\bar{g}_n(\theta_0)\right]\\&=[M_l(\theta_{0s})+o_p(1)]-[\mathcal{O}_p(1)\times\mathcal{O}_p(1/\sqrt{n})]\xrightarrow{P}M_l(\theta_{0s}).\end{aligned}\qquad\text{(A.1)}$$

On the other hand, from assumptions W(a) and D' we know that for $l=1,2$, the r-th column of

$\widehat{G}_{lwn}(\theta_0)$, denoted by $\widehat{G}^{(r)}_{lsn}(\theta_0)$, is such that for $q = (l-1)\nu_1 + r$,

$$\begin{aligned}
\sqrt{n}\widehat{G}^{(r)}_{lwn}(\theta_0) &= \sqrt{n}\bar{G}^{(r)}_{lwn}(\theta_0) - \widehat{V}_{qg}(\theta_0)\widehat{V}^{-1}_{gg}(\theta_0)\sqrt{n}\bar{g}_n(\theta_0), \\
&= \left[\sqrt{n}\left(\bar{G}^{(r)}_{lwn}(\theta_0) - E[\bar{G}^{(r)}_{lwn}(\theta_0)]\right) - \widehat{V}_{qg}(\theta_0)\widehat{V}^{-1}_{gg}(\theta_0)\sqrt{n}\bar{g}_n(\theta_0)\right] + \sqrt{n}E[\bar{G}^{(r)}_{lwn}(\theta_0)], \\
&\xrightarrow{d} [\Psi^{[r]}_l - V^{[r]}_{lg}(\theta_0)V^{-1}_{gg}(\theta_0)\Psi_g] + \widetilde{M}^{(r)}_{lw}(\theta_0) = G^{*[r]}_l \text{ (say)} \qquad \text{(A.2)}
\end{aligned}$$

where $\Psi^{[r]}_l$ denotes the $k \times 1$ block containing the $(r-1)k+1$-th to rk-th rows of Ψ_l; $V^{[r]}_{lg}(\theta_0) = \operatorname{plim} \widehat{V}_{qg}(\theta_0)$ denotes the $k \times k$ block containing the $(r-1)k+1$-th to rk-th rows of $V_{lg}(\theta_0)$. Note that, by construction and from assumption $D'2$, $\Psi^{[r]}_l - V^{[r]}_{lg}(\theta_0)V^{-1}_{gg}(\theta_0)\Psi_g$ and Ψ_g are asymptotically jointly normally distributed, uncorrelated, and hence independent. Now, for $l = 1, 2$, denoting $G^*_l := [G^{*[1]}_l, G^{*[2]}_l, \ldots, G^{*[\nu_{lw}]}_l, M_l(\theta_{0s})]$, it follows from (A.1), (A.2) and assumption $D'1$ that

$$\mathfrak{LM}^{\text{eff}}_{n1}(\theta_{01}, \theta_{02}) \xrightarrow{d} \Psi'_g V^{-1/2}_{gg}(\theta_0) P\left(N\left(V^{-1/2'}_{gg}(\theta_0)G^*_2\right) V^{-1/2'}_{gg}(\theta_0)G^*_1\right) V^{-1/2'}_{gg}(\theta_0)\Psi_g \sim \chi^2_{\nu_1},$$

conditionally on G^*_1, G^*_2, and hence unconditionally because each element of these two matrices is independent of Ψ_g.

(ii) This follows from Lemma 3.1 once we note that for n large enough $\theta^0 \in \mathcal{N}$ and that the conditions of the Lemma are also satisfied here. In particular, now for $l = 1, 2$ and $q = (l-1)\nu_1 + \nu_{lw}$,

$$\begin{aligned}
\widehat{G}_{ln}(\theta^0) &= \bar{G}_{ln}(\theta^0) - \left[\widehat{V}_{q+1,g}(\theta^0)\widehat{V}^{-1}_{gg}(\theta^0)\bar{g}_n(\theta^0), \ldots, \widehat{V}_{q+\nu_{ls},g}(\theta^0)\widehat{V}^{-1}_{gg}(\theta^0)\bar{g}_n(\theta^0)\right] \\
&= \bar{G}_{ln}(\theta^0) - \left[\left(\frac{1}{\sqrt{n}}\widehat{V}_{q+1,g}(\theta^0)\widehat{V}^{-1}_{gg}(\theta^0)\right)\sqrt{n}\bar{g}_n(\theta^0), \ldots, \left(\frac{1}{\sqrt{n}}\widehat{V}_{q+\nu_{ls},g}(\theta^0)\widehat{V}^{-1}_{gg}(\theta^0)\right)\sqrt{n}\bar{g}_n(\theta^0)\right] \\
&= [M_l(\theta_0) + o_p(1)] - \left[\mathcal{O}_p(1/\sqrt{n}) \times \left\{\sqrt{n}\bar{g}_n(\theta_0) + M_1(\theta_{0s})d_1 + M_2(\theta_0)d_{n2}\right\}\right] \\
&\xrightarrow{P} M_l(\theta_0),
\end{aligned}$$

and the probability limit, which does not depend on the deviation from θ_0, is the same as G_l in the statement of Lemma 3.1. (The subscript s is irrelevant here because no parameter is weakly identified, i.e., $\nu_w = \nu_{1w} = \nu_{2w} = 0$.) ■

Proof of Theorem 4.2: (i) First note that assumption $D'2$ gives $S_n(\theta_0) \xrightarrow{d} \Psi_g' V_{gg}^{-1}(\theta_0)\Psi_g \sim \chi_k^2$ and this holds for any $\theta_0 \in \Theta$ satisfying assumptions Θ, W and D'. Therefore, $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$, defined in (4.3), contains θ_{02} with probability approaching $1-\zeta$ when H^0 is true (i.e., $\theta_1^0 = \theta_{01}$). Hence the proof follows in exactly the same way as that of Theorem 3.2(i) by using Lemma 4.1(i).

(ii) This proof has two parts - the first part establishes that under the conditions of the Theorem, the region $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$, if non-empty, can only contain points in the $\sqrt{n}$-neighborhood of θ_{02}. The second part uses this result and follows the same technique as in the proof of Theorem 3.2(ii) to establish our final result.

[Part 1:] The result will be proved based on the following two observations. First note that from assumptions W and $D'2$, we know that $S_n(\theta_1^0,\theta_2)/n - m'(\theta_{01},\theta_2)V_{gg}^{-1}(\theta_{01},\theta_2)m(\theta_{01},\theta_2) = o_p(1)$ uniformly for $\theta_2 \in \Theta_2^0 := \{\theta_2^0 \in \Theta_2 : \theta_2^0 = \theta_{02} + d_{n2} \text{ where } d_{n2} \neq o_p(1)\}$. Second, note that from assumptions W and D', we know that for any fixed $s \in (0,1/2)$, $n^{2s-1}S_n(\theta_1^0,\theta_2) - n^s(\theta_2-\theta_{02})'M_2'(\theta_0)V_{gg}^{-1}(\theta_0)M_2(\theta_0)n^s(\theta_2-\theta_{02}) = o_p(1)$ uniformly for $\theta_2 \in \Theta_2^s := \{\theta_2^0 \in \Theta_2 : \theta_2^0 = \theta_{02} + d_{n2}/n^s \text{ where } d_{n2} \neq o_p(1)\}$. Therefore, by assumptions $W(b)$ and $D'2$ i.e., by respectively using $m(\theta) = 0 \Longleftrightarrow \theta = \theta_0$, $V_{gg}(\theta)$ is positive definite for $\theta \in \widetilde{\mathcal{N}}$ and $M_2(\theta_0)$ is full column rank, it follows that the probability limits of $S_n(\theta_1^0,\theta_2)/n$ and $n^{2s-1}S_n(\theta_1^0,\theta_2)$ are greater than 0 uniformly for θ_2 in respectively Θ_2^0 and Θ_2^s (for $s \in (0,1/2)$). Therefore, the statistic $S_n(\theta_1^0,\theta_2^0)$, inverting which $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$ is obtained, diverges to $+\infty$ in both cases. Hence any $\theta_2^0 = \theta_{02} + d_{n2}/n^s$ (where $d_{n2} \neq o_p(1)$ and $s \in [0,1/2)$) has a zero probability of being contained in $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$ as $n \to \infty$.[16] This proves the result that $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$, if non-empty, can only contain points in the $\sqrt{n}$-neighborhood of θ_{02}.

[Part 2:] Define the following events:

$$
\begin{aligned}
A_n &:= \left\{\mathcal{C}_{2n}(1-\zeta,\theta_1^0) \text{ is empty}\right\} \\
B_n &:= \left\{\theta_2^0 \in \mathcal{C}_{2n}(1-\zeta,\theta_1^0) \Rightarrow \theta_2^0 = \theta_{02} + d_{2n}/\sqrt{n} \in \Theta_2 \text{ for some } d_{2n} = \mathcal{O}_p(1)\right\} \\
C_n &:= \left\{\inf_{\theta_2^0 \in \mathcal{C}_{2n}(1-\zeta,\theta_1^0)} \mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0,\theta_2^0) > \chi_{\nu_1}^2(1-\tau)\right\} \\
D_n &:= \left\{\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0,\theta_{02}) > \chi_{\nu_1}^2(1-\tau)\right\}.
\end{aligned}
$$

[16] There is no need to consider $\theta_2^0 = \theta_{02} + d_{n2}/n^s$ for $s < 0$, because for sufficiently large n this will be outside Θ_2 (compact) and hence, by definition, cannot be contained in $\mathcal{C}_{2n}(1-\zeta,\theta_1^0)$.

An implication of what has been just proved in [Part 1] is that B_n and A_n^c are asymptotically the same event. Hence conditioning on B_n instead of on A_n^c and vice versa does not change the asymptotic probability of any event. Now note that for any true value $\theta_0 \in \Theta$ satisfying assumptions Θ, W and D', the asymptotic rejection probability of the infeasible test of the hypothesis $H^0 : \theta_1 = \theta_1^0$ is

$$\begin{aligned}
\lim_n P_{\theta_0}(D_n) &= \lim_n P_{\theta_0}(D_n \cap A_n) + \lim_n P_{\theta_0}(D_n \cap A_n^c) \\
&\leq \lim_n P_{\theta_0}(A_n) + \lim_n P_{\theta_0}(D_n \cap A_n^c) \\
&= \lim_n P_{\theta_0}(A_n) + \lim_n P_{\theta_0}(A_n^c) P_{\theta_0}(D_n | A_n^c) \\
&= \lim_n P_{\theta_0}(A_n) + \lim_n P_{\theta_0}(A_n^c) P_{\theta_0}(D_n | B_n) \text{ (from [Part 1])} \\
&= \lim_n P_{\theta_0}(A_n) + \lim_n P_{\theta_0}(A_n^c) P_{\theta_0}(C_n | B_n) \text{ (from Lemma 4.1(ii))} \\
&= \lim_n P_{\theta_0}(A_n) + \lim_n P_{\theta_0}(A_n^c) P_{\theta_0}(C_n | A_n^c) \text{ (from [Part 1])} \\
&= \lim_n P_{\theta_0}(A_n) + \lim_n P_{\theta_0}(C_n \cap A_n^c),
\end{aligned}$$

which is the asymptotic rejection probability of the new projection-based score test. The proof of the third equality from the bottom follows that of Theorem 3.2(ii) by utilizing Lemma 4.1(ii) that establishes the invariance of the asymptotic distribution of $\mathfrak{LM}_{n1}^{\text{eff}}(\theta_1^0, \theta_2)$ to $\sqrt{n}$-local deviations from the true θ_{02}. ■

Acknowledgements

We are grateful to R. Gallant (co-editor), an anonymous associate editor, three anonymous referees, T. Richardson and J. Robins for their extremely helpful comments and important suggestions. We also thank E. Renault, F. Kleibergen, S. Mavroeidis, J. M. Dufour, R. Startz, M. Carrasco, D. Andrews, J. Hill, S. Park, participants at UBC, U. Montreal, NCSU, UNC Chapel Hill, Queen's, Waterloo, U. Washington, the ESRC Study Group Conference (2007), Midwest Econometrics Conference (2007) and the Winter (2008) and Summer(2009) meetings of the Econometric Society for helpful suggestions.

References

Bera, A. K., 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.

Chaudhuri, S., 2008. Projection-Type Score Tests for Subsets of Paramaters. Ph.D. thesis, University of Washington.

Chaudhuri, S., Renault, E., 2011. Finite-sample improvements of score tests by the use of implied probabilities from Generalized Empirical Likelihood. Tech. rep., University of North Carolina, Chapel Hill.

Chaudhuri, S., Richardson, T., Robins, J., Zivot, E., 2010. Split-Sample Score Tests in Linear Instrumental Variables Regression. Econometric Theory 26, 1820–1837.

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., 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., Taamouti, M., 2005a. Further Results on Projection-Based Inference in IV Regressions with Weak, Collinear or Missing Instruments, discussion Paper.

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

Dufour, J. M., 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., 2010. On the Asymptotic Size Distortion of Tests When Instruments Locally Violate the Exogeneity Assumption. Tech. rep., University of California, San Diego.

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

Kim, K. I., 2009. Inference for Subsets of Parameters in Partially Identified Models. Tech. rep., University of Minnesota.

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

Kleibergen, F., Mavroeidis, S., 2009. Inference on Subsets of Parameters in GMM without Assuming Identification. Tech. rep., Brown University Working Paper.

Mikusheva, A., 2010. Robust Confidence Sets in the Presence of Weak Instruments. Journal of Econometrics 157, 236–247.

Moon, H. R., Schorfheide, F., 2009. Estimation with Overidentifying Inequality Moment Conditions. Journal of Econometrics 153, 136–154.

Newey, W. K., Smith, R. J., 2004. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. Econometrica 72, 219–255.

Newey, W. K., West, K. D., 1987. Hypothesis Testing with Efficient Method of Moments Estimation. International Economic Review 28, 777–787.

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

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

Staiger, D., Stock, J. H., 1997. Instrumental Variables Regression with Weak Instruments. Econometrica 65, 557–586.

Stock, J. H., Wright, J. H., 2000. GMM with Weak Identification. Econometrica 68, 1055–1096.

Zivot, E., Chaudhuri, S., 2009. Comment: Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve by F. Kleibergen and S. Mavroeidis. Journal of Business and Economic Statistics 27, 328–331.

B Tables and Figures

	$\nu_{2w} = 1, \nu_{2s} = 0$	$\nu_{2w} = 0, \nu_{2s} = 1$
$\nu_{1w} = 1$, $\nu_{1s} = 0$	**WI-Case I** θ_1 : weakly identified θ_2 : weakly identified	**WI-Case II** θ_1 : weakly identified θ_2 : (strongly) identified
$\nu_{1w} = 0$, $\nu_{1s} = 1$	**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 special cases of weak identification for Monte-Carlo study.

		plug-in score/ subset-K	usual-proj score	new-proj score	sub-JKLM	new-proj score	sub-JKLM	new-proj score	sub-JKLM
k	WI-Case	$\epsilon = 5\%$	$\epsilon = 5\%$	$\zeta = .5\%$ $\tau = 4.5\%$	$\zeta = .5\%$ $\tau = 4.5\%$	$\zeta = 1\%$ $\tau = 4\%$	$\zeta = 1\%$ $\tau = 4\%$	$\zeta = 2\%$ $\tau = 3\%$	$\zeta = 2\%$ $\tau = 3\%$
2	I	2.6	0.4	2.4	-	2	-	1.4	-
2	II	5.2	1.5	4.7	-	4	-	3	-
2	III	2.6	0.4	2.4	-	2	-	1.4	-
2	IV	5.2	1.5	4.7	-	4	-	4	-
4	I	5.3	0.7	2.2	4.9	2.2	4.6	2	4.2
4	II	5.3	1.4	4.6	5.5	4.4	5.4	3.8	5.5
4	III	4.2	0.7	2.6	3.9	2.5	3.7	2.5	3.7
4	IV	5.1	1.4	4.6	5.2	4.3	5.1	3.7	5.3
8	I	7.9	1	2	7.6	2	7.3	2.4	6.6
8	II	6.2	1.9	5.2	6.3	5	6.2	4.7	6.3
8	III	6.4	1	2.1	6.3	2.1	6.1	2.6	5.7
8	IV	5	1.5	4.8	5.3	4.6	5.4	4.4	5.9

Table 2: Reported are the rates (in percentage) at which the tests reject the true value θ_{01} of the parameter of interest θ_1. Results are based on 5000 Monte-Carlo Trials with $k = 4, 8$ instruments/moments and sample size $n = 100$. The known upper bound for the asymptotic size of all these tests is 5% which is enforced either by choosing $\epsilon = 5\%$ directly for one-step tests like subset-K or usual-proj score or by choosing ζ and τ such that $\zeta + \tau = 5\%$ for the rest of the tests (which are all two step by nature). The frequencies are missing for sub-JKLM when $k = 2$ because the test is not defined in just-identified models.

	WI-Case I			WI-Case II				WI-Case III			WI-Case IV		
$\theta_1^0 - \theta_{01}$	$k=2$	$k=4$	$k=8$	$k=2$	$k=4$	$k=8$	$\theta_1^0 - \theta_{01}$	$k=2$	$k=4$	$k=8$	$k=2$	$k=4$	$k=8$
-5	0.1	0.3	0.6	2.62	5.2	5.2	-0.2	3.06	7.2	8.4	85.06	99.4	99.9
-4.5	0.1	0.2	0.5	2.4	5.0	5.0	-0.18	2.94	7.2	8.9	74.62	97.7	99.3
-4	0.1	0.2	0.5	2.24	4.6	4.7	-0.16	2.76	7.1	9.2	60.9	92.9	96.4
-3.5	0.1	0.2	0.5	2.1	4.1	4.3	-0.14	2.5	6.8	9.5	46.5	82.5	87.8
-3	0.1	0.2	0.5	1.9	3.7	3.9	-0.12	2.2	6.5	9.5	31.46	64.4	69.6
-2.5	0.08	0.2	0.4	1.68	3.3	3.5	-0.1	1.46	5.6	9.1	18.9	41.4	44.4
-2	0.08	0.2	0.3	1.52	2.7	2.8	-0.08	0.96	3.7	7.6	8.92	21.1	23.0
-1.5	0.08	0.1	0.2	1.24	1.8	2.1	-0.06	0.34	1.9	3.7	3.72	8.4	9.3
-1	0.06	0.0	0.2	0.78	1.1	1.3	-0.04	0.12	0.5	1.1	1.32	2.2	3.0
-0.5	0.04	0.1	0.1	0.52	0.5	0.8	-0.02	0.04	0.1	0.3	0.36	0.5	1.2
0	0.04	0.1	0.1	0.14	0.2	0.5	0	0.04	0.1	0.2	0.14	0.4	0.5
0.5	0.76	1.3	1.2	1.66	2.9	3.4	0.02	0.14	0.2	0.5	0.42	0.8	0.9
1	1.68	2.8	2.3	17.88	33.4	30.9	0.04	0.26	0.7	1.0	1.52	2.7	3.0
1.5	0.52	2.2	2.6	20.02	38.1	37.1	0.06	0.78	1.2	1.8	5.4	10.0	9.9
2	0.26	1.6	2.3	15.26	30.3	29.2	0.08	1.26	2.1	2.6	14.28	26.6	25.6
2.5	0.18	1.1	2.0	12.04	24.8	23.6	0.1	1.98	2.8	2.9	30.24	52.7	51.3
3	0.16	0.8	1.6	9.96	21.1	20.5	0.12	2.54	3.2	3.3	53.04	78.7	75.8
3.5	0.16	0.7	1.5	8.88	18.6	18.2	0.14	2.82	3.5	3.6	73.72	93.6	91.8
4	0.16	0.7	1.4	7.92	17.0	16.3	0.16	3.08	3.8	3.9	88.9	98.7	98.3
4.5	0.16	0.6	1.3	7.2	15.6	15.1	0.18	3.24	4.0	4.1	96.64	99.9	99.8
5	0.14	0.6	1.3	6.76	14.5	14.2	0.2	3.36	4.2	4.3	99.44	100.0	100.0

Table 3: Reported are the rates (in percentage) of obtaining an empty first step confidence region $\mathcal{C}_{2n}(1-\zeta, \theta_1^0)$ where $\zeta = .5\%$, θ_1^0 and θ_{01} are respectively the hypothesized and the true value of the parameter of interest θ_1. Results are based on 5000 Monte-Carlo Trials with $k = 2, 4, 8$ instruments/moments and sample size $n = 100$.

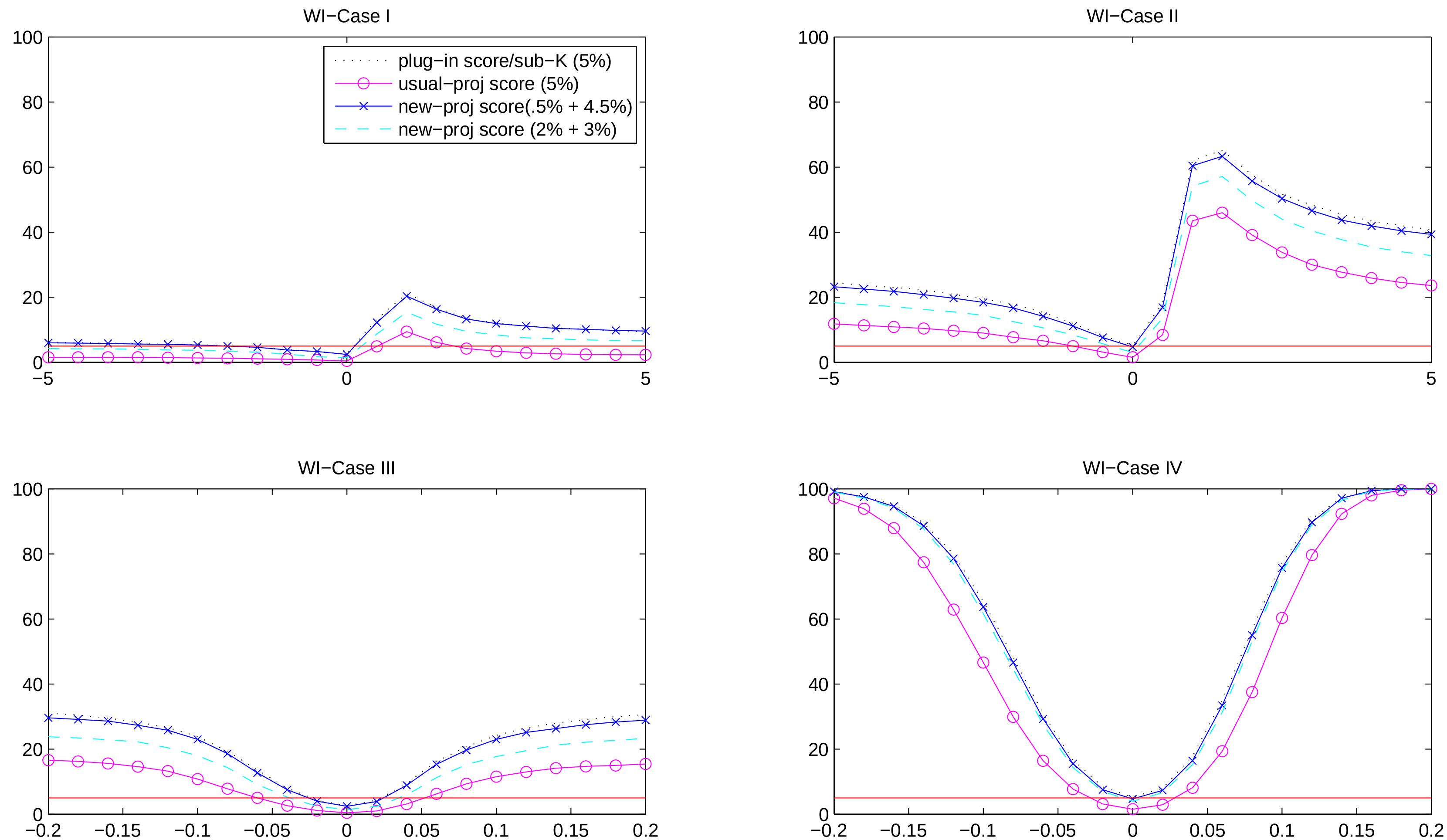

Figure 1: Plotted are the rejection rates of different tests against the horizontal axis = hypothesized θ_1 - true θ_1. Results are based on 5000 Monte-Carlo Trials with $k = 2$ instruments/moments and sample size $n = 100$. The known upper bound for the asymptotic size of all these tests is 5%. The JKLM test is not defined for just identified models.

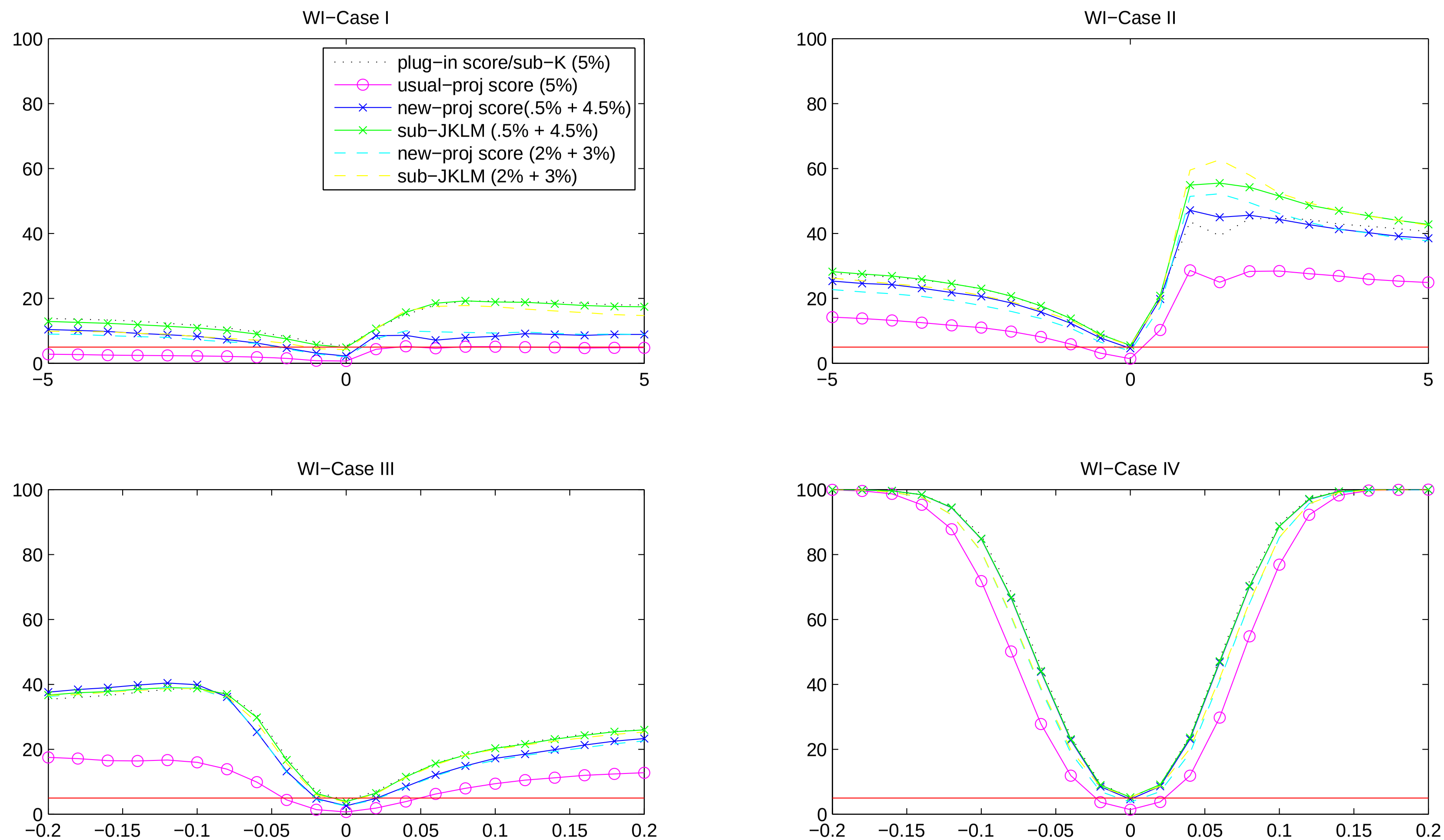

Figure 2: Plotted are the rejection rates of different tests against the horizontal axis = hypothesized θ_1 - true θ_1. Results are based on 5000 Monte-Carlo Trials with $k = 4$ instruments/moments and sample size $n = 100$. The known upper bound for the asymptotic size of all these tests is 5%.

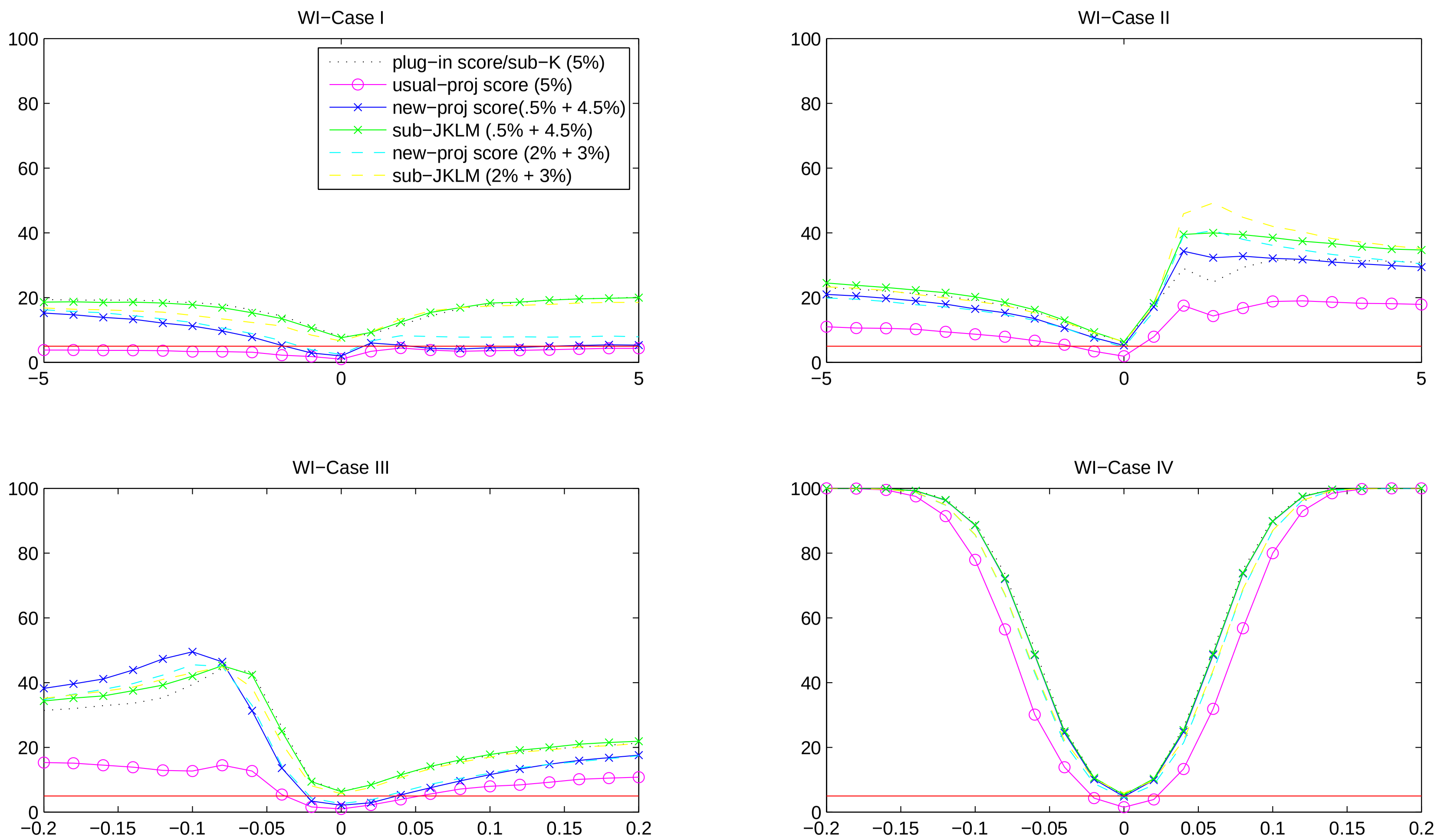

Figure 3: Plotted are the rejection rates of different tests against the horizontal axis = hypothesized θ_1 - true θ_1. Results are based on 5000 Monte-Carlo Trials with $k = 8$ instruments/moments and sample size $n = 100$. The known upper bound for the asymptotic size of all these tests is 5%.

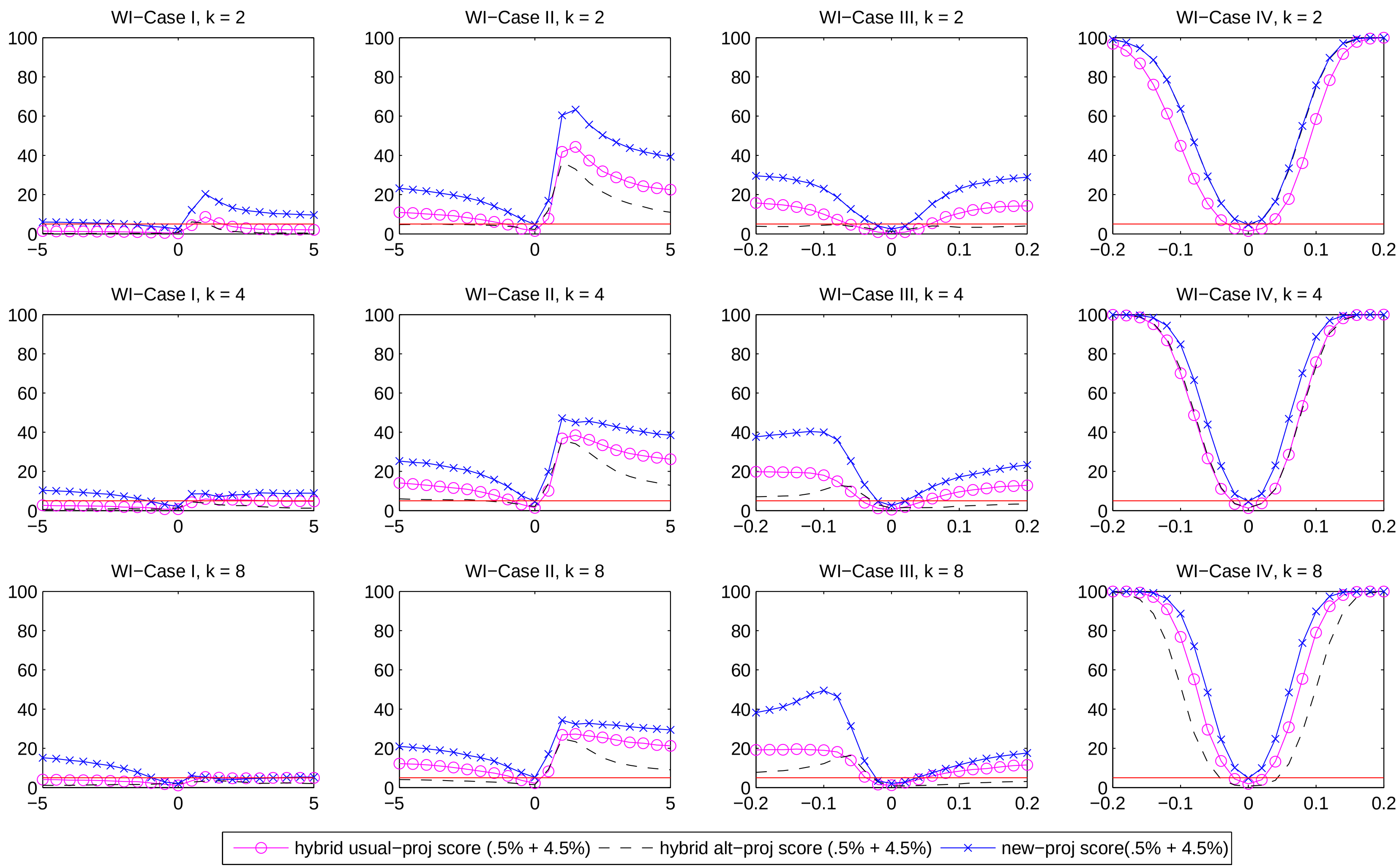

Figure 4: Plotted are the rejection rates of the hybrid tests described in (4.6) and (4.7) and the new projection-based score test against the horizontal axis = hypothesized θ_1 - true θ_1. Results are based on 5000 Monte-Carlo Trials with $k = 2, 4, 88$ instruments/moments and sample size $n = 100$. For all three, $\zeta = .5\%$ and $\tau = 4.5\%$ and hence the known upper bound for the asymptotic size of all three tests is 5%.