Finite-sample improvements of score tests by the use of implied probabilities from Generalized Empirical Likelihood *†

Saraswata Chaudhuri‡and Eric Renault§

Preliminary version: Comments are welcome

Date: December 3, 2011.

Abstract

We are interested in score tests for parameter vectors and sub-vectors defined by moment restrictions. We provide a general setup to conduct score tests by utilizing the additional information obtained from the Generalized Empirical Likelihood framework in the form of implied probabilities. Although most tests considered here are first-order asymptotically equivalent, we find by a series of simulation experiments that use of the empirical likelihood implied probabilities matches the estimated asymptotic size best with the nominal levels of the tests. Such tests can be computationally difficult when testing sub-vectors. We suggest a convenient test based on Neyman (1959)'s $C(\alpha)$ statistic that is asymptotically equivalent and performs similarly in finite samples. When elements of the parameter vector not specified by the null hypothesis are weakly identified, the conventional tests do not have correct asymptotic size and the direction of size-distortion is not clear. To prevent any uncontrolled upward size-distortion, we extend the projection-based test proposed by Chaudhuri and Zivot (2011) to the setup of the current paper. This test performs well in terms of finite-sample size and power in our simulation experiments.

*Part of this paper was previously circulated as "Projection-based GEL score tests for subsets of parameters with possible weak identification".

†We gratefully acknowledge helpful comments from A. Bester, S. Bhamidi, M. Caner, P. Guggenberger, C. Hansen, J. Hill, V. Krishna and the seminar participants at Chicago GSB, NC State, Ohio State, UNC, Joint Statistical meetings (2009), Midwest Econometrics Group (2010, 2011).

‡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, CB 3305, University of North Carolina, Chapel Hill, NC 27519. Telephone: 919-966-5326. Fax: 919-966-4986. Email: renault@email.unc.edu.

1 Introduction

Newey and Smith (2004) demonstrated that in terms of bias, the optimal (infeasible) Generalized Method of Moments (GMM) estimator is more closely mimicked by the Generalized Empirical Likelihood (GEL) estimators (see Smith (1997)) than the efficient two step GMM (2S-GMM) estimator (see Hansen (1982)).While both GEL and 2S-GMM are suitable for estimation and inference on parameters defined by moment restrictions, the property stated above makes GEL higher-order-asymptotically more attractive than the computationally convenient efficient 2S-GMM.

The question is: How much of these higher-order gains of GEL are actually reflected in finite samples of reasonable size? Typically simulation studies are tools to explore such questions. However, to our knowledge, there are not many such studies covering a variety of (econometrically relevant) cases that would be sufficient to get a conclusive answer on the benefit of using the GEL methods over 2S-GMM.[1] Perhaps as a consequence of this lack of simulation evidence, GEL methods still seem to be much less popular than 2S-GMM among applied researchers in economics (even post Newey and Smith (2004)).[2]

A primary goal of our paper is to fill this gap by seeking to explore the benefits/perils of using various classes of GEL methods over one other and 2S-GMM through a series of large-scale Monte-Carlo experiments covering cases where GEL have been theoretically shown to be beneficial. In these Monte-Carlo experiments, we also consider cases where 2S-GMM is known to perform better than GEL: e.g., when estimation bias is not a concern, or when the moment vector in (2.1) is thick tailed (see for e.g., Newey and Smith (2004) and Guggenberger (2008)).

A practical problem with the GEL methods is the computational cost. GEL estimation of parameters involve solving a saddle-point optimization that becomes increasingly difficult computationally when there are multiple parameters and moment restrictions. Partly to avoid this computational burden, we focus on score tests when the parameters are completely ($H_0 : \theta = \theta_0$) or partially ($H_{10} : \theta_1 = \theta_{10}$) specified by the null hypothesis where θ is the parameter vector defined by (2.1) and θ_1 is a sub-vector of θ.[3] We note that the desirable higher-order properties of the

[1]One strand of the literature that has explored the use of a particular class of GEL, i.e., the Continuous Updating GMM (CU-GMM), is the weak instrument/identification literature. See for example, Staiger and Stock (1997), Stock and Wright (2000), Kleibergen (2002), Kleibergen (2005), Dufour and Taamouti (2005a), etc. Exceptions, that actually compare the various class of GEL (occasionally in other contexts) are Mittelhammer et al. (2005), Guggenberger and Smith (2005), Guggenberger and Hahn (2005), Guggenberger (2008) and Caner (2010).

[2]An ad hoc search of the keywords "generalized method of moments", "generalized empirical likelihood", "continuous updating generalized method of moments", "empirical likelihood", "exponential titling" yielded respectively (19, 0, 2, 0, 0) in the Journal of Political Economy (2004-2011), (21, 0, 1, 0, 0) in the Quarterly Journal of Economics (2004-2011) and (34, 0, 0, 0, 1) in the American Economic Review (2004-2008). There were some other occurrences of the term "empirical likelihood" but in unrelated contexts.

[3]In terms of point-estimation, a practical approach to avoid the computational problem is to use approximation-based estimators that become asymptotically equivalent to the GEL estimator. See, for example, Antoine et al. (2007) and Fan et al. (2011). We will treat the first reference elaborately in the sequel.

GEL estimators over 2S-GMM (following the arguments of Newey and Smith (2004)) are likely to extend to the GEL score tests by reweighing the estimators of the Jacobian and the variance matrix of the moment vector using the so-called "implied probabilities".[4] Therefore, any evidence of the superior finite-sample performance of GEL score tests over 2S-GMM score tests is of importance to the practitioners.

The saddle-point optimization cannot, however, be avoided by the conventional plug-in score tests when interest lies only on a subset of parameters, i.e., when the null hypothesis is of the form $H_{10} : \theta_1 = \theta_{10}$. In such cases the conventional plug-in score tests require the estimated value of the nuisance parameters to be plugged in the score statistics.[5] To avoid such computations, we propose the use of a GEL score statistic that is analogous to Neyman (1959)'s $C(\alpha)$ statistic and can be asymptotically equivalent to the conventional plug-in score statistic.

There is an additional theoretical problem with the GEL tests for subsets of parameters. Under an important scenario considered in this paper we allow the parameters to be weakly identified following the framework of Stock and Wright (2000). In case any element of the nuisance parameters is weakly identified, one cannot obtain a consistent estimator of that element (even when the null hypothesis is true) and as a result the conventional plug-in GEL score tests and the $C(\alpha)$-type tests do not have correct asymptotic size.[6] We say that the asymptotic size of a test is incorrect when it is different from the nominal level that is used to determine the critical value for the test. We recall here that the tests for subsets of parameters discussed in Kleibergen (2005) and Guggenberger and Smith (2005)(GS-05, henceforth) assume that the nuisance parameters are strongly (i.e., not weakly) identified. While recently Kleibergen and Mavroeidis (2009) contend that such score tests (based on CU-GMM, a particular class of GEL) are downward size-distorted, i.e., asymptotic size is less than nominal size; Chen and Guggenberger (2011) contend that they can be upward size-distorted (typically considered a more serious problem in economics). In the view of this confusion and deeming downward size-distortion as the lesser problem, it seems that the conventional projection-based tests, such as (extensions of) those advocated by, among others, Dufour and Taamouti (2005b), are the safest choices for practitioners.

However, it is also known that these conventional projection-based tests for subsets of parameters can be needlessly conservative (even in the absence of weak identification). To the best of our

[4]We expect the higher-order gains of the GEL estimators in terms of bias are likely to result in better size properties of the GEL score tests. To the best of our knowledge, GEL estimators do not result in efficiency gains, unless bias corrected. Hence we do not expect the GEL score tests to have an advantage in terms of power.

[5]We should qualify the usage of the term "nuisance parameters". While moment restriction models are intrinsically semi-parametric in nature and involve infinite dimensional nuisance parameters, our usage of the term is rather simplistic. In the context of (2.1), we refer to the sub-vector of θ that is not specified by the null hypothesis as the nuisance parameters.

[6]Newey and West (1987)'s 2S-GMM score test of $H_0 : \theta = \theta_0$ or $H_{10} : \theta_1 = \theta_{10}$ has incorrect asymptotic size when any element of the entire parameter vector is weakly identified, and hence is even less robust to weak identification.

knowledge, the projection-based test proposed by Chaudhuri and Zivot (2011), in the context of CU-GMM, is the least conservative projection test. They showed that this test: (i) is less conservative than the conventional projection-based tests, (ii) always guards against uncontrolled upward size-distortion, and (iii) in the absence of weak identification, can be made asymptotically as powerful as the conventional plug-in based tests. In this paper we extend this test to the entire GEL class and thus obtain a GEL test for subsets of parameters, which we henceforth call the GEL-projection test. This GEL test is not upward size-distorted (unlike the conventional plug-in based GEL tests) and at the same time is more powerful than the conventional projection-based test. This is the methodological contribution of our paper.

The rest of the paper is organized as follows. In Section 2, we describe the framework of our paper and summarize the results obtained by simulations in the subsequent sections. In Section 3, we consider score tests for null hypotheses of the form $H_0 : \theta = \theta_0$, (i.e., for the entire parameter vector) and conduct two simulation experiments to conduct a study of the relative performance of the 2S-GMM and GEL score tests. In Section 4, we do the same for null hypotheses of the form $H_{10} : \theta_1 = \theta_{10}$ (i.e., for a sub-vector). In addition, we also discuss the GEL-projection test, establish its asymptotic properties, and compare its performance in finite samples with the other tests. Proofs of all the theoretical results are collected in the Technical Appendix.

Notations used throughout the paper: For any $a \times b$ matrix A, $\|A\| := \sqrt{\text{trace}(A'A)}$. 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 such that $A = A^{\frac{1}{2}} A^{\frac{1}{2}\prime}$. $\chi^2_{p,1-\alpha}$ denotes the $1-\alpha$-th quantile of a central χ^2 distribution with p degrees of freedom.

2 Framework and summary of simulation results

Suppose that we have observations $w_1, w_2, \ldots, w_n \in \mathcal{W}$ from the unknown probability distribution $F_{\theta^0,\eta^0} \in \mathcal{F} := \{F_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^p\}$ where for a given θ, the elements of η are a set of unknown nuisance parameters in $\mathcal{A}(\theta)$, a space of possibly infinite dimension and possibly constrained by θ. Now, suppose that the unknown "true" value $\theta^0 \in \text{interior}(\Theta)$ is uniquely identified by a set of moment restrictions given by

$$E[g(w, \theta)] = 0 \Leftrightarrow \theta = \theta^0, \tag{2.1}$$

where $g : \mathcal{W} \times \Theta \mapsto \mathbb{R}^k$ is a known function.[7] Being true to the literature on moment restrictions models, we let the expectation on the LHS of (2.1) to be with respect to the family of unknown probability distributions $F_{\theta^0,\eta}$ for all $\eta \in \mathcal{A}(\theta^0)$ (and not just F_{θ^0,η^0}). This notion of uniformity is

[7] We only consider unconditional moment restrictions. Extension of these results to conditional moment restrictions, while of theoretical relevance, is straightforward once the usual technical and computational issues are taken care of.

a key ingredient of the current paper and will be of paramount importance when we consider the size of tests in the presence of weak identification.

In this paper we consider various forms of score tests on θ (i.e., $H_0 : \theta = \theta_0$), and subsequently on its sub-vector θ_1 (i.e., $H_{10} : \theta_1 = \theta_{10}$), based on the first-order conditions of the efficient 2S-GMM and GEL objective functions. The principle behind the design of the score statistics is same as that of the 2S-GMM score statistic in Newey and West (1987). However, with the insight from Newey and Smith (2004), we also incorporate the additional information obtained from the GEL framework by using the GEL implied probabilities to obtain efficient estimators of the Jacobian and the variance matrix of the moment vector $g(w, \theta)$ (or, interchangeably, its sample average).

For the sake of completeness, we first briefly describe the necessary part of the GEL framework following Newey and Smith (2004) and then describe how to use the GEL implied probabilities to design the score statistics. To fix the idea, in this section, we focus on the entire vector θ, i.e., on null hypotheses of the form $H_0 : \theta = \theta_0$.

2.1 GEL implied probabilities and GEL score statistic:

The GEL class of estimators of θ^0 is indexed by the function ρ (see Assumption ρ below) and is defined as

$$
\begin{aligned}
&\widehat{\theta}_{\rho,n} := \arg\min_{\theta\in\Theta} \sup_{\lambda\in\Lambda_n(\theta)} \hat{Q}_{\rho,n}(\theta,\lambda) \\
\text{where}\quad &\hat{Q}_{\rho,n}(\theta,\lambda) := \frac{1}{n}\sum_{i=1}^{n} \rho(\lambda' g_i(\theta)) - \rho(0), \\
\text{and}\quad &\Lambda_n(\theta) := \{\lambda \in \mathbb{R}^k : \lambda' g_i(\theta) \in \mathcal{O}, \forall\ i = 1, \ldots, n\}.
\end{aligned}
$$

The choice of $\rho(.)$ leads to various types of GEL such as CU-GMM or Euclidean empirical likelihood (EEL) ($\rho(v) = -(1+v)^2/2, \mathcal{O} = \mathbb{R}$), empirical likelihood (EL) ($\rho(v) = \ln(1-v), \mathcal{O} = (-1, \infty)$), exponential tilting (ET) ($\rho(v) = -\exp[v], \mathcal{O} = \mathbb{R}$), etc. all of which satisfy Assumption ρ.

Assumption ρ: (GEL function)

$\rho : \mathcal{O} \mapsto \mathbb{R}$ is a continuous function such that

(i) ρ is concave on its domain $\mathcal{O}$ which is an open interval containing 0.

(ii) ρ is twice continuously differentiable on its domain. Defining $\rho_r(v) := \partial^r \rho(v)/\partial v^r$ for $r = 1, 2$ and $\rho_r := \rho_r(0)$, let $\rho_1 = \rho_2 = -1$ (standardization for convenience).

The desirable higher-order properties of the GEL estimators are precisely due to the GEL first order condition which, assuming differentiability of the moment vector $g(w, \theta)$ with respect to θ, is given by

$$o_P\left(\frac{1}{\sqrt{n}}\right) = \left[\sum_{i=1}^{n}\pi_{\rho,i,n}(\widehat{\theta}_{\rho,n})G_i(\widehat{\theta}_{\rho,n})\right]'\left[\sum_{i=1}^{n}\kappa_{\rho,i,n}(\widehat{\theta}_{\rho,n})g_i(\widehat{\theta}_{\rho,n})g_i'(\widehat{\theta}_{\rho,n})\right]^{-1}\bar{g}_n(\widehat{\theta}_{\rho,n}) \quad (2.2)$$

where for given θ and $\rho(.)$, $g_i(\theta) := g(w_i,\theta)$, $\bar{g}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} g_i(\theta)$, $G_i(\theta) := \frac{\partial}{\partial\theta'}g_i(\theta)$,

$$\lambda_{\rho,n}(\theta) := \arg\sup_{\lambda\in\Lambda_n(\theta)} \hat{Q}_{\rho,n}(\theta,\lambda), \quad (2.3)$$

$$\pi_{\rho,i,n}(\theta) := \frac{\rho_1(\lambda'_{\rho,n}(\theta)g_i(\theta))}{\sum_{j=1}^{n}\rho_1(\lambda'_{\rho,n}(\theta)g_j(\theta))} : \text{implied probabilities from GEL at a generic } \theta, \quad (2.4)$$

$$\kappa_{\rho,i,n}(\theta) := \frac{\kappa_\rho(\lambda'_{\rho,n}(\theta)g_i(\theta))}{\sum_{j=1}^{n}\kappa_\rho(\lambda'_{\rho,n}(\theta)g_j(\theta))}, \quad \kappa_\rho(v) := \frac{\rho_1(v)+1}{v} \text{ if } v\neq 0, \kappa_\rho(0) = -1.$$

Observe that $\rho(.)$ corresponding to EL leads to $\pi_{\rho,i,n}(\theta) = \kappa_{\rho,i,n}(\theta)$ for $i = 1,\ldots,n$. It is because of this along with a nice property of the implied probabilities $\pi_{\rho,i,n}(\theta)$ (discussed in Lemma-2.1 and Corollary-2.2 below) that results in the superior higher-order properties of the EL estimator (among the GEL class) discussed in Newey and Smith (2004).

Lemma 2.1 *Consider any $\theta \in interior(\Theta)$ such that*
(i) $c_n := \max_{1\le i\le n}\|g_i(\theta)\| = o_p(\sqrt{n})$,
(ii) $E[\bar{g}_n(\theta)] = O(1/n^s)$ for $s \ge 1/2$,[8]
(iii) $\sqrt{n}\,(\bar{g}_n(\theta) - E[\bar{g}_n(\theta)]) = O_P(1)$ (i.e., assume $V := Avar\,(\bar{g}_n(\theta)) = O(1)$ and hence as a consequence of (ii), $\bar{g}_n(\theta) = O_P(n^{-\min\{1/2,s\}}) = O_P(n^{-1/2})$),[9]
(iv) $b_{min} \le \gamma_{min}(\theta) \le \gamma_{max}(\theta) \le b_{max}$, almost surely (w) where $\gamma_{max}(\theta)$ and $\gamma_{min}(\theta)$ are the maximum and minimum eigen values of $\bar{V}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} g_i(\theta)g_i'(\theta)$, and b_{max} and b_{min} are (finite) positive constants,[10]
(v) there exists a (finite) positive constant b such that for each $v \in \mathcal{O}$, $|\rho_2(v) - \rho_2(0)| \le b \times |v|$.
Then the following results hold as $n \to \infty$ under assumptions ρ and (i)-(v):

(A) $\lambda_{\rho,n}(\theta)$ defined in (2.3) is such that $\lambda_{\rho,n}(\theta) = -\bar{V}_n^{-1}(\theta)\bar{g}_n(\theta) + o_P\left(n^{-1/2}\right)$,

(B) $\pi_{\rho,i,n}(\theta)$ defined in (2.4) is such that for a given $i = 1,\ldots,n$,

$$\pi_{\rho,i,n}(\theta) = \pi_{EEL,i,n}(\theta) + o_P\left(n^{-3/2}\right)$$

where $\pi_{EEL,i,n}(\theta)$'s are the implied probabilities from EEL with the closed-form expression

$$\pi_{EEL,i,n}(\theta) = \frac{1}{n}\left[1 - (g_i(\theta) - \bar{g}_n(\theta))'\bar{V}_n^{-1}(\theta)\bar{g}_n(\theta)\right] = \frac{1}{n} + O_P\left(n^{-3/2}\right).$$

[8] We are focusing on moment restrictions that are weak or worse.
[9] This will be the rationale behind our definition of the space of hypothesized values of θ in Section 3.
[10] Under (ii) and (iii), $\bar{V}_n(\theta) - V = o_P(1)$. Additionally, (iv) gives $\bar{V}_n^{-1}(\theta) - V^{-1} = o_P(1)$.

Remark: Note from (B) that the difference between the EEL and GEL implied probabilities is of a smaller order than that between the EEL implied probabilities and the naive empirical probabilities $\{1/n\}$. This suggests us to claim that the use of the GEL implied probabilities to re-weight observations results in equivalence up to one higher order and at the same time point out a wedge between the use of the GEL implied probabilities versus the naive empirical probabilities. However, this result, in itself, is not sufficient for such a claim because (B) is not uniform in $i = 1, \ldots, n$. We provide a proof of this claim in Corollary-2.2.

Corollary 2.2 *Consider any $\theta \in \Theta$ such that all the assumptions in Lemma-2.1 hold. Now consider n i.i.d. realizations $\{Y_{1,n}, \ldots, Y_{n,n}\}$ of a $d \times 1$ random vector Y_n. Denote $\bar{Y}_n = \sum_{i=1}^n Y_{i,n}/n$. Assume that*

(vi) $\bar{Y}_n - \mu_n \xrightarrow{P} 0$, $\frac{1}{n}\sum_{i=1}^n (Y_{i,n} - \mu_n)\left[(g_i(\theta) - \bar{g}_n(\theta))', Y_{i,n}'\right] \xrightarrow{P} [\Omega_{Yg}, \Omega_{YY}]$ *and that*

$$\begin{pmatrix} \sqrt{n}(\bar{Y}_n - \mu_n) \\ \sqrt{n}(\bar{g}_n(\theta) - E[\bar{g}_n(\theta)]) \end{pmatrix} \xrightarrow{d} N\left(0_{(d+k)\times 1}, \begin{bmatrix} \Omega_{YY} & \Omega_{Yg} \\ \Omega_{Yg}' & V \end{bmatrix}\right), \tag{2.5}$$

where $\Omega_{Yg} := ACov(\bar{Y}_n, \bar{g}_n(\theta)), \Omega_{YY} := AVar(\bar{Y}_n)$ *are finite.*[11]

(vii) $E\|Y_i\|^4 < \infty$ *and* $E\|g_i(\theta)\|^4 < \infty$.

Then the following results hold as $n \to \infty$ under assumptions ρ and (i)-(vi):

(A) $$\begin{pmatrix} \sqrt{n}\sum_{i=1}^n \pi_{EEL,i,n}(\theta)\,(Y_i - \mu_n) \\ \sqrt{n}(\bar{g}_n(\theta) - E[\bar{g}_n(\theta)]) \end{pmatrix} \xrightarrow{d} N\left(0_{(d+k)\times 1}, \begin{bmatrix} \Omega_{YY} - \Omega_{Yg}V^{-1}\Omega_{Yg}' & 0 \\ 0 & V \end{bmatrix}\right),$$

(B) $\sqrt{n}\sum_{i=1}^n \pi_{\rho,i,n}(\theta)\,(Y_i - \mu_n) - \sqrt{n}\sum_{i=1}^n \pi_{EEL,i,n}(\theta)\,(Y_i - \mu_n) \xrightarrow{P} 0.$

Remarks: (A) states that the EEL implied probabilities provide a revision over the naive empirical probabilities and, in case $\bar{g}_n(\theta)$ contains any information about $\bar{G}_n(\theta)$, this revision leads to gain in efficiency. Moreover, it also makes the re-weighted average of $Y_{i,n}$ asymptotically uncorrelated (and, hence, independent because of normality) of the moment vector $\bar{g}_n(\theta)$. This is the primary source of the improved higher-order properties of the GEL methods (see Newey and Smith (2004), ABR-07). This is also an important property that has been exploited in the literature to design score tests for weakly identified parameters (see Kleibergen (2005), Chaudhuri and Zivot (2011)). On the other hand, the next result, (B), states that up to the first-order the implied probabilities from all the members of the GEL class characterized by the function $\rho(.)$ gives the same revision. (B) will be the key justification for our general treatment of the various GEL score tests considered later in this paper instead of considering them separately for each GEL member.

[11] The assumption on the existence of a CLT is not essential for our argument, but is made for the sake of a reference in the subsequent sections of this paper. However, the convergence to and the existence of the asymptotic variances and covariances are essential. Standard extension to independent but not identically distributed data is possible.

Based on these observations, we consider a slightly modified and more general version of the GEL first-order condition in (2.2). In other words, we consider a modified version of estimating equations for θ. This is given by

$$l_{n,\theta}\left(\theta;\pi^G(\theta),\pi^V(\theta)\right) := \left[\sum_{i=1}^n \pi_i^G(\theta)G_i(\theta)\right]' \left[\sum_{i=1}^n \pi_i^V(\theta)V_i(\theta)\right]^{-1} \sqrt{n}\bar{g}_n(\theta) = o_P(1), \qquad (2.6)$$

where the notations are the same as before except that we take $V_i(\theta) := g_i(\theta)[g_i(\theta) - \bar{g}_n(\theta)]'$ in the deviation from mean form (shrinkage, see Hall (2000), Chaudhuri and Renault (2011)) with a hope of better finite-sample performance. The choice of weights $\pi_i^G(\theta)$ for the Jacobian and $\pi_i^V(\theta)$ for the weighting/variance matrix is dictated by the particular method used and this is what makes the setup general (encompassing all the existing feasible methods based on moment restrictions).[12,13] We illustrate this with some theoretically well known examples:

- 2S-GMM: $\pi_i^G(\theta) = \pi_i^V(\theta) = 1/n$ for all $i = 1, \ldots, n$ and a preliminary consistent estimator $\widetilde{\theta}$ in the expression for $V_i(\theta)$,
- EEL/CU-GMM: $\pi_i^G(\theta) = \pi_{EEL,i,n}(\theta)$, $\pi_i^V(\theta) = 1/n$ for all $i = 1, \ldots, n$,
- 3S-EEL: $\pi_i^G(\theta) = \pi_{EEL,i,n}(\theta)$, $\pi_i^V(\theta) = \pi_{EEL,i,n}(\theta)$ for all $i = 1, \ldots, n$ and a preliminary consistent estimator $\widetilde{\theta}$ in the expressions for $G_i(\theta)$ and $V_i(\theta)$,[14]
- EL: $\pi_i^G(\theta) = \pi_i^V(\theta) = \pi_{EL,i,n}(\theta)$ for all $i = 1, \ldots, n$,
- more generally, any GEL $\rho(.)$: $\pi_i^G(\theta) = \pi_{\rho,i,n}(\theta)$, $\pi_i^V(\theta) = \kappa_{\rho,i,n}(\theta)$ for all $i = 1, \ldots, n$.

In fact, our setup can also accommodate mixed methods like ETEL (see Schennach (2007)) by means of adding more notation to characterize the two different types of GEL, say $\widehat{\rho}(.)$ and $\bar{\rho}(.)$ (both satisfying Assumption ρ), in the expressions of the implied probabilities (in (2.4)):

$$\pi_{\widehat{\rho},\bar{\rho},i,n}(\theta) := \frac{\bar{\rho}_1(\lambda'_{\widehat{\rho},n}(\theta)g_i(\theta))}{\sum_{j=1}^n \bar{\rho}_1(\lambda'_{\widehat{\rho},n}(\theta)g_j(\theta))} \text{ and similarly } \kappa_{\widehat{\rho},\bar{\rho},i,n}(\theta) := \frac{\kappa_{\bar{\rho}}(\lambda'_{\widehat{\rho},n}(\theta)g_i(\theta))}{\sum_{j=1}^n \kappa_{\bar{\rho}}(\lambda'_{\widehat{\rho},n}(\theta)g_j(\theta))},$$

$$\text{where } \lambda_{\widehat{\rho},n}(\theta) := \arg \sup_{\lambda\in\Lambda_n(\theta)} \hat{Q}_{\widehat{\rho},n}(\theta,\lambda) \text{ [compare with (2.3)].}$$

For the specific case of ETEL, $\widehat{\rho} \equiv \rho_{ET}$ and $\bar{\rho} \equiv \rho_{EL}$. Since we are primarily interested in the size of score tests under the assumption of correctly specified models, we do not consider these mixed methods in the current paper.[15] Nevertheless, our results in Lemma-2.1 and Corollary-2.2 apply

[12] $\pi_i^G(\theta)$ and $\pi_i^V(\theta)$ can and will depend on sample size n. However, the dependence is suppressed for notational convenience with a hope that it will not be unduly confusing to the readers.

[13] Note that, in contrast, the infeasible optimal GMM estimator solves the first order condition, $E[G_i(\theta^0)]Var^{-1}(g_i(\theta^0))\bar{g}_n(\theta) = o_P(n^{-1/2})$.

[14] It is obvious that the shrinkage version of the EEL implied probabilities put forward by ABR-07 and Dovonon (2008) can be handled accordingly.

[15] A treatment of mixed methods, that is similar in spirit to our paper, can be found in Chaudhuri and Min (2012) in the context of doubly robust Augmented Inverse Probability Weighting estimators of average treatment effects.

to all the GEL and mixed methods (for mixed methods like ETEL, Lemma-2.1(A) is the key) and hence all the theoretical results in this paper apply equally to a wide variety of methods.

The corresponding score statistic for testing null hypotheses of the form $H_0 : \theta = \theta_0$ can be designed, following Newey and West (1987), as a quadratic form of $n^{-1/2} l_{n,\theta}\left(\theta_0; \pi^G(\theta_0), \pi^V(\theta_0)\right)$ defined in (2.6) with respect to the inverse of its asymptotic variance (under H_0). The asymptotic variance is likely to be unknown and needs to be replaced by a feasible estimator. In the spirit of using the implied probabilities in $l_{n,\theta}\left(\theta_0; \pi^G(\theta_0), \pi^V(\theta_0)\right)$, we do the same for the estimator of the asymptotic variance. Related technicalities are presented in Sections 3 and 4.

2.2 The motivation and summary of our findings:

In essence, all the methods considered are characterized by the weights used for estimating the Jacobian and the variance matrix. Re-weighting with the GEL implied probabilities kills the correlation (up to one higher order) of the average moment vector $\bar{g}_n(\theta)$ with the estimators of the Jacobian and the variance matrix. As a result, the approximation

$$E\left[l_{n,\theta}\left(\theta_0; \pi^G(\theta_0), \pi^V(\theta_0)\right)\right] = 0$$

is more accurate under $H_0 : \theta = \theta_0$. This is likely to result in less size-distortion of the score test in finite samples. There are at least two scenarios where this is known to be useful: (1) when the parameters θ are weakly identified, and (2) when the moment vector $g_i(\theta)$ is skewed. We consider both to motivate the questions asked in this paper, and then summarize below the answers found by simulation experiments.

Scenario 1: When θ is weakly identified, without any re-weighting, the estimator of the Jacobian can be correlated with the average moment vector causing the score test to be size-distorted in over-identified models (see Wang and Zivot (1998)). The solution of using the CU-GMM score statistic, which is equivalent to re-weighting the Jacobian estimator with the EEL implied probabilities, was proposed by Kleibergen (2005). Note that in this case the proposed solution is to use the relevant quadratic form of $l_{n,\theta}\left(\theta_0; \pi^G(\theta_0) = \pi_{EEL,i,n}(\theta_0), \pi^V(\theta_0) = 1/n\right)$ as the score statistic. For the purpose of our paper, it is important to observe the implicit and automatic re-scaling (by $\sqrt{n}$) in the presence of weakly identified θ:

$$\sqrt{n} l_{n,\theta}\left(\theta_0; \pi_{EEL,i,n}(\theta_0), \frac{1}{n}\right) = \left[\sqrt{n}\sum_{i=1}^{n} \pi_{EEL,i,n}(\theta_0) G_i(\theta)\right]' \left[\frac{1}{n}\sum_{i=1}^{n} V_i(\theta)\right]^{-1} \sqrt{n}\bar{g}_n(\theta_0).$$

Corollary-2.2(B) suggests that using any other GEL implied probabilities instead of EEL for re-weighting the Jacobian estimator does not change the first-order asymptotic results. Corollary-

2.2(A) suggests that using any GEL implied probabilities instead of the naive empirical probabilities $(1/n)$ for re-weighting the variance matrix estimator does not change the second-order asymptotic results. Also recall that in the spirit of the moment restrictions models in (2.1), under $H_0 : \theta = \theta_0$, these results are supposed to be uniform with respect to the possibly infinite dimensional nuisance parameters $\eta \in \mathcal{A}(\theta_0)$. Monte-Carlo experiment-I in Section 3 studies the accuracy of these results implied by Corollary-2.2 in samples of reasonable size and under various specifications of η.

Scenario 2: When the moment vector is skewed, without any re-weighting, the estimator of the variance matrix is correlated with the average moment vector causing the score test to be size-distorted (see Altonji and Segal (1996), Horowitz (1998)). A solution is to re-weight the variance matrix estimator with the EEL implied probabilities. This was proposed by ABR-07 and was formulated into a score statistic by Guay and Pelgrin (2008). Note that in this case the proposed solution is to use as the score statistic the relevant quadratic form of

$$l_{n,\theta}\left(\theta_0; \frac{1}{n}, \pi_{EEL,i,n}(\theta_0)\right) = \left[\frac{1}{n}\sum_{i=1}^{n} G_i(\theta)\right]' \left[\sum_{i=1}^{n} \pi_{EEL,i,n}(\theta_0) V_i(\theta)\right]^{-1} \sqrt{n}\bar{g}_n(\theta_0).$$

Corollary-2.2 suggests that using any other GEL implied probabilities or $(1/n)$ instead of EEL for re-weighting the variance estimator does not change the second-order asymptotic results. Monte-Carlo experiment-II in Section 3 studies the accuracy of these results implied by Corollary-2.2 in samples of reasonable size and under various specifications of η.

As a representative of GEL other than EEL, we select EL and use it throughout the paper. In addition to the results in Newey and Smith (2004), another reason behind this choice is the well known optimality properties of the EL likelihood ratio (LR) test (and its similarity with the LR test based on maximum likelihood estimation). See Kitamura (2006) for a survey. In the same article (see page 44), Kitamura also asks if this nice property of EL carry over to the score type tests. Our setup provides a platform to examine this question.

We find that in the presence of weak identification re-weighting the Jacobian estimator with the EEL implied probabilities provides significant gain over 2S-GMM score tests. However, although innocuous for first-order asymptotics, additionally re-weighting the variance estimator by the EEL implied probabilities can cause severe size-distortion in small samples. On the other hand, in the case of the EL implied probabilities, re-weighting both the Jacobian and the variance matrix estimators provide the most accurate approximation, although re-weighting only the Jacobian estimator becomes equally good in relatively large samples.

In the presence of skewed moment vector, using the EL implied probabilities to re-weight both the Jacobian and the variance matrix estimators provide the most accurate approximation. Using EEL implied probabilities produce bad approximation which can be even worse than that of 2S-

GMM in relatively small samples. We note here that Chaudhuri and Renault (2011) found a partly similar result in the context of a simulation experiment in covariance structure models in the sense that EL performed better than EEL, which in turn performed better than 2S-GMM.

Over all, the asymptotic equivalence of the GEL class only seems to hold when the sample size is relatively large. Perhaps a way to justify the break down of asymptotic equivalence is to note that the result on the order of magnitude of the difference in GEL implied probabilities in Lemma–2.1(A) is not uniform over observations $i = 1, \ldots, n$. While shrinkage in the spirit of ABR-07 and Dovonon (2008) can correct for the negative EEL implied probabilities, often, without proper trimming, certain influential observations can result in large EEL implied probabilities and thus distort the asymptotic equivalence results and subsequently the behavior of the score test. Note that the EL implied probabilities are between 0 and 1 by construction and, in a correctly specified model, are quite stable (Schennach (2007)). Probably this is the reason behind the better performance of score tests based on the EL implied probabilities.

Monte-Carlo experiments III and IV in Section 4 study the same for testing a sub-vector of θ, i.e., for null hypotheses of the form $H_{10} : \theta_1 = \theta_{10}$. For θ_2 such that $\theta := (\theta_1', \theta_2')'$, the sub-vector score tests require that an estimator of θ_2 (obtained by minimizing with respect to θ_2 the corresponding objective function constrained by $\theta_1 = \theta_{10}$) be plugged in the score statistic. We call them generically the conventional plug-in based score tests. Since such computations can be difficult with EL implied probabilities, we suggest the use of Neyman's $C(\alpha)$ form and replace θ_2 by any easy to obtain estimator that is $\sqrt{n}$-consistent under H_0 and local alternatives. This results in an asymptotically equivalent score test (when they are known to work) and our simulations in Monte-Carlo experiments III and IV show that it performs comparably to the conventional plug-in based score tests.

However, as mentioned in the Introduction, it is not clear if any of these plug-in based tests is not upward size-distorted when θ_2 is weakly identified (see Chen and Guggenberger (2011)). Hence we also extend the projection-based test of Chaudhuri and Zivot (2011) to the GEL setup described above. The main issue with this test is that: being of a projection type, in finite samples it may not be as powerful as the plug-in based tests when there is actually no problem of weak identification. This does not seem to be a serious problem in our simulation results from Monte-Carlo experiment III.

3 Testing for entire parameter vector, $H_0 : \theta = \theta_0$

3.1 Various Score tests and their asymptotic properties

As discussed in the last section, following Newey and West (1987), the score statistic for testing $H_0 : \theta = \theta_0$ is designed as a quadratic form of $l_{n,\theta}(\theta_0; \pi^G(\theta_0), \pi^V(\theta_0))/\sqrt{n}$ in (2.6) with respect to the inverse of an estimator of its asymptotic variance. The asymptotic variance is estimated by $\mathcal{I}_{n,\theta}(\theta_0; \pi^G(\theta_0), \pi^V(\theta_0))$ where, for any θ, we define similar to (2.6),

$$\mathcal{I}_{n,\theta}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right) := \left[\sum_{i=1}^{n} \pi_i^G(\theta) G_i(\theta)\right]' \left[\sum_{i=1}^{n} \pi_i^V(\theta) V_i(\theta)\right]^{-1} \left[\sum_{i=1}^{n} \pi_i^G(\theta) G_i(\theta)\right]. \quad (3.1)$$

Therefore, the generic form of the score statistic is $\mathfrak{LM}_{n,\theta}\left(\theta_0; \pi^G(\theta_0), \pi^V(\theta_0)\right)$ where

$$\mathfrak{LM}_{n,\theta}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right) := l'_{n,\theta}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right) \mathcal{I}^{-1}_{n,\theta}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right) l_{n,\theta}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right). \quad (3.2)$$

The score test rejects $H_0 : \theta = \theta_0$ at the nominal level α if

$$\mathfrak{LM}_{n,\theta}\left(\theta_0; \pi^G(\theta_0), \pi^V(\theta_0)\right) > \chi^2_{p,1-\alpha}.$$

While one is free to use different sets of weights to re-weight the estimator of the Jacobian and/or the variance matrix appearing in $l_{n,\theta}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right)$ and $\mathcal{I}_{n,\theta}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right)$ respectively, for the purpose of our simulations we choose to use the same set of weights. Also keeping in mind that three sets of weights that we are particularly interested in are the naive empirical probabilities and the implied probabilities from EEL and EL, the particular score statistics employed here are:

(A) 2S-GMM: $\mathfrak{LM}_{n,\theta}\left(\theta_0; 1/n, 1/n\right)$ – see Newey and West (1987).

(B1) EEL-Hybrid-1: $\mathfrak{LM}_{n,\theta}\left(\theta_0; \pi_{EEL,i,n}(\theta_0), 1/n\right)$ – see Kleibergen (2005).

(B2) EEL-Hybrid-2: $\mathfrak{LM}_{n,\theta}\left(\theta_0; 1/n, \pi_{EEL,i,n}(\theta_0)\right)$ – see Guay and Pelgrin (2008).

(B3) EEL-Hybrid-3: $\mathfrak{LM}_{n,\theta}\left(\theta_0; \pi_{EEL,i,n}(\theta_0), \pi_{EEL,i,n}(\theta_0)\right)$ – motivated from ABR-07.

(C1) EL-Hybrid-1: $\mathfrak{LM}_{n,\theta}\left(\theta_0; \pi_{EL,i,n}(\theta_0), 1/n\right)$ – see GS-05.

(C2) EL-Hybrid-2: $\mathfrak{LM}_{n,\theta}\left(\theta_0; 1/n, \pi_{EL,i,n}(\theta_0)\right)$ – EL version of Guay and Pelgrin (2008).

(C3) EL-Hybrid-3: $\mathfrak{LM}_{n,\theta}\left(\theta_0; \pi_{EL,i,n}(\theta_0), \pi_{EL,i,n}(\theta_0)\right)$ – see GS-05.[16]

Since we allow for weak identification, it is well known that the score tests in (A), (B2) and (C2) do not have correct asymptotic size. It is also known, at least since Kleibergen (2005) and GS-05, that

[16]Since GS-05 do not explicitly state which variance estimator is used in their equations (3.5) and (3.6), both EL-Hybrid-1 and EL-Hybrid-3 are encompassed in their framework.

the other tests have correct asymptotic size. For the sake of completeness, we briefly list below the asymptotic properties of these other GEL score tests, i.e., (B1), (B3), (C1), (C3), under standard assumptions that also allow for weakly identified θ. See GS-05 for a comprehensive treatment of these tests.

Assumption Θ: (parameters and parameter spaces)

$\Theta = \Theta_w \times \Theta_s$ and $\theta^0 = (\theta_w^{0'}, \theta_s^{0'})'$ where $\theta_w^0 \in \text{interior}(\Theta_w) \subset \mathbb{R}^{p_w}$, $\theta_w^0 \in \text{interior}(\Theta_s) \subset \mathbb{R}^{p_s}$ and Θ_w, Θ_w are compact. (The cross-product form of Θ is not necessary but taken for convenience.)

Definition of hypothesized value θ_0:

$\theta_0 \in \Theta^n := \text{interior}\,(\Theta_w \times \Theta_s^n)$ where $\Theta_s^n := \{\theta_s^n = \theta_s^0 + d_s/\sqrt{n}$ for some $d_s \in \mathbb{R}^{p_s}\} \subseteq \Theta_s$.

Assumption ID: (weak identification)

The expectation of the average moment vector is $E[\bar{g}_n(\theta)] = m_n^w(\theta)/\sqrt{n} + m(\theta_s)$ where

(i) $m_n^w(\theta) : \Theta \mapsto \mathbb{R}^k$ is a continuous function such that $m_n^w(\theta) \to m^w(\theta)$ and $M_n^w(\theta) := \partial m_n^w(\theta)/\partial\theta' \to M^w(\theta)$ uniformly in $\theta \in \Theta^n$. $m^w(\theta^0) = 0$, $m^w(\theta)$ and $M^w(\theta)$ are uniformly bounded for $\theta \in \Theta^n$ and continuous in θ_s at $\{\theta_s^0\}$.

(ii) $m(\theta_s) : \Theta_s \mapsto \mathbb{R}^k$ is a continuous function and $m(\theta_s) = 0$ if and only if $\theta_s = \theta_s^0$. $M(\theta_s) := \partial m(\theta_s)/\partial\theta_s'$ such that it is uniformly bounded for $\theta_s \in \Theta_s^n$, and continuous with full column rank at $\theta_s = \theta_s^0$.

Assumption S: (moment vector and its derivative)

(i) $max_{1 \le i \le n} \sup_{\theta\in\Theta^n} \|g_i(\theta)\| = o_p(\sqrt{n})$.

(ii.a) $g_i(\theta)$ is twice continuously differentiable in $\theta \in \Theta^n$.

(ii.b) $\bar{G}_n(\theta) := \partial\bar{g}_n(\theta)/\partial\theta' = [\bar{G}_{wn}(\theta), \bar{G}_{sn}(\theta)] = E[\bar{G}_n(\theta)] + o_p(1)$ uniformly in $\theta \in \Theta^n$ where $E[\bar{G}_n(\theta)] = \partial E[\bar{g}_n(\theta)]/\partial\theta' = M_n^w(\theta)/\sqrt{n} + [0, M(\theta_s)]$ by imposing interchangeability of the order of differentiation and integration (and from Assumption ID).

(ii.c) $\partial vec(\bar{G}_{w,n}(\theta))/\partial\theta_s' = \mathcal{G}_w(\theta) + o_p(1)$ uniformly in $\theta \in \Theta^n$ where $\mathcal{G}_w(\theta)$ is uniformly bounded for $\theta \in \Theta^n$ and is continuous in θ_s at $\{\theta_s^0\}$.

(iii) For $\theta_w \in \text{interior}(\Theta_w)$, let $\theta_{ws^0} := (\theta_w', \theta_s^{0'})'$. Assume that

$$\begin{bmatrix} \underset{k\times 1}{\Psi_{g,n}(\theta_{ws^0})} \\ \underset{kp_w\times 1}{\Psi_{w,n}(\theta_{ws^0})} \end{bmatrix} := \sqrt{n} \begin{bmatrix} \bar{g}_n(\theta_{ws^0}) - E[\bar{g}_n(\theta_{ws^0})] \\ vec(\bar{G}_{wn}(\theta_{ws^0}) - E[\bar{G}_{wn}(\theta_{ws^0})]) \end{bmatrix} \Rightarrow \begin{bmatrix} \underset{k\times 1}{\Psi_g(\theta_{ws^0})} \\ \underset{kp_w\times 1}{\Psi_w(\theta_{ws^0})} \end{bmatrix}$$

where $[\Psi_g'(\theta_{ws^0}), \Psi_w'(\theta_{ws^0})]$ is a mean zero Gaussian process with covariance matrix

$$\Delta(\theta_{w_a s^0}, \theta_{w_b s^0}) = \begin{bmatrix} \underset{k\times k}{\Delta_{gg}(\theta_{w_a s^0}, \theta_{w_b s^0})} & \underset{k\times kp_w}{\Delta_{gw}(\theta_{w_a s^0}, \theta_{w_b s^0})} \\ \underset{kp_w\times k}{\Delta_{wg}(\theta_{w_a s^0}, \theta_{w_b s^0})} & \underset{kp_w\times kp_w}{\Delta_{ww}(\theta_{w_a s^0}, \theta_{w_b s^0})} \end{bmatrix}$$

for any two $\theta_{w_a}, \theta_{w_b} \in \Theta_w$. $\Psi_g(\theta_{ws^0})$ and $\Psi_w(\theta_{ws^0})$ are uniformly bounded in probability for $\theta \in \text{interior}(\Theta_w) \times \{\theta_s^0\}$.

(iv.a) $V_{gg}(\theta)$ and $V_{wg}(\theta)$ are respectively $k \times k$ and $kp_w \times k$ matrices such that they are bounded for $\theta \in \Theta_w \times \{\theta_s^0\}$ and are continuous in θ_s at $\{\theta_s^0\}$. $V_{gg}(\theta)$ (and $V_{gg}^{-1}(\theta)$) is also positive definite for $\theta \in \Theta_w \times \{\theta_s^0\}$. Furthermore, $V_{wg}(\theta_{ws^0}) = \Delta_{wg}(\theta_{ws^0}, \theta_{ws^0}) = \Delta'_{gw}(\theta_{ws^0}, \theta_{ws^0}) = V'_{gw}(\theta_{ws^0})$ and $V_{gg}(\theta_{ws^0}) = \Delta_{gg}(\theta_{ws^0}, \theta_{ws^0})$.

(iv.b) $\widetilde{V}_{gg}(\theta) := \sum_{i=1}^n g_i(\theta))[g_i(\theta) - \bar{g}_n(\theta)]'/n \equiv \bar{V}_n(\theta)$ and $\widetilde{V}_{wg}(\theta) := \sum_{i=1}^n vec(G_{w,i}(\theta))[g_i(\theta) - \bar{g}_n(\theta)]'/n$ are respectively $k \times k$ and $kp_w \times k$ matrices that are continuous in θ_2 at $\{\theta_s^0\}$ and uniformly bounded in probability for $\theta \in \Theta^n$. Furthermore, $\widetilde{V}_{gg}(\theta) = V_{gg}(\theta) + o_p(1)$ and $\widetilde{V}_{wg}(\theta) = V_{wg}(\theta) + o_p(1)$ uniformly and $\widetilde{V}_{gg}(\theta)$ is positive definite almost surely for $\theta \in \Theta^n$.

Theorem 3.1 *For $\theta \in \Theta^n$ define $\widetilde{G}_n(\theta) := [\widetilde{G}_{wn}(\theta), \bar{G}_{sn}(\theta)]$ where $\widetilde{G}_{wn}(\theta)$ is a $k \times p_w$ matrix such that $vec(\widetilde{G}_{wn}(\theta)) := \sqrt{n} vec(\bar{G}_{wn}(\theta)) - \widetilde{V}_{wg}(\theta)\widetilde{V}_{gg}^{-1}(\theta)\sqrt{n}\bar{g}_n(\theta)$. Then the following results hold as $n \to \infty$ under assumptions Θ, ID, S and ρ:*

(A) If $\pi_i^G(\theta) = \pi_{\rho,i,n}(\theta)$ and $\pi_i^V(\theta) = 1/n$ or $\pi_{\rho,i,n}(\theta)$ for all $i = 1, \ldots, n$, then

$$\mathfrak{LM}_{n,\theta}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right) = \widetilde{l}'_{n,\theta}(\theta)\widetilde{\mathcal{I}}_{n,\theta}^{-1}(\theta)\widetilde{l}_{n,\theta}(\theta) + o_p(1)$$

uniformly in $\theta \in \Theta^n$ where $\widetilde{l}_{n,\theta}(\theta) := \widetilde{G}'_n(\theta)\widetilde{V}_{gg}^{-1}(\theta)\sqrt{n}\bar{g}_n(\theta)$ and $\widetilde{\mathcal{I}}_{n,\theta}(\theta) := \widetilde{G}'_n(\theta)\widetilde{V}_{gg}^{-1}(\theta)\widetilde{G}_n(\theta)$.

(B) Let the hypothesized value θ_0 be such that $\theta_0 := \theta_{ws^n} = (\theta'_w, \theta_s^{n'})' \in \Theta^n$ where $\theta_w \in$ is fixed and $\theta_s^n = \theta_s^0 + d_s/\sqrt{n}$ for some fixed d_s. Define $\widetilde{G}(\theta_w, d_s) := [\widetilde{G}_w(\theta_w, d_s), M(\theta_s^0)]$ where $\widetilde{G}_w(\theta_w, d_s)$ is a $k \times p_w$ matrix such that

$$vec(\widetilde{G}_w(\theta_w, d_s)) := \underbrace{\Psi_{w.g}(\theta_{ws^0}) + \left[\mathcal{G}_w(\theta_{ws^0})d_s - V_{wg}(\theta_{ws^0})V_{gg}^{-1}(\theta_{ws^0})[m^w(\theta_{ws^0}) + M(\theta_s^0)d_s]\right]}_{\text{Compare with Corollary-2.2(A)}},$$

$\theta_{ws^0} := (\theta'_w, \theta_s^{0'})'$ and where $\Psi_{w.g}(\theta_{ws^0}) := \Psi_w(\theta_{ws^0}) - V_{wg}(\theta_{ws^0})V_{gg}^{-1}(\theta_{ws^0})\Psi_g(\theta_{ws^0})$ is independent of $\Psi_g(\theta_{ws^0})$ by construction. If $\pi_i^G(\theta) = \pi_{\rho,i,n}(\theta)$ and $\pi_i^V(\theta) = 1/n$ or $\pi_{\rho,i,n}(\theta)$ for all $i = 1, \ldots, n$, then

$$\mathfrak{LM}_{n,\theta}\left(\theta_0; \pi^G(\theta_0), \pi^V(\theta_0)\right) \xrightarrow{d} \widetilde{g}'(\theta_w, d_s)V_{gg}^{-1/2}(\theta_{ws^0})P\left(V_{gg}^{-1/2'}(\theta_{ws^0})\widetilde{G}(\theta_w, d_s)\right)V_{gg}^{-1/2'}(\theta_{ws^0})\widetilde{g}(\theta_w, d_s)$$

where $\widetilde{g}(\theta_w, d_s) := \Psi_g(\theta_{ws^0}) + [m^w(\theta_{ws^0}) + M(\theta_s^0)d_s)]$.

Part(A) of the theorem states that the score statistics in which the estimator of the Jacobian is re-weighted by the GEL implied probabilities are first-order asymptotically equivalent under our assumptions. This, naturally, is a sufficient condition for the first-order asymptotic equivalence of the corresponding score tests.

Part(B) of the theorem specifies the hypothesized value θ_0 and states the asymptotic distribution of the score statistic at this particular value. When $\theta_w = \theta_w^0$ and $d_s = 0$, i.e., at $\theta_0 = \theta^0$, the score statistic $\mathfrak{LM}_{n,\theta}\left(\theta_0; \pi^G(\theta_0), \pi^V(\theta_0)\right) \xrightarrow{d} \chi_p^2$ conditional on $\Psi_{w.g}(\theta^0)$, and hence unconditionally. Therefore, the score test has correct asymptotic size. On the other hand, for deviations from the truth in the form of $\theta_w \neq \theta_w^0$ and/or $d_s \neq 0$, the statistic $\mathfrak{LM}_{n,\theta}\left(\theta_0; \pi^G(\theta_0), \pi^V(\theta_0)\right)$, conditional on $\Psi_{w.g}(\theta_{ws^0})$, converges in distribution to a non-central χ_p^2 with non-centrality parameter $\mu(\theta_w, d_s)'\mu(\theta_w, d_s)$ where

$$\begin{aligned} \mu(\theta_w, d_s) &:= \left[\widetilde{G}'(\theta_w, d_s)V_{gg}^{-1}(\theta_{ws^0})\widetilde{G}(\theta_w, d_s)\right]^{-1/2'} \widetilde{G}'(\theta_w, d_s)V_{gg}^{-1}(\theta_{ws^0})[m^w(\theta_{ws^0}) + M(\theta_s^0)d_s)] \\ &= \left[M'(\theta_s^0)V_{gg}^{-1}(\theta_{ws^0})M(\theta_s^0)\right]^{1/2'} d_s \text{ if } p_w = 0. \end{aligned}$$

Therefore, in the presence of both weakly and strongly identified parameters the score test (conditional on $\Psi_{w.g}(\theta_{ws^0})$) has nontrivial power against $\sqrt{n}$-deviations along the strongly identified components. Along the weakly identified components, the score test can have nontrivial power only against fixed deviations from the true value. The second line of the above equality leads to the standard (unconditional) power function of a score test when all the elements of θ are strongly identified, and holds for all the score tests (A), (B1)-(B3), (C1)-(C3) considered here.

Finally, we note that unlike the problem with weak identification, which manifests itself in the first-order asymptotics ($O(1)$), the problem with the skewed moment vector only manifests in the second order ($O(n^{-1/2})$), and hence a convenient and convincing theoretical modeling of it is more difficult. For this reason, simulation based study of their effect in finite samples seems to be useful. We study the finite sample properties (rejection rates for true and false values of θ) of the score tests based on all the above score statistics in Monte-Carlo Experiments I and II.

3.2 Monte-Carlo Experiment I

3.2.1 Design

This is similar to Design III of GS-05 that studies the robustness of the score test to the skewness in the distribution of the structural errors in a linear IV model. We modify the design slightly so that the regressor is actually endogenous, while preserving the skewed distribution of the structural error. We draw i.i.d. copies of $w_i = (y_i, X_i, Z_i')$ for $i = 1, \ldots, n$ from the following DGP-I:

$$\begin{aligned} y &= X\theta^0 + u, \\ X &= Z'\Pi + \vartheta, \\ u &= \rho\vartheta + \sqrt{\frac{1-\rho^2}{2}}(e^2 - 1), \end{aligned}$$

and $(e, \vartheta) \sim N(0, I_2)$ independent of $Z \sim N(1_k, I_k)$. We define $\Pi := C/\sqrt{n}$ where $C = c1_k$ is generated such that the concentration parameter $\mu = \Pi' \sum_{i=1}^{n} Z_i Z_i' \Pi / k$ is 0 (complete un-identification), 1 (weak identification) or 10 (strong identification). The sample size n is chosen to be 100 (small sample) and 1000 (reasonable for a micro-econometric application). The number of instruments is chosen as $k = 2, 4, 8, 16$ in a way that attempts to capture the fixed (number of) moment asymptotics, the many weak moment asymptotics (for $n = 100$, $k = 8, 16$ and $\mu = 1$) and the many strong moment asymptotics (for $n = 100$, $k = 8, 16$ and $\mu = 10$). The level of endogeneity, ρ, of X and the skewness of the moment vector are made to vary inversely – skewness is 0 when $\rho = 1$. This helps to disentangle the effects of these two factors – endogeneity of the regressor and skewness of the structural error – on the inference of the structural coefficient θ. In particular, we choose $\rho = 0$ (skewness ≈ 11.3), $\rho = .5$ (skewness ≈ 7.4) and $\rho = .9$ (skewness $\approx .9$) to signify low, medium and high level of endogeneity (high, medium and low level of skewness) respectively.

In this case the moment vector is

$$g_i(\theta) \equiv g(w_i, \theta) = Z_i(y_i - X_i\theta), \text{ where } w_i = (y_i, X_i, Z_i')' \text{ is } i.i.d. \; \forall \; i = 1, \ldots, n, \tag{3.3}$$

and the specification of endogeneity, skewness, strength of instruments, etc. described above models certain relevant elements of the infinite dimensional nuisance parameters η to check the robustness of the approximation results in Theorem-3.1. The score tests only use the information/restriction (2.1) based on this moment vector. We ignore the other facets (such as independence of Z and u) of the DGP described above because the researcher is often reluctant to make these assumptions a priori. In particular, while estimating the asymptotic variance, we only assume that $Asym\ Var[\sqrt{n}\bar{g}_n(\theta^0)] = E[Z_i Z_i'(y_i - X_i\theta^0)^2]$ and not $E[Z_i Z_i'] \times Var[y_i - X_i\theta^0]$ although the latter is also true for DGP - I.[17] Hence the EEL Hybrid-1 version should not be confused with Kleibergen (2002)'s K test.

[17] A logical extension will be to ignore the i.i.d. nature of the data and use a HAC variance matrix for the average moment vector. However, we avoid this so as not to confound the answers to our primary questions by issues (like the choice of bandwidth) related to the HAC estimator.

3.2.2 Results

Based on 5000 Monte-Carlo trials, we report in Table-1 (for n = 100) and Table-2 (for n = 1000) the empirical rejection rates of the true value $\theta = \theta^0 (= 1)$ for all the tests with nominal level 5 %.

The 2S-GMM score test performs well where it is supposed to do so. However, with weak identification and/or skewed moment vector, it is severely size-distorted – the upward size-distortion is more prominent in case of the former. (The problem gets worse when the number of moment restrictions increases.) With the level of skewness and weak identification specified in this design, it seems that proper weighting is more necessary for the estimator of the Jacobian than for the estimator of the variance matrix, and is borne out by comparing the results corresponding to the Hybrid-1 versus the Hybrid-2 versions of the EEL and EL score tests. This should not be surprising because weak identification manifests as distortion in the first-order asymptotics ($\mathcal{O}(1)$) while skewness in the the second order ($\mathcal{O}(n^{-1/2})$). We find that the Hybrid-3 version of the EL score test provides more improvement over the Hybrid-1 version. However, contrary to ABR-07's result of the higher-order equivalence between EL and 3S-EEL, this is not true for EEL. Overall, considering the size-distortion due to endogeneity/weak identification, skewed moment vector and many moments (weak and strong), we find that the Hybrid-3 version of the EL score test performs the best. This is even more true when the sample size is large (n = 1000) as is found commonly in micro-econometric applications.

3.3 Monte-Carlo Experiment II

3.3.1 Design

This design is commonly employed in papers studying the properties of GEL-type estimators. See, for example, Hall and Horowitz (1996), Imbens et al. (1998) (Model 2), Kitamura (2001), Schennach (2007) and Dovonon (2008) (Design D) for discussions on the rationale behind this design. We draw i.i.d. copies of $w_i = (X_{1i}, X_{2i}, \ldots, X_{ki})$ for $i = 1, \ldots, n$ from the DGP-II:

$$X_1 \sim N(0, .16), X_2 \sim N(0, .16), X_j \sim \chi^2_1 \text{ for } j = 3, \ldots, k,$$

and $X_1, X_2, \ldots, X_k$ are independent. The moment vector is defined as

$$\begin{aligned} g_i(\theta) &\equiv g(w_i, \theta) = Z_i r(X_{1i}, X_{2i}, \theta), \\ \text{where } Z_i &= (1, X_{2i}, X_{3i}, \ldots, X_{ki})', \\ r(X_{1i}, X_{2i}, \theta) &= \exp\left[-.72 - (X_{1i} + X_{2i})\theta + 3X_{2i}\right] - 1 \end{aligned}$$

and the moment restrictions in (2.1) are satisfied for $\theta = \theta^0 = 3$.

We choose the sample size $n = 100, 200, 500, 1000$ and the number of moments $k = 3, 10$.

3.3.2 Results

Based on 5000 Monte-Carlo trials, we report in Table-3 the empirical rejection rates of the true value $\theta = \theta^0 (= 1)$ for all the tests with nominal level 5%.

While, in general, the finite-sample size for none of the score tests is very close to the nominal level (which, in this experiment is the first-order asymptotic size for all the score tests), the EL Hybrid-3 score test performs uniformly better than the rest for all specifications. Note that, unlike in Monte-Carlo Experiment I, here EEL Hybrid-3 can provide size-refinement over EEL Hybrid-1, and this is observed for relatively larger sample sizes. In this experiment, there is no problem of endogeneity but the moment vector is skewed. Hence proper weighting in the estimator of the variance matrix (i.e., Hybrid-2) should be more important than the same for the Jacobian (i.e., Hybrid-1). This is evident for the EL score tests but only supported for the EEL score tests with large sample size and large number of moments. The performance of 2S-GMM score test is poor but better than that of the EEL score tests and the EL Hybrid-1 score test. The ranking between the 2S-GMM score test and the EL Hybrid-2 score test is not clear. We conclude by noting again that the EL Hybrid-3 score test performs the best, unambiguously.

4 Testing for sub-vectors, $H_{10} : \theta_1 = \theta_{10}$

The last section provided evidence of the superior performance of score tests for the entire parameter vector, i.e, of $H_0 : \theta = \theta_0$, when the EL implied probabilities are used to re-weight the estimators of both the Jacobian and the variance matrix. However, the computational advantage of score tests, especially with the use of the EL implied probabilities, is less when we are only interested in a sub-vector of θ. This is because the elements of θ that are not specified by the null hypothesis need to be estimated.

4.1 Various Score tests: computation and other considerations

We briefly describe below the procedure for designing the conventional score statistic. Then we suggest an alternative and computationally convenient form of the score statistic following Neyman (1959)'s $C(\alpha)$ principle. This alternative form is considered throughout the paper as the plug-in score statistic. Finally we extend the projection-based score statistic proposed by Chaudhuri and Zivot (2011) to the GEL setup of our paper. Unlike the plug-in based tests, this GEL-projection test guarantees against uncontrolled upward size-distortion when any element of θ is weakly identified.

Consider the partition of $\theta = (\theta_1', \theta_2')'$ where θ_1 is $p_1 \times 1$ and θ_2 is $p_2 \times 1$ with $p = p_1 + p_2$. Accordingly, let θ_1^0 and θ_2^0 be such that $\theta^0 = (\theta_1^{0'}, \theta_2^{0'})'$ satisfies the moment restrictions given in (2.1). The null hypothesis specifies $\theta_1 = \theta_{10}$. However, the unknown "nuisance parameter" θ_2 is unspecified and the user needs to fix some value of θ_2 to plug-in, along with θ_{01}, in the score statistic given in (3.2). The standard practice is to plug-in $\widetilde{\theta}_{0n} = (\theta_{10}', \widetilde{\theta}_{2n}'(\theta_{10}))'$, where $\widetilde{\theta}_{2n}(\theta_{10}) := \widetilde{\theta}_{2n}\left(\theta_{10}; \pi^G(\widetilde{\theta}_{0n}), \pi^V(\widetilde{\theta}_{0n})\right)$ is the estimator of θ_2 under the restriction $\theta_1 = \theta_{10}$ and solves the (approximate) first-order conditions, i.e., the last p_2 rows of (2.2) with θ_1 replaced by θ_{10},:[18]

$$l_{n,2}\left(\widetilde{\theta}_{0n}; \pi^G(\widetilde{\theta}_{0n}), \pi^V(\widetilde{\theta}_{0n})\right) := \left[\sum_{i=1}^n \pi_i^G(\widetilde{\theta}_{0n}) G_{2,i}(\widetilde{\theta}_{0n})\right]' \left[\sum_{i=1}^n \pi_i^V(\widetilde{\theta}_{0n}) V_i(\widetilde{\theta}_{0n})\right]^{-1} \sqrt{n}\bar{g}_n(\widetilde{\theta}_{0n}) = o_P(1). \tag{4.1}$$

In connection with (2.2), note that we have used the partition $G_i(\theta) := \partial g_i(\theta)/\partial\theta' = [G_{1,i}(\theta), G_{2,i}(\theta)]$ where $G_{l,i}(\theta) := \partial g_i(\theta)/\partial\theta_l'$ is the $k \times p_l$ matrix of partial derivatives of $g_i(\theta)$ with respect to θ_l (for $l = 1, 2$). So $l_{n,2}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right)$ constitutes of the last $p_2 = p - p_1$ rows of $l_{n,\theta}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right)$. One can similarly define $l_{n,1}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right)$ as the first p_1 rows of $l_{n,\theta}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right)$ that corresponds to the $k \times p_1$ sub-matrix $G_{1,i}(\theta)$ of $G_i(\theta)$.

The conventional plug-in based score test rejects $H_{10} : \theta_1 = \theta_{10}$ at the nominal level α if $\mathfrak{LM}_{n,\theta}\left(\widetilde{\theta}_{0n}; \pi^G(\widetilde{\theta}_{0n}), \pi^V(\widetilde{\theta}_{0n})\right) > \chi^2_{p_1,1-\alpha}$. Therefore, with the conventional form of the score statistic, one cannot hope to exploit the information in the EL implied probabilities and at the same time avoid the EL estimation of θ_2.

To avoid such estimations, we suggest the use of an alternative form of the score statistic that can also be viewed as Neyman's $C(\alpha)$ statistic. This statistic is defined as

$$\mathfrak{LM}_{n,1.2}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right) := l_{n,1.2}'\left(\theta; \pi^G(\theta), \pi^V(\theta)\right) \mathcal{I}_{n,11.2}^{-1}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right) l_{n,1.2}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right). \tag{4.2}$$

The notations used in (4.2) are as follows. For the triplet $\xi := \left(\theta; \pi^G(\theta), \pi^V(\theta)\right)$,

$$\begin{aligned} l_{n,1.2}(\xi) &:= l_{n,1}(\xi) - \mathcal{I}_{n,12}(\xi)\mathcal{I}_{n,22}^{-1}(\xi) l_{n,2}(\xi), \\ \mathcal{I}_{n,11.2}(\xi) &:= \mathcal{I}_{n,11}(\xi) - \mathcal{I}_{n,12}(\xi)\mathcal{I}_{n,22}^{-1}(\xi)\mathcal{I}_{n,21}(\xi), \\ \text{where } \mathcal{I}_{n,ll'}\left(\theta; \pi^G(\theta), \pi^V(\theta)\right) &:= \left[\sum_{i=1}^n \pi_i^G(\theta) G_{l,i}(\theta)\right]' \left[\sum_{i=1}^n \pi_i^V(\theta) V_i(\theta)\right]^{-1} \left[\sum_{i=1}^n \pi_i^G(\theta) G_{l',i}(\theta)\right] \end{aligned}$$

[18]Unless, there is a chance of confusion we suppress the dependence of the estimator of θ_2 on the weights used for the estimator of the Jacobian and the variance matrix in (4.1). In cases where this dependence is made explicit, we suppress the dependence on the hypothesized value θ_{10} for the parameter of interest because that is obvious.

for $l, l' = 1, 2$. When (4.1) holds exactly (and not just up to $o_P(1/\sqrt{n})$), we obtain the following numerical equivalence,

$$\mathfrak{LM}_{n,1.2}\left(\widetilde{\theta}_{0n}; \pi^G(\widetilde{\theta}_{0n}), \pi^V(\widetilde{\theta}_{0n})\right) = \mathfrak{LM}_{n,\theta}\left(\widetilde{\theta}_{0n}; \pi^G(\widetilde{\theta}_{0n}), \pi^V(\widetilde{\theta}_{0n})\right).$$

The computational advantage of this alternative form was exploited by Chaudhuri and Zivot (2011). In the context of 2S-GMM and EEL-Hybrid-1 and in the absence of weak identification, they showed that for any hypothesized value θ_{10} in the $\sqrt{n}$-neighborhood of the true value θ_1^0,

$$\mathfrak{LM}_{n,1.2}\left((\theta_1^0, \theta_{2n}); \pi^G(\theta_1^0, \theta_{2n}), \pi^V(\theta_1^0, \theta_{2n})\right) = \mathfrak{LM}_{n,1.2}\left((\theta_1^0, \theta_2^0); \pi^G(\theta_1^0, \theta_2^0), \pi^V(\theta_1^0, \theta_2^0)\right) + o_P(1)$$

where θ_{2n} is any $\sqrt{n}$-consistent estimator of θ_2^0. It is obvious that under the same scenario this nice property extends to all the score tests considered in our paper. We prefer this $C(\alpha)$ form over the conventional form because the former allows the practitioner, who wishes to use the EL(GEL) implied probabilities for re-weighting but wants to avoid the EL(GEL) estimation of θ_2, to simply plug-in the computationally convenient 2S-GMM estimator of θ_2 and obtain asymptotically equivalent results in cases where the conventional plug-in score tests have correct asymptotic size.[19]

In the current paper we will refer to the GEL plug-in score test as the test that rejects $H_{10} : \theta_1 = \theta_{10}$ at the nominal level α if

$$\mathfrak{LM}_{n,1.2}\left(\widetilde{\theta}_{0n}; \pi^G(\widetilde{\theta}_{0n}), \pi^V(\widetilde{\theta}_{0n})\right) > \chi^2_{p_1, 1-\alpha},$$

i.e., the test that plugs in $\widetilde{\theta}_{2n}(\theta_{10})$ obtained as the solution of (4.1).

However, it is also known since GS-05 that when the elements of θ_2 are weakly identified, neither the conventional plug-in score test nor the GEL plug-in score test has correct asymptotic size.[20] As mentioned in the Introduction, there seems to be some confusion in the literature regarding the asymptotic size of these plug-in tests. In the context of CU-GMM, while Kleibergen and Mavroeidis (2009) contend that such score tests are downward size-distorted, Chen and Guggenberger (2011) contend that they can be upward size-distorted.

In the view of this confusion we think it is safer to use some kind of projection methods for such tests of sub-vectors. Accordingly we extend here the projection-based test proposed by Chaudhuri and Zivot (2011) to the entire GEL framework in the full generality characterized by (2.2). We

[19]This convenience is similar in spirit to the higher-order asymptotic equivalence between ABR-7's 3S-EEL estimator with the EL estimator. Recall that ABR-07's asymptotic equivalence, unfortunately, did not translate into similar finite-sample properties of the EEL-Hybrid-3 and EL-Hybrid-3 score tests in the last section. Therefore, it will be important to check the accuracy of this computation-facilitating approximation in the subsequent Monte-Carlo experiments.

[20]Recall that if the empirical probabilities $(1/n)$ are used for $\pi^G(\widetilde{\theta}_{0n})$, then the test has incorrect asymptotic size when any element of θ is weakly identified.

call this the GEL-projection test. Chaudhuri and Zivot (2011) demonstrated that this particular kind of projection is less conservative than the usual projection-based tests (also see Zivot and Chaudhuri (2009), Chaudhuri et al. (2010)). This GEL-projection test, in its generic form, rejects $H_{10} : \theta_1 = \theta_{10}$ if

$$\inf_{\theta_2 \in \mathcal{C}_{2n}(1-\tau,\theta_{10})} \mathfrak{LM}_{n,1.2}\left((\theta_{01},\theta_2);\pi^G(\theta_{01},\theta_2),\pi^V(\theta_{01},\theta_2)\right) > \chi^2_{p_1,1-\alpha} \tag{4.3}$$

where $\mathcal{C}_{2n}(1-\tau,\theta_{10})$ is a region, possibly dependent on θ_{10} and n, such that it contains θ_2^0 with probability approaching (at least) $1-\tau$ whenever $\theta_1^0 = \theta_{10}$. Additionally, for the asymptotic equivalence with the plug-in GEL score tests in the absence of weak identification, we would require that if $\theta_1^0 = \theta_{10} + d_1/\sqrt{n}$, then $\mathcal{C}_{2n}(1-\tau,\theta_{10})$ belongs in the $\sqrt{n}$-neighborhood of θ_2^0 almost surely. Finding such a region is not difficult. When θ_2 is weakly identified one can invert a weak-identification robust test such as the Anderson-Rubin-type tests, Kleibergen (2005)'s K test, or more generally the tests described in GS-05. When θ_2 is not weakly identified, one can simply obtain a Wald-type confidence region based on the estimator $\widetilde{\theta}_{2n}(\theta_{10}) := \widetilde{\theta}_{2n}\left(\theta_{10};\pi^G(\widetilde{\theta}_{0n}),\pi^V(\widetilde{\theta}_{0n})\right)$ with a convenient choice of $\pi^G(\widetilde{\theta}_{0n})$ and $\pi^V(\widetilde{\theta}_{0n})$.

4.2 GEL-projection test: asymptotic properties

The conventional plug-in tests have already been extensively studied in Kleibergen (2005), GS-05 , Kleibergen and Mavroeidis (2009) and Chen and Guggenberger (2011). Since it is not clear that they are never size-distorted upwards under our setup, in this section we do not report their asymptotic properties. However, there is some thing apparently new about our claim to the computational advantage of the alternative form of plug-in tests. We claimed that when the conventional plug-in tests are known to have correct asymptotic size, i.e., when θ_2 is not weakly identified, one can safely replace θ_2 by, say, its computationally convenient 2S-GMM estimator and at the same time use the EL implied probabilities to design the sub-vector score test. We prove this result in Lemma-4.1.

Other than this, we focus exclusively on the projection-GEL test in this sub-section. By virtue of projection, and using the results of the full-vector tests in GS-05, it is straightforward to show by Bonferroni arguments that the asymptotic size of this test can always be bounded by any pre-specified level $(\tau + \alpha)$.[21] What is important about this test is the following: This test is asymptotically equivalent to the plug-in based tests when there is no problem of weak identification. No other projection-based test shares this property. Hence in our discussion (theoretical and simulation-based) of the projection-GEL test we will focus on this particular property.

The assumptions under which the asymptotic properties are stated are same as those for the

[21]Note that no test has correct asymptotic size under the assumptions our paper.

test of $H_0 : \theta = \theta_0$. However, since our interest is in the sub-vector test under the scenario where the elements of θ_1 and θ_2 can both be weakly and strongly (not weakly) identified, it will be useful to point out the connection by specifying the regrouping and partition of the spaces, vectors and matrices defined in assumptions Θ, ID and S and definition Θ^n. We list them below:

(Θ): $\theta := (\theta_1', \theta_2')'$, $\theta_1 := (\theta_{1w}', \theta_{1s}')'$, $\theta_2 := (\theta_{2w}', \theta_{2s}')'$ and accordingly the partition $\Theta_1 = \Theta_{1w} \times \Theta_{1s}$ and $\Theta_2 = \Theta_{2w} \times \Theta_{2s}$, $\Theta_w := \Theta_{1w} \times \Theta_{2w}$, $\Theta_s = \Theta_{1s} \times \Theta_{2s}$ where $\Theta_{lj} \subseteq \mathbb{R}^{p_{lj}}$ is compact for $l = 1, 2$, $j = w, s$.

(Θ^n): Same as Θ.

(ID): $M^w(\theta) = [M_{1w}^w(\theta), M_{1s}^w(\theta), M_{2w}^w(\theta), M_{2s}^w(\theta)]$ where $M_{lj}^w(\theta) := \partial m^w(\theta)/\partial\theta_{lj}'$ for $l = 1, 2$ and $j = w, s$. $M(\theta_s) = [M_1(\theta_s), M_2(\theta_s)]$ where $M_l(\theta_s) := \partial m(\theta_s)/\partial\theta_{ls}'$ for $l = 1, 2$.

(S): (ii.b) $E[\bar{G}_n(\theta)] = \partial E[\bar{g}_n(\theta)]/\partial\theta' = M_n^w(\theta)/\sqrt{n} + [0, M_1(\theta_s), 0, M_2(\theta_s)]$.

(S): (ii.c) Holds for $\partial vec(\bar{G}_{lw,n}(\theta))/\partial\theta_s'$ for $l = 1, 2$.

(S): (iii) The partition of $\Psi_w = [\Psi_{1w}', \Psi_{2w}']'$, $\Delta_{gw} = [\Delta_{g1}, \Delta_{g2}] = \Delta_{wg}'$ and $\Delta_{ww} = (\Delta_{ll'})_{l,l'=1,2}$ 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.

(S): (iv.a) Same as S(iii).

(S): (iv.b) Same as S(iii).

In Lemma-4.1 we show the computational convenience provided by the alternative form of plug-in tests. As the reference test, we consider the EL-Hybrid-3 just for the sake of being specific. The result extends to entire GEL class of tests when the reference is known to have correct asymptotic size.

Lemma 4.1 *Let the hypothesized value θ_{10} be such that $\theta_{10} := (\theta_{1w}', \theta_{1s}^{n'})' \in \Theta^n$ where $\theta_{1s}^n = \theta_{1s}^0 + d_1/\sqrt{n}$ for some fixed d_1. Assume that $p_{2w} = 0$ i.e., there is no weakly identified element in θ_2.*[22] *Consider the following two estimators of θ_2 under the restriction imposed by $H_{10} : \theta_1 = \theta_{10}$:*

(i) $\widehat{\theta}_{n2} := \widetilde{\theta}_{n2}(\theta_{10}, \pi^G, \pi^V)$ *obtained by solving* (4.1) *for θ_2 with $\pi^G = 1/n$ or any $\pi_{\rho,i,n}$ and $\pi^V = 1/n$ or any $\kappa_{\rho,i,n}$,* [23]

(ii) $\theta_{n2}^{EL} := \widetilde{\theta}_{n2}(\theta_{10}, \pi_{EL,i,n}, \kappa_{EL,i,n})$ *obtained by solving* (4.1) *for θ_2, i.e., the EL estimator.*

Then the following holds as $n \to \infty$ under assumptions Θ, ID, S and ρ:

$$
\begin{aligned}
&\mathfrak{LM}_{n,1.2}\left((\theta_{10}, \theta_{n2}^{EL}); \pi_{EL,i,n}(\theta_{10}, \theta_{n2}^{EL}), \kappa_{EL,i,n}(\theta_{10}, \theta_{n2}^{EL})\right) \\
= \ &\mathfrak{LM}_{n,1.2}\left((\theta_{10}, \widehat{\theta}_{n2}); \pi_{EL,i,n}(\theta_{10}, \widehat{\theta}_{n2}), \kappa_{EL,i,n}(\theta_{10}, \widehat{\theta}_{n2})\right) + o_P(1).
\end{aligned}
$$

[22]This is crucial for our proof and also for the fact that the plug-in GEL tests are known to have correct asymptotic size under this assumption (see Kleibergen (2005) and GS-05).

[23]The recursive nature of the relationship in the expressions of $\pi_{\rho,i,n}$ and $\kappa_{\rho,i,n}$ is suppressed for notational convenience.

Remarks: This result follows directly from the proof Theorem 6 in GS-05 once we note that $\sqrt{n}\left(\theta_{n2}^{EL} - \widehat{\theta}_{n2}\right) = o_P(1)$ (see Theorem 1 of Qin and Lawless (1994) and Lemma A1 of Stock and Wright (2000)). In fact, in the absence of any weakly identified parameters, i.e., when $p_w = 0$, a stronger result holds: one can use any θ_2 in the $\sqrt{n}$-neighborhood of θ_2^0 to obtain the same asymptotic equivalence. This is the result that follows from Neyman (1959) and is exploited in Theorem-4.2(B) below following Chaudhuri and Zivot (2011). On the other hand, the result in Lemma-4.1 is the modification necessary to allow for weakly identified elements in the hypothesized parameter θ_1.

In the next result, we describe the relevant asymptotic properties of the GEL-projection test.

Theorem 4.2 *The following result holds as $n \to \infty$ under assumptions Θ, ID, ρ and S:*

(A) If the region $\mathcal{C}_{2,n}(1-\tau, \theta_{10})$ is such that $\liminf_{n\to\infty} P_{\theta^0}\left\{\theta_2^0 \in \mathcal{C}_{2,n}(1-\tau, \theta_{01})\right\} \geq 1-\tau$ when $\theta_{10} = \theta_1^0$ then the asymptotic size of the projection-based GEL score test defined in (4.3) cannot exceed $\tau + \alpha$.

(B) Let the hypothesized value θ_{10} be such that $\theta_{10} := (\theta_{1w}', \theta_{1s}^{n'})' \in \Theta^n$ where $\theta_{1s}^n = \theta_{1s}^0 + d_1/\sqrt{n}$ for some fixed d_1 and assume that $p_w = 0$ i.e., when there are no weakly identified parameters. Let the region $\mathcal{C}_{2,n}(1-\tau, \theta_{10})$ be non-empty and inside Θ_2^n almost surely. Then the projection-based GEL score test defined in (4.3) is asymptotically equivalent to the infeasible (because it uses unknown true θ_2^0) GEL score test that rejects $H_{10} : \theta_1 = \theta_{10}$ if $\mathfrak{LM}_{n,1.2}\left((\theta_{01}, \theta_2^0); \pi^G(\theta_{01}, \theta_2^0), \pi^V(\theta_{01}, \theta_2^0)\right) > \chi^2_{p_1, 1-\alpha}$.

Remarks:

(i) The first part of the theorem establishes that, like the other projection-based tests suggested in the literature (see Dufour (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005b, 2007), etc.), the asymptotic size of the GEL-projection test can also be bounded from above by any pre-specified upper bound $\tau + \alpha$. To the best of our knowledge, this property has not yet been proved for the plug-in-based GEL score tests of GS-05 when elements of θ_2 can be weakly identified.

(ii) The asymptotic equivalence of the projection-based score test with the infeasible score test (that plugs in the unknown true value θ_2^0 of θ_2) described in the second part is important. Note that the plug-in based GEL score tests of GS-05 are asymptotically equivalent to this infeasible test in the absence of weak identification. By virtue of the $C(\alpha)$ form of the test statistic and the restricted nature of the projection, one does not lose any asymptotic power in this regular case, i.e., when there is no problem of weak identification. This is not possible with the other projection-based tests. See Chaudhuri and Zivot (2011) for more discussion on these results.

(iii) We have not specified what $\mathcal{C}_{2,n}(1-\tau, \theta_{01})$ is. The weak identification robust confidence regions for θ_2, such as those obtained by inverting the K or the MQLR tests or their GEL counterparts, are possible candidates for such regions. Typically in addition to our assumptions, one would require that $M_2(\theta_s)$ is full column rank and $V_{gg}(\theta)$ is positive definite for $\theta \in \Theta_w \times \Theta_{1s}^n \times \Theta_{2s}$ for the condition required in Theorem 4.2(ii) to be satisfied. Computational advantage over the typical sub-vector tests can be achieved if one uses, for example, $\mathcal{C}_{2,n}(1-\tau, \theta_{01})$ based on EEL, in sub-vector test based on the EL implied probabilities.

4.3 Monte-Carlo Experiment III

4.3.1 Design

There is no reason to believe that the problems with the score tests for the entire parameter vector will disappear when testing sub-vectors. Given the additional complications of inference for sub-vectors, such as consistent estimation of the nuisance parameters, we first consider a very simple design by abstracting from the issues of weak identification. In this experiment we mainly focus on the issues related to the variance estimator of the average moment vector. We draw i.i.d. copies of w_i for $i = 1, \ldots, n$ from the DGP-III:

$$w_i \sim \text{Gamma}(\beta^0 = 1, \delta^0 = 2)$$

and wish to use the first two moments of the Gamma distribution, i.e., $E[w_i] = \beta^0\delta^0$ and $E[w_i^2] = \beta^0\delta^{0^2} + \beta^{0^2}\delta^{0^2}$ to conduct the score tests. However, in order to avoid the constrained estimation of the nuisance parameters (note that, by definition, the shape parameter β and the scale parameter δ of a Gamma distribution have to be positive) we re-parameterize them as

$$\theta_1 = \exp[\beta] \text{ and } \theta_2 = \exp[\delta],$$

giving $\theta_1^0 = 0$ and $\theta_2^0 = .6931$ (approximated). Accordingly, the moment vector is defined as

$$g_i(\theta) \equiv g(w_i, (\theta_1, \theta_2)) = \begin{bmatrix} w_i - \exp[\theta_1 + \theta_2] \\ w_i^2 - \exp[\theta_1 + 2\theta_2] - \exp[2\theta_1 + 2\theta_2] \end{bmatrix},$$

and satisfies the moment restrictions in (2.1) for $\theta = \theta^0 = (\theta_1^{0'}, \theta_2^{0'})'$. Now note that, for this experiment

$$G_i(\theta) = [G_{1,i}(\theta), G_{2,i}(\theta)] = -\begin{bmatrix} \exp[\theta_1 + \theta_2] & \exp[\theta_1 + \theta_2] \\ \exp[\theta_1 + 2\theta_2] + 2\exp[2\theta_1 + 2\theta_2] & 2\exp[\theta_1 + 2\theta_2] + 2\exp[2\theta_1 + 2\theta_2] \end{bmatrix}$$

is a constant (for given θ_1, θ_2), and hence there is no problem (e.g. size-distortion) due to weak identification. So weights for the estimator of the Jacobian is moot. However, the moment vectors are skewed – the first element of $g_i(\theta^0)$ has skewness 2, while the second element has skewness 6.6 (approx). Moreover, the fourth moments of the two elements of the moment vector are large - 144 (kurtosis = 36) and 8687616 (kurtosis = 84.84) respectively. Hence problem with the variance estimator might be an issue and, therefore, proper weighting for the estimator of the variance matrix need not be inconsequential. In this experiment we ignore the Hybrid-1 versions (which is same as 2S-GMM) and focus on the relative efficacy of the Hybrid-2 (equivalent to Hybrid-3 in this case) EEL and EL score tests over the 2S-GMM score test.

We choose the sample size $n = 100, 1000$.

4.3.2 Results

Based on 5000 Monte-Carlo trials, we report in Table-4 (columns 3-5) the empirical rejection rates of the true value $\theta_1 = \theta_1^0 (= 1)$ for the 2S-GMM, EEL Hybrid-2 (3) and EL Hybrid-2 (3) score tests with nominal level 5%. Since we observe that the finite-sample size of all these tests is much closer to the nominal level as compared to what we saw in the previous experiments, we also report the rejection rate (finite-sample power) for a grid of false hypothesized values in the same table.

A word on computation is in order here. The restricted EL estimator of θ_2 is computed using the Matlab code of John Zedlewski (web link – http://www.people.fas.harvard.edu/~jzedlews/matelike14.zip). The availability of this package greatly improved the computational efficiency. Nevertheless, even in this simple framework, we find it extremely difficult to the compute the EL estimator of the nuisance parameter θ_2 for some of the false values of the parameter of interest θ_1, especially when the sample size is small, $n = 100$. As a result, we are unable to the report the finite-sample power of the EL Hybrid-2 test with $n = 100$ because repeating the exercise of finding the restricted EL estimator 5000 times at each hypothesized value θ_{10} of θ_1 seemed quite formidable to us and the final solution was not stable.

The $C(\alpha)$ form of the score statistic in (4.2) becomes useful in this scenario. Recall that unlike the asymptotic size of the conventional plug-in score tests, the same based on the $C(\alpha)$ statistic is not affected when the nuisance parameter is replaced by any $\sqrt{n}$-consistent estimator.[24] In this case, we consider the EL Hybrid-2 (3) score test for which the nuisance parameter θ_2 is replaced by the restricted 2S-GMM estimator $\widetilde{\theta}_{2n}(\theta_{10}; 1/n, 1/n)$ instead of restricted EL estimator $\widetilde{\theta}_{2n}\left(\theta_{10}; \pi_{EL,i,n}(\widetilde{\theta}_{0n}), \pi_{EL,i,n}(\widetilde{\theta}_{0n})\right)$. We report the finite-sample rejection rates of EL Hybrid-2 (3) score test (with the 2S-GMM estimator plugged-in) in column 6 of Table-4.

[24] For this experiment the $\sqrt{n}$-local asymptotic power of the score test based on the $C(\alpha)$ statistic is also not affected by such replacements of the nuisance parameter θ_2. This holds because there is no problem of weak identification.

To compare the reported finite-sample power of these tests with respect to a fixed benchmark we report in column 7 of the same table an estimate of the infeasible asymptotic power obtained from

$$P_{\theta^0}\left\{\chi_1^2\left(\text{ncp}=n\left(\theta_{10}-\theta_1^0\right)' G_1' V^{-1/2} N\left(V^{-1/2'} G_2\right) V^{-1/2'} G_1\left(\theta_{10}-\theta_1^0\right)\right)>\chi_{1,.95}^2\right\}. \qquad (4.4)$$

We note that the 2S-GMM score test does not perform as poorly as it did in the other experiments. The EEL Hybrid-2(3) score test performs the worst (unlike in the previous experiments).[25] The problem is compounded by the undesired decline in power for positive deviations from the truth. This problem was already noted by Kleibergen (2005) in a different context. However, his solution of using the so-called JKLM test in conjunction with the EEL score test cannot be used here because that requires an over-identified model whereas ours is just-identified. Finally we note that the performances of the EL Hybrid-2(3) score test, both with the EL and the 2S-GMM estimator of the nuisance parameter θ_2, are comparable. This indicates that not much is lost in terms of finite-sample size and power of the EL score test by plugging-in for the nuisance parameter the 2S-GMM estimator as opposed to the computationally more demanding EL estimator.

Although the size-distortions in this experiment are not as large as compared to the rest, there is still some upward size-distortion. In Table-5 we report the finite-sample size and power of the corresponding projection tests. The finite-sample size of these tests are much closer to the nominal level of 5% and the finite-sample power is not much lower than that of the corresponding plug-in tests (as was stated in Theorem-4.2(B)). This clearly demonstrates the benefits of using the projection test even when there is no problem of weak identification.

4.4 Monte-Carlo Experiment IV

4.4.1 Design

Here we focus mainly on weak identification of the sub-vector of interest θ_1 and do not allow θ_2 to be weakly identified. Under this scenario the conventional GEL plug-in tests have correct asymptotic size. This design follows Chaudhuri and Zivot (2011) where they study the projection-based tests for subsets of parameters in a linear instrumental variables regression. We draw i.i.d. copies of

[25] Since the EEL probabilities can be negative and thus cannot guarantee the positive (semi) definiteness of the variance estimator, finding the restricted estimator of θ_2 by minimizing the EEL objective function becomes a problem. To avoid this we used the shrinkage estimator of θ_2 by using the modified EEL implied probabilities proposed by Dovonon (2008). Asymptotically it should not matter.

$w_i = (y_i, X_{1i}, X_{2i}, Z_i')$ for $i = 1, \ldots, n$ from the following DGP-IV:

$$y_i = X_{1i}\theta_1^0 + X_{2i}\theta_2^0 + u_i,$$
$$X_{1i} = Z_i'\Pi_1 + \vartheta_{1i},$$
$$X_{2i} = Z_i'\Pi_2 + \vartheta_{2i},$$

where $u_i, \vartheta_{1i}, \vartheta_{2i}$ are jointly normal with mean zero, unit variance, $Cov(u_i, \vartheta_{1i}) = Cov(u_i, \vartheta_{2i}) = .8$ and $Cov(\vartheta_{1i}, \vartheta_{2i}) = .3$, and are independent of $Z_i \sim N(0_k, I_k)$. We define $\Pi_j := C_j/\sqrt{n}$ for $j = 1, 2$, where $C_1 = c_1 1_k$ is generated such that $\mu_1 := \Pi_1' \sum_{i=1}^n Z_i Z_i' \Pi_1/k$ is approx 4/3 or 20, and $C_2 = c_2 1_k$ is generated such that $\mu_2 := \Pi_2' \sum_{i=1}^n Z_i Z_i' \Pi_2/k$ is 20. These quantities, μ_1 and μ_2, are not exactly the elements of the so-called concentration matrix because the covariance between ϑ_1 and ϑ_2 has not been taken into account. However, individually μ_1 and μ_2 can be regarded as the concentration parameters (as in Monte-Carlo Experiment I) and with everything else same, the identification of θ_1 (θ_2) gets weaker as μ_1 (μ_2) decreases.

The true values $\theta_1^0 = .5$ and $\theta_2^0 = 1$ satisfy the moment restrictions in (2.1) for the moment vector

$$g_i(\theta) \equiv g(w_i, (\theta_1, \theta_2)) = Z_i(y_i - X_{1i}\theta_1 - X_{2i}\theta_2)$$

for $i = 1, \ldots, n$. The number of instruments is chosen as $k = 4$ and the sample size is chosen $n = 1000$.[26] Note that while we allow the number of moments (instruments) k to be as large as 16, it is still small relative to the sample size in this experiment.

4.4.2 Results

Based on 5000 Monte-Carlo trials, we report in Table-6 the empirical rejection rates of the true value $\theta = \theta^0 (= .5)$ for the EEL Hybrid-1, EEL Hybrid-3, EL Hybrid-1 and EL Hybrid-3 score tests (tests that are known to have correct asymptotic size under this scenario). As in Monte-Carlo Experiment 1, while estimating the asymptotic variance, we only assume that $Asym\,Var[\sqrt{n}\bar{g}_n(\theta^0)] = E[Z_i Z_i'(y_i - X_{1i}\theta_1^0 - X_{2i}\theta_2^0)^2]$ and not $E[Z_i Z_i'] \times Var[y_i - X_{1i}\theta_1^0 - X_{2i}\theta_2^0]$ although the latter is also true for DGP - IV. Hence the EEL Hybrid-1 version should not be confused with Kleibergen (2004)'s subset-K test.

While there is not much difference, in this design the Hybrid-3 version performs better than the Hybrid-1 version of the EEL score test. For EL score test also the rejection rate (both true and

[26]Results for other choices of k are available upon request. Unlike the other experiments, we do not present here the results for $n = 100$. We faced too many problems with the estimation of the restricted EL estimator of the nuisance parameter θ_2 to be reasonably certain about the reliability of the simulations. However, the results for EEL with $n = 100$ are available in Chaudhuri and Zivot (2011).

false) of the Hybrid-3 version is more than the Hybrid-1 version. Unlike in the other Monte-Carlo Experiments, here EEL dominates EL both in terms of (less) finite-sample size and (more) finite-sample power. In fact, the only case where EL (Hybrid-3 only) is more powerful than EEL is when the number of moment restrictions is relatively large (16) and θ_1 is weakly identified. However, EL also has a large upward size-distortion here and hence cannot be recommended. (As in Monte-Carlo Experiment I, simulation results for EL with many moments are less trustworthy because chances of making an error in computing the implied probabilities is really high.)

To assess the effectiveness of the $C(\alpha)$ form and the projection-GEL test, we report in Table-7 the finite-sample size and power of these tests based on EL for $k = 4$. These tests perform remarkably better than the conventional plug-in test (reported in column 3 of the same table).

A Appendix: Proofs

The proofs in this Appendix are direct applications of the results in Lemma-2.1 and Corollary-2.2. Hence we prove these two results elaborately.

Proof of Lemma 2.1:

(A) A mean-value expansion of the RHS of the (approximate) first-order condition of the maximization problem in (2.3) gives,

$$\begin{aligned} o_P\left(\frac{1}{\sqrt{n}}\right) &= \frac{1}{n}\sum_{i=1}^{n}\rho_1\left(\lambda'_{\rho,n}(\theta)g_i(\theta)\right)g_i(\theta) \\ &= \frac{1}{n}\sum_{i=1}^{n}\rho_1(0)g_i(\theta) + \frac{1}{n}\sum_{i=1}^{n}\rho_2(0)g_i(\theta)g_i'(\theta)\lambda_{\rho,n}(\theta) + R_{\lambda,n}(\theta) \\ &= -\bar{g}_n(\theta) - \bar{V}_n(\theta)\lambda_{\rho,n}(\theta) + R_{\lambda,n}(\theta), \end{aligned} \tag{A.1}$$

where $\bar{v}_i$ are the mean-values satisfying $|\bar{v}_i| \leq |\lambda'_{\rho,n}(\theta)g_i(\theta)|$ for all $i = 1, \ldots, n$, and the remainder term $R_{\lambda,n}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[\rho_2(\bar{v}_i) - \rho_2(0)\right]g_i(\theta)g_i'(\theta)\lambda_{\rho,n}(\theta)$. Ignoring the contribution of $R_{\lambda,n}(\theta)$ in (A.1), i.e., ignoring the term $\bar{V}_n^{-1}(\theta)R_{\lambda,n}(\theta)$ gives

$$\lambda_{\rho,n}(\theta) = -\bar{V}_n^{-1}(\theta)\bar{g}_n(\theta) + \bar{V}_n^{-1}(\theta) \times o_P\left(\frac{1}{\sqrt{n}}\right) = O_P\left(\frac{1}{\sqrt{n}}\right).$$

since $\bar{V}_n(\theta)$ and $\bar{V}_n^{-1}(\theta)$ are assumed to be $O_P(1)$ in (iv) and $\bar{g}_n(\theta) = O_P(n^{-1/2})$ by (ii) and (iii). Precisely for this reason, if we can show that $\|R_{\lambda,n}(\theta)\| = o_P(n^{-1/2})$ that will be sufficient for result (A). This is what we prove below.

$$\begin{aligned} \|R_{\lambda,n}(\theta)\| &= \left\|\left(\frac{1}{n}\sum_{i=1}^{n}\left[\rho_2(\bar{v}_i) - \rho_2(0)\right]g_i(\theta)g_i'(\theta)\right)\lambda_{\rho,n}(\theta)\right\| \\ &\leq \left\|\frac{1}{n}\sum_{i=1}^{n}\left[\rho_2(\bar{v}_i) - \rho_2(0)\right]g_i(\theta)g_i'(\theta)\right\| \times \|\lambda_{\rho,n}(\theta)\| \\ &\leq \max_{1\leq i\leq n}|\rho_2(\bar{v}_i) - \rho_2(0)| \times \left\|\frac{1}{n}\sum_{i=1}^{n}g_i(\theta)g_i'(\theta)\right\| \times \|\lambda_{\rho,n}(\theta)\| \text{ since } g_i(\theta)g_i'(\theta) \text{ psd} \\ &\leq b \times \max_{1\leq i\leq n}|\bar{v}_i| \times \|\bar{V}_n(\theta)\| \times \|\lambda_{\rho,n}(\theta)\| \\ &\leq b \times \max_{1\leq i\leq n}|g_i'(\theta)\lambda_{\rho,n}(\theta)| \times \gamma_{\max}(\theta) \times \|\lambda_{\rho,n}(\theta)\| \\ &\leq b \times \max_{1\leq i\leq n}\|g_i(\theta)\| \times b_{max} \times \|\lambda_{\rho,n}(\theta)\|^2 \text{ using } (iv), \\ &\leq b \times b_{\max} \times O_P(c_n) \times \|\lambda_{\rho,n}(\theta)\|^2 = O_P\left(c_n\|\lambda_{\rho,n}(\theta)\|^2\right) = o_P\left(n^{-1/2}\right), \end{aligned} \tag{A.2}$$

by repeated use of Cauchy-Schwartz and triangle inequalities and because $c_n = o_P(\sqrt{n})$ and $\|\lambda_{\rho,n}(\theta)\| = O_P(n^{-1/2})$. Therefore, (A.2) gives our desired result and hence the result (A).

(B) Expanding the numerator and denominator of the RHS of (2.4), and using the result obtained in (A) we get for any given $i = 1, \ldots, n$

$$
\begin{aligned}
\pi_{\rho,i,n}(\theta) &= \frac{\frac{1}{n}\left[\rho_1(0) + \rho_2(0)\lambda'_{\rho,n}(\theta)g_i(\theta) + \{\rho_2(\bar{v}_i) - \rho_2(0)\}\,\lambda'_{\rho,n}g_i(\theta)\right]}{\frac{1}{n}\sum_{j=1}^{n}\left[\rho_1(0) + \rho_2(0)\lambda'_{\rho,n}g_i(\theta) + \{\rho_2(\bar{v}_j) - \rho_2(0)\}\,\lambda'_{\rho,n}g_j(\theta)\right]} \\
&= \frac{\frac{1}{n}\left[\rho_1(0) - \rho_2(0)g'_i(\theta)\left\{\bar{V}_n^{-1}(\theta)\bar{g}_n(\theta) + o_P(n^{-1/2})\right\} + \{\rho_2(\bar{v}_i) - \rho_2(0)\}\,\lambda'_{\rho,n}(\theta)g_i(\theta)\right]}{\frac{1}{n}\sum_{j=1}^{n}\left[\rho_1(0) - \rho_2(0)g'_i(\theta)\left\{\bar{V}_n^{-1}(\theta)\bar{g}_n(\theta) + o_P(n^{-1/2})\right\} + \{\rho_2(\bar{v}_j) - \rho_2(0)\}\,\lambda'_{\rho,n}(\theta)g_j(\theta)\right]} \\
&= \frac{\frac{1}{n}\left[1 - (g_i(\theta) - \bar{g}_n(\theta))'\bar{V}_n^{-1}(\theta)\bar{g}_n(\theta)\right] + R_{NUM,i,n}}{1 - \bar{g}'_n(\theta)\bar{V}_n^{-1}(\theta)\bar{g}_n(\theta) + R_{DEN,n}} \qquad \text{(A.3)}
\end{aligned}
$$

where the remainder terms in the numerator and the denominator are respectively

$$
\begin{aligned}
R_{NUM,i,n} &:= \frac{1}{n}\{\rho_2(\bar{v}_i) - \rho_2(0)\}\,\lambda'_{\rho,n}(\theta)g_i(\theta) - \frac{1}{n}\rho_2(0)g'_i(\theta) \times o_P(n^{-1/2}) + \frac{1}{n}\bar{g}'_n(\theta)\bar{V}_n^{-1}(\theta)\bar{g}_n(\theta), \\
R_{DEN,n} &:= \frac{1}{n}\sum_{j=1}^{n}\left[\{\rho_2(\bar{v}_j) - \rho_2(0)\}\,\lambda'_{\rho,n}(\theta)g_j(\theta) - \rho_2(0)g'_i(\theta) \times o_P(n^{-1/2})\right].
\end{aligned}
$$

First note that i is given (fixed) in the remainder term $R_{NUM,i,n}$. Now exactly following the steps as in (A) to deal with the remainder term we obtain for a given $i = 1, \ldots, n$

$$
\begin{aligned}
|R_{NUM,i,n}| &\leq \frac{1}{n}|\rho_2(\bar{v}_i) - \rho_2(0)| \times \|\lambda_{\rho,n}(\theta)\| \times \|g_i(\theta)\| + \frac{1}{n}\|g_i(\theta)\| \times o_P\left(\frac{1}{\sqrt{n}}\right) + \frac{1}{n}\bar{g}'_n(\theta)\bar{V}_n^{-1}(\theta)\bar{g}_n(\theta) \\
&\leq \frac{1}{n}b \times |\lambda'_{\rho,n}(\theta)g_i(\theta)| \times \|\lambda_{\rho,n}(\theta)\| \times \|g_i(\theta)\| + \|g_i(\theta)\| \times o_P\left(\frac{1}{n^{3/2}}\right) + \frac{1}{n}\|\bar{g}_n(\theta)\|^2 \times \gamma_{min}^{-1}(\theta) \\
&\leq \frac{1}{n}b \times \|\lambda_{\rho,n}(\theta)\|^2 \times \|g_i(\theta)\|^2 + \|g_i(\theta)\| \times o_P\left(\frac{1}{n^{3/2}}\right) + \frac{1}{n}\|\bar{g}_n(\theta)\|^2 \times b_{min}^{-1} \text{ using } (iv), \\
&= O_P\left(\frac{1}{n^{1+1}}\right) \times O_P(1) + O_P(1) \times o_P\left(\frac{1}{n^{3/2}}\right) + O_P\left(\frac{1}{n^{1+1}}\right) \\
&= o_P\left(\frac{1}{n^{3/2}}\right) \qquad \text{(A.4)}
\end{aligned}
$$

because $\|g_i(\theta)\| = O_P(1)$ by (i), $\|\bar{g}_n(\theta)\| = O_P(n^{-1/2})$ by (ii) and (iii), and $\lambda_{\rho,n}(\theta) = O_P(n^{-1/2})$ by (A). We will use a similar technique to find the order of magnitude of $|R_{DEN,n}|$. However, there is an important difference: now we cannot work with a given $i = 1, \ldots, n$ and hence what will be important here is the order of magnitude of $c_n := \max_{1 \leq j \leq n}\|g_j(\theta)\|$ and not simply $\|g_i(\theta)\|$. So, proceeding as before, we obtain that

$$
\begin{aligned}
|R_{DEN,n}| &\leq \frac{1}{n}\left|\sum_{j=1}^{n}\{\rho_2(\bar{v}_j) - \rho_2(0)\}\,\lambda'_{\rho,n}(\theta)g_j(\theta)\right| + \|\bar{g}_n(\theta)\| \times o_P(n^{-1/2}) \\
&\leq \max_{1 \leq j \leq n}|\rho_2(\bar{v}_j) - \rho_2(0)| \times \|\bar{g}_n(\theta)\| \times \|\lambda_{\rho,n}(\theta)\| + \|\bar{g}_n(\theta)\| \times o_P(n^{-1/2}) \\
&\leq b \times \max_{1 \leq j \leq n}|\lambda'_{\rho,n}(\theta)g_j(\theta)| \times \|\bar{g}_n(\theta)\| \times \|\lambda_{\rho,n}(\theta)\| + \|\bar{g}_n(\theta)\| \times o_P(n^{-1/2}) \\
&\leq b \times \max_{1 \leq j \leq n}\|g_j(\theta)\| \times \|\bar{g}_n(\theta)\| \times \|\lambda_{\rho,n}(\theta)\|^2 + \|\bar{g}_n(\theta)\| \times o_P(n^{-1/2}) \\
&= o_P\left(n^{1/2-3/2}\right) + o_P(n^{-1}) = o_P(n^{-1})
\end{aligned}
$$

because $c_n := \max_{1\le j\le n}\|g_j(\theta)\| = o_P(\sqrt{n})$ by (i) and $\lambda_{\rho,n}(\theta) = O_P(n^{-1/2})$ by (A). Also, $\bar{g}_n'(\theta)\bar{V}_n^{-1}(\theta)\bar{g}_n(\theta)$ in the denominator of (A.3) is $O_P(n^{-1})$ because $\|\bar{g}_n(\theta)\| = O_P(n^{-1/2})$ by (ii) and (iii). Therefore, the entire denominator of (A.3) is $1+O_P(n^{-1})$. Hence the result (B) follows from (A.3) and (A.4). ■

Proof of Corollary 2.2:

(A) This follows directly from (vi) and the definition of the $\pi_{EEL,i,n}(\theta)$.

(B) Since our result in Lemma-2.1(B) is not uniform in $i = 1,\dots,n$ we cannot appeal to $\max_{1\le i\le n}|\pi_{\rho,i,n}(\theta) - \pi_{EEL,i,n}(\theta)|$ after applying Cauchy-Schwartz inequality. Instead, we directly work with the expression of the difference $\{\pi_{\rho,i,n}(\theta) - \pi_{EEL,i,n}(\theta)\} = R_{NUM,i}(= R_{NUM,i}/(1+o_P(1))$ as was obtained in (A.3). Also for notational simplicity, only in this proof, denote $\widetilde{Y}_{i,n} := Y_{i,n} - \mu_n$, $g_i := g_i(\theta)$, $\bar{g}_n := \bar{g}_n(\theta)$, $\bar{V}_n := \bar{V}_n(\theta)$ and $\lambda := \lambda_{\rho,n}(\theta)$. Accordingly, using Lemma-2.1(A), and (ii)-(vi), we obtain

$$
\begin{aligned}
& \left\|\sqrt{n}\sum_{i=1}^n \widetilde{Y}_{i,n}\left\{\pi_{\rho,i,n}(\theta) - \pi_{EEL,i,n}(\theta)\right\}\right\| \\
= & \left\|\sqrt{n}\sum_{i=1}^n \widetilde{Y}_{i,n}R_{NUM,i,n}\right\| \\
\le & \left\|\frac{1}{n}\sum_{i=1}^n\left\{\rho_2(\bar{v}_i) - \rho_2(0)\right\}\widetilde{Y}_{i,n}g_i'\right\| \times \sqrt{n}\|\lambda\| + o_P(1)\times\left\|\frac{1}{n}\sum_{i=1}^n\widetilde{Y}_{i,n}g_i'\right\| + \bar{g}_n'\bar{V}_n^{-1}\bar{g}_n\times\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\widetilde{Y}_{i,n}\right\| \\
\le & \sqrt{\frac{1}{n}\sum_{i=1}^n\left\{\rho_2(\bar{v}_i) - \rho_2(0)\right\}^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^n\left\|\widetilde{Y}_{i,n}g_i'\right\|^2}\times\sqrt{n}\,\|\lambda\| + o_P(1)\times\left\|\frac{1}{n}\sum_{i=1}^n\widetilde{Y}_{i,n}g_i'\right\| + \bar{g}_n'\bar{V}_n^{-1}\bar{g}_n\times\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\widetilde{Y}_{i,n}\right\| \\
\le & b\times\sqrt{\frac{1}{n}\sum_{i=1}^n|\lambda' g_i|^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^n\left\|\widetilde{Y}_{i,n}g_i'\right\|^2}\times\sqrt{n}\,\|\lambda\| + o_P(1)\times\left\|\frac{1}{n}\sum_{i=1}^n\widetilde{Y}_{i,n}g_i'\right\| + \bar{g}_n'\bar{V}_n^{-1}\bar{g}_n\times\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\widetilde{Y}_{i,n}\right\| \\
\le & b\times\sqrt{\|\lambda\|^2\frac{1}{n}\sum_{i=1}^n\|g_i\|^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^n\left\|\widetilde{Y}_{i,n}\right\|^2\|g_i'\|^2}\times\sqrt{n}\,\|\lambda\| + o_P(1)\times\left\|\frac{1}{n}\sum_{i=1}^n\widetilde{Y}_{i,n}g_i'\right\| + \bar{g}_n'\bar{V}_n^{-1}\bar{g}_n\times\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\widetilde{Y}_{i,n}\right\| \\
\le & b\times\left(\frac{1}{n}\sum_{i=1}^n\|g_i\|^2\right)^{\frac{1}{2}}\times\left(\frac{1}{n}\sum_{i=1}^n\|g_i\|^4\right)^{\frac{1}{4}}\times\left(\frac{1}{n}\sum_{i=1}^n\|\widetilde{Y}_{i,n}\|^4\right)^{\frac{1}{4}}\times\sqrt{n}\,\|\lambda\|^2 \\
& + o_P(1)\times\left\|\frac{1}{n}\sum_{i=1}^n\widetilde{Y}_{i,n}g_i'\right\| + \bar{g}_n'\bar{V}_n^{-1}\bar{g}_n\times\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\widetilde{Y}_{i,n}\right\| \\
= & O_P(1)\times O_P(1)\times O_P(1)\times O_P(n^{-1/2}) + o_P(1)\times O_P(1) + O_P(n^{-1})\times O_P(1) = o_P(1),
\end{aligned}
$$

from the standard arguments. ■

Lemma-A.1 lists some intermediate results that are useful in proving the lemmas and theorems stated in Sections 3 and 4 of the text.

Lemma A.1 *The following results hold as $n\to\infty$ under assumptions Θ, ID, S and ρ:*

(i.a) $\sqrt{n}\sum_{i=1}^{n}\pi_{GEL,i,n}(\theta)G_{lw,i}(\theta)=\widetilde{G}_{lwn}(\theta)+o_p(1)$ *and* $\sum_{i=1}^{n}\pi_i(\theta,\rho)G_{ls,i}(\theta)=\widetilde{G}_{lsn}(\theta)+o_p(1)$ *uniformly in* $\theta\in\Theta^n$ *for* $l=1,2$. *(*$\widetilde{G}_{lwn}(\theta)$ *and* $\widetilde{G}_{lsn}(\theta)$ *are defined in (i.b).)*

(i.b) For $\theta\in\Theta^n$ *and* $l=1,2$ *let* $\widetilde{G}_{lsn}(\theta)$ *be a* $k\times p_{ls}$ *matrix such that* $\widetilde{G}_{lsn}(\theta):=\bar{G}_{lsn}(\theta)$ *and* $\widetilde{G}_{lwn}(\theta)$ *be a* $k\times p_{lw}$ *matrix such that* $vec(\widetilde{G}_{lwn}(\theta)):=\sqrt{n}vec(\bar{G}_{lwn}(\theta))-\widetilde{V}_{lg}(\theta)\widetilde{V}_{gg}^{-1}(\theta)\sqrt{n}\bar{g}_n(\theta)$. *For any given sequence* $\theta_{ws^n}=((\theta_{1w}',\theta_{1s}^{n'}),(\theta_{2w}',\theta_{2s}^{n'}))'\in\Theta^n$ *such that for* $l=1,2$, $\theta_{ls}^n=\theta_{ls}^0+d_1/\sqrt{n}$ *and* $d_s=(d_1',d_2')'$, *the following hold:*

(b1) $vec(\widetilde{G}_{lwn}(\theta_{ws^n}))\xrightarrow{d}\Psi_{l.g}(\theta_{ws^0})+\left[\mathcal{G}_{lw}(\theta_{ws^0})d_s-V_{lg}(\theta_{ws^0})V_{gg}^{-1}(\theta_{ws^0})[\widetilde{m}(\theta_{ws^0})+M(\theta_s^0)d_s]\right]$ *where* $\Psi_{l.g}(\theta_{ws^0}):=\Psi_{lw}(\theta_{ws^0})-V_{lg}(\theta_{ws^0})V_{gg}^{-1}(\theta_{ws^0})\Psi_g(\theta_{ws^0})$ *is independent of* $\Psi_g(\theta_{ws^0})$,

(b2) $\widetilde{G}_{lsn}(\theta_{ws^n})\xrightarrow{P}M_l(\theta_s^0)$.

(ii) $\sum_{i=1}^{n}\pi_{GEL,i,n}(\theta)V_i(\theta)=\widetilde{V}_{gg}(\theta)+o_p(1)\xrightarrow{P}V_{gg}(\theta_{ws^0})$ *uniformly in* $\theta\in\Theta^n$.

Proof:

Point-wise convergence for any $\theta\in\Theta^n$ follows when we consider all these matrices column by column and then use Corollary 2.2. Uniformity in $\theta\in\Theta^n$ follows from our assumptions ID and S. ■

Proof of Theorem 3.1:

(A) For $\theta\in\Theta^n$, we know that

$$\begin{aligned}\sup_{\theta\in\Theta^n}\|\bar{g}_n(\theta)\| &\leq \sup_{\theta_w\in\Theta_w}\|\bar{g}_n(\theta_{ws^0})\|+\sup_{\theta\in\Theta^n}\|\bar{G}_{ns}(\theta)\|\times O_P(n^{-1/2})\\ &\leq \sup_{\theta_w\in\Theta_w}\frac{1}{\sqrt{n}}\{\|\Psi_{g,n}(\theta_{ws^0})\|+\|m_n^w(\theta_{ws^0})-m^w(\theta_{ws^0})\|+\|m^w(\theta_{ws^0})\|\}+O_P(n^{-1/2})\times\\ &\quad\times\sup_{\theta\in\Theta^n}\left\{\|\bar{G}_{ns}(\theta)-E[\bar{G}_{ns}(\theta)]\|+\|M_{ns}^w(\theta)-M_s^w(\theta)\|/\sqrt{n}+\|M_s^w(\theta)\|/\sqrt{n}+\|M(\theta_s)\|\right\}\\ &= \frac{1}{\sqrt{n}}\Big\{\{O_P(1)+o(1)+O_P(1)\}+\{o_p(1)+o_p(n^{-1/2})+O_P(n^{-1/2})+O_P(1)\}\Big\}\\ &= O_P(n^{-1/2})\end{aligned}\tag{A.5}$$

by using assumptions ID, S (iii) and (iv.a), and assumption S (ii.b) respectively. Therefore, the result follows directly from Lemma A.1(ii.a).

(B) For $\theta_0=\theta_{ws^n}\in\Theta^n$ (as defined in the statement of the Theorem) it follows from assumptions ID, S(ii.b) and (iii) that

$$\begin{aligned}\sqrt{n}\bar{g}_n(\theta_0) &= \sqrt{n}\bar{g}_n(\theta_{ws^0})+M(\theta_s^0)d_s+o_p(1)\\ &= \sqrt{n}\left(\bar{g}_n(\theta_{ws^0})-E[\bar{g}_n(\theta_{ws^0})]\right)+m^w(\theta_{ws^0})+M(\theta_s^0)d_s+o_p(1)\\ &\xrightarrow{d} \Psi_g(\theta_{ws^0})+[m^w(\theta_{ws^0})+M(\theta_s^0)d_s]=\widetilde{g}(\theta_w,d_s).\end{aligned}\tag{A.6}$$

The result then directly follows from Lemma A.1(ii.b) and assumption S(iv.b). ■

Proof of Lemma 4.1:

The first-order asymptotic behavior of $\mathfrak{LM}_{n,1.2}\left((\theta_{10}, \theta_{n2}^{EL}); \pi_{EL,i,n}(\theta_{10}, \theta_{n2}^{EL}), \kappa_{EL,i,n}(\theta_{10}, \theta_{n2}^{EL})\right)$ is given in Theorem 6 of GS-05. To see the consequences of replacing θ_{n2}^{EL} by $\widehat{\theta}_{n2}$, first note that under the assumptions of this lemma,

$$\sqrt{n}\left(\theta_{n2}^{EL} - \widehat{\theta}_{n2}\right) = o_P(1), \tag{A.7}$$

$$\sqrt{n}\left(\theta_{n2}^{EL} - \theta_2^0\right) = O_P(1) \tag{A.8}$$

(see, for e.g., Theorem 1 of Qin and Lawless (1994) and Lemma A1 of Stock and Wright (2000)). This implies that all the results in Lemma-2.1 and Corollary-2.2 hold irrespective of whether θ_{n2}^{EL} or $\widehat{\theta}_{n2}$ is used for θ_2 when $\theta_1 = \theta_{10}$ as specified by this lemma.[27]

For the sake of completeness, now let us consider the relevant $O_P(1)$ components of the statistic $\mathfrak{LM}_{n,1.2}\left((\theta_{10}, \widehat{\theta}_{n2}); \pi_{EL,i,n}(\theta_{10}, \widehat{\theta}_{n2}), \kappa_{EL,i,n}(\theta_{10}, \widehat{\theta}_{n2})\right)$ one by one. We show below that these components are asymptotically equivalent (up to $o_P(1)$) with the corresponding components of the statistic $\mathfrak{LM}_{n,1.2}\left((\theta_{10}, \theta_{n2}^{EL}); \pi_{EL,i,n}(\theta_{10}, \theta_{n2}^{EL}), \kappa_{EL,i,n}(\theta_{10}, \theta_{n2}^{EL})\right)$.

Guided by (A.8), consider any θ_{n2} such that $\sqrt{n}(\theta_{n2} - \theta_2^0) = O_P(1)$. The asymptotic equivalence is trivial for the weighted averages that are not scaled up (multiplied) by $\sqrt{n}$, i.e., for $\widetilde{G}_{1sn}(\theta_{10}, \theta_{n2})$, $\widetilde{G}_{2sn}(\theta_{10}, \theta_{n2})$, $\sum_{i=1}^{n} \pi_{EL,i,n}(\theta) V_i(\theta_{10}, \theta_{n2})$ and $\widetilde{V}_{1g}(\theta_{10}, \theta_{n2})\widetilde{V}_{gg}^{-1}(\theta_{10}, \theta_{n2})$ (the first three are defined in Lemma-A.1). This holds by virtue of boundedness and continuity of their non-random probability limits. So we focus on the sclaed up terms $\sqrt{n}\bar{g}_n(\theta_{10}, \theta_{n2})$ and $\sqrt{n}\widetilde{G}_{1sn}(\theta_{10}, \theta_{n2})$:

$$\sqrt{n}\bar{g}_n(\theta_{10}, \theta_{n2}) = \sqrt{n}\bar{g}_n(\theta_{10}, \theta_2^0) + M_{2s}(\theta_s^0)\underbrace{\sqrt{n}(\theta_{n2} - \theta_2^0)}_{\text{key term}} + o_P(1) \tag{A.9}$$

$$\begin{aligned}
\sqrt{n}\widetilde{G}_{1sn}(\theta_{10}, \theta_{n2}) &= \sqrt{n}vec(\bar{G}_{1wn}(\theta_{10}, \theta_{n2})) - \widetilde{V}_{1g}(\theta_{10}, \theta_2^0)\widetilde{V}_{gg}^{-1}(\theta_{10}, \theta_2^0)\sqrt{n}\bar{g}_n(\theta_{10}, \theta_{n2}) \\
&= \left[\sqrt{n}vec(\bar{G}_{lwn}(\theta_{10}, \theta_2^0)) - V_{1g}(\theta_{10}, \theta_2^0)V_{gg}^{-1}(\theta_{10}, \theta_2^0)\sqrt{n}\bar{g}_n(\theta_{10}, \theta_2^0)\right] \\
&\quad + \left[\mathcal{G}_{1w}^{(2)}(\theta_{10}, \theta_2^0) - V_{1g}(\theta_{10}, \theta_2^0)V_{gg}^{-1}(\theta_{10}, \theta_2^0)M_{2s}(\theta_s^0)\right]\underbrace{\sqrt{n}(\theta_{n2} - \theta_2^0)}_{\text{key term}} + o_P(1)
\end{aligned} \tag{A.10}$$

where $\mathcal{G}_{1w}^{(2)}(\theta_{10}, \theta_2^0)$ is a $kp_{1w} \times p_{2s}$ matrix containing the last p_{2s} columns of the $kp_{1w} \times p_s$ matrix $\mathcal{G}_{1w}(\theta_{10}, \theta_2^0)$. As can be seen from (A.9) and (A.10), both $\sqrt{n}\bar{g}_n(\theta_{10}, \theta_{n2})$ and $\sqrt{n}\widetilde{G}_{1sn}(\theta_{10}, \theta_{n2})$ remain asymptotically equivalent (up to $o_P(1)$) when θ_2 is replaced by θ_{n2}^{EL} or $\widehat{\theta}_{n2}$ once we note (A.7) and (A.8). ■

[27] The main reason behind this is that the remainder terms that are explicitly considered in Lemma-2.1 and Corollary-2.2 remain of the same order.

Remarks: Note that when, additionally, $p_{1w} = 0$, one can replace θ_{n2}^{EL} by any θ_{n2} such that $\sqrt{n}(\theta_{n2} - \theta_2^0) = O_P(1)$. This is because, in the absence of the term $\sqrt{n}\widetilde{G}_{1sn}(\theta_{10}, \theta_{n2})$, the only effect of this replacement is on $\sqrt{n}\bar{g}_n(\theta_{10}, \theta_{n2})$. However, by virtue of the $C(\alpha)$ form, the occurrence of $\sqrt{n}\bar{g}_n(\theta_{10}, \theta_{n2})$ is always (immediately) preceded by $N\left(\widetilde{V}_{gg}^{-1/2'}(\theta^0)M_{2s}(\theta^0)\right)\widetilde{V}_{gg}^{-1/2'}(\theta^0) + o_P(1)$. Hence the dependence of (A.9) on $\sqrt{n}(\theta_{n2} - \theta_2^0)$ vanishes up to order $o_P(1)$. This was the original justification behind Neyman's $C(\alpha)$ form.

Proof of Theorem 4.2: Unlike in the case of Theorem 3.1, here we provide more details for the proof because, to our knowledge, this proof for the GEL family has not yet appeared in any published papers.

(i) First, by taking $\theta_w = \theta_w^0$ and $d_s = 0$ i.e., for $\theta_{ws^n} = \theta^0$, we note from Lemma A.1(ii.b), (A.6) and Assumption S(iv.b) that

$$
\begin{aligned}
&\mathfrak{LM}_{n,1.2}\left(\theta^0; \pi^G(\theta^0), \pi^V(\theta^0)\right) \\
\xrightarrow{d}\ & \widetilde{g}'(\theta_w^0, 0)V_{gg}^{-1/2}(\theta^0)P\left(N\left(V_{gg}^{-1/2'}(\theta^0)\widetilde{G}_2(\theta^0)\right)V_{gg}^{-1/2'}(\theta^0)\widetilde{G}_1(\theta^0)\right)V_{gg}^{-1/2'}(\theta^0)\widetilde{g}(\theta_w^0, 0),
\end{aligned}
$$

where, by construction, $\widetilde{g}(\theta_w^0, 0) := \Psi_g(\theta^0)$ is independent of $\widetilde{G}_1(\theta^0)$ and $\widetilde{G}_2(\theta^0)$. Note that for $l = 1, 2$, $\widetilde{G}_l(\theta^0) := [\widetilde{\Psi}_{l.g}(\theta^0), M_l(\theta_s^0)]$ where $\widetilde{\Psi}_{l.g}(\theta^0)$ is a $k \times p_{l_w}$ matrix such that $vec(\widetilde{\Psi}_{l.g}(\theta^0)) = \Psi_{l.g}(\theta^0)$ and $Cov(\Psi_g(\theta^0), \Psi_{l.g}(\theta^0)) = 0$ by construction (element-by-element). Therefore, from Assumption S(ii), it follows that $\mathfrak{LM}_{n,1.2}\left(\theta^0; \pi^G(\theta^0), \pi^V(\theta^0)\right) \xrightarrow{d} \chi^2_{p_1}$ conditional on $\Psi_{1.g}(\theta^0)$ and $\Psi_{2.g}(\theta^0)$, and hence unconditionally. Hence, conditional on the event that $\mathcal{C}_{2n}(1-\tau, \theta_1^0)$ contains θ_2^0, $\inf_{\theta_2 \in \mathcal{C}_{2n}(1-\tau,\theta_1^0)} \mathfrak{LM}_{n,1.2}\left((\theta_1^0, \theta_2); \pi^G(\theta_1^0, \theta_2), \pi^V(\theta_1^0, \theta_2)\right) \le \mathfrak{LM}_{n,1.2}\left(\theta^0; \pi^G(\theta^0), \pi^V(\theta^0)\right)$ where the RHS has just been proved to converge in distribution to a central $\chi^2_{p_1}$ distribution. Under the condition of the theorem, the event that $\mathcal{C}_{2n}(1-\tau, \theta_1^0)$ contains θ_2^0 occurs with probability approaching $1-\tau$ (at least). Therefore, a simple application of the Bonferroni inequality gives that the asymptotic size of the projection-based GEL score test cannot exceed $1-(1-\tau)(1-\alpha) \le \tau + \alpha$.

(ii) If $p_w = 0$, i.e., if $\theta_1 = \theta_{1s}$ and $\theta_2 = \theta_{2s}$ then for any $\widetilde{\theta}_n := (\theta_{01}', \theta_{2s}^{n'})'$ where $\theta_{2s}^n = \theta_s^0 + d_2/\sqrt{n}$ for some $d_2 = O_P(1)$ (not required to be fixed), we know from Lemma A.1(ii.a), Assumption S(iv.b) and a mean-value expansion of $\sqrt{n}\bar{g}_n(\widetilde{\theta}_n)$ (around $(\theta_{01}', \theta_{2s}^{0'})'$ and using Assumption ID(ii)) that

$$\begin{aligned}
& \mathfrak{LM}_{n,1.2}\left(\widetilde{\theta}_n; \pi^G(\widetilde{\theta}_n), \pi^V(\widetilde{\theta}_n)\right) \\
= \; & n\left[\bar{g}_n(\theta_{01}, \theta_2^0) + M_2(\theta_s^0)d_2\right]' V_{gg}^{-1/2}(\theta^0) P\left(N\left(V_{gg}^{-1/2'}(\theta^0)M_2(\theta_s^0)\right) V_{gg}^{-1/2'}(\theta^0)M_1(\theta_s^0)\right) V_{gg}^{-1/2'}(\theta^0) \\
& \times \left[\bar{g}_n(\theta_{01}, \theta_2^0) + M_2(\theta_s^0)d_2\right] + o_p(1) \\
= \; & n\bar{g}_n'(\theta_{01}, \theta_2^0) V_{gg}^{-1/2}(\theta^0) P\left(N\left(V_{gg}^{-1/2'}(\theta^0)M_2(\theta_s^0)\right) V_{gg}^{-1/2'}(\theta^0)M_1(\theta_s^0)\right) V_{gg}^{-1/2'}(\theta^0)\bar{g}_n(\theta_{01}, \theta_2^0) + o_p(1)
\end{aligned} \quad \text{(A.11)}$$

because $N\left(V_{gg}^{-1/2'}(\theta^0)M_2(\theta_s^0)\right) V_{gg}^{-1/2'}(\theta^0)M_2(\theta_s^0)d_2 = 0$. The most important thing to note here is that $\sqrt{n}$-deviation (random or non-random) from the true value θ_2^0 of the nuisance parameter does not matter asymptotically, i.e., it does not matter asymptotically if we replace θ_2 by the true θ_2^0 or any $p_{2s} \times 1$ vector in its $\sqrt{n}$-neighborhood – a feature of the $C(\alpha)$ form of the statistic used in this paper. As a consequence, a test that replaces θ_2 by any $p_{2s} \times 1$ vector in the $\sqrt{n}$-neighborhood of θ_2^0 is asymptotically equivalent to the infeasible test (that uses the unknown θ_2^0) defined in the statement of the theorem.

Now since $\mathcal{C}_{2n}(1-\tau, \theta_{01}) \in \Theta_{1s}^n$ is assumed to be non-empty and inside Θ_2^n almost surely, by construction, $\theta_{2s}^{\inf}(\theta_{01}) := \arg\inf_{\theta_2 \in \mathcal{C}_{2n}(1-\tau,\theta_{01})} \mathfrak{LM}_{n,1.2}\left((\theta_{01}, \theta_2); \pi^G(\theta_{01}, \theta_2), \pi^V(\theta_{01}, \theta_2)\right)$ is inside Θ_{2s}^n almost surely. Hence, except in the negligible set (with probability 0), by definition of Θ_{2s}^n there exists a $\left\{d_{2s,n}^{\inf}(\theta_{01})\right\}_n = O_P(1)$ such that $\theta_{2s}^{\inf}(\theta_{01}) = \theta_{2s}^0 + d_{2s,n}^{\inf}(\theta_{01})/\sqrt{n}$. Therefore, the asymptotic equivalence with the infeasible test follows from (A.11). ■

References

Altonji, J. and Segal, L. M. (1996). Small Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business and Economic Statistics*, 14: 353–366.

Antoine, B., Bonnal, H., and Renault, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. *Journal of Econometrics*, 138: 461–487.

Caner, M. (2010). Exponential tilting with weak instruments: Estimation and testing. *Oxford Bulletin of Economics and Statistics*, 72: 307–326.

Chaudhuri, S. and Renault, E. (2011). Shrinkage of Variance for Minimum Distance based T ests. Technical report, University of North Carolina, Chapel Hill.

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

Chaudhuri, S. and Zivot, E. (2011). A new method of projection-based inference in GMM with weakly identified nuisance parameters. *Journal of Econometrics*, 164: 239–251.

Chen, L. and Guggenberger, P. (2011). On the Asymptotic Size of Subvector Tests in Linear Instrumental Variables Model. Technical report, University of California, San Diego.

Dovonon, P. (2008). Large Sample Properties of the Three-Step Euclidean Likelihood Estimators under Model Misspecification. CIREQ, University of Montreal, Barclays Wealth.

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

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

Dufour, J. M. and Taamouti, M. (2005a). Further Results on Projection-Based Inference in IV Regressions with Weak, Collinear or Missing Instruments. Discussion Paper.

Dufour, J. M. and Taamouti, M. (2005b). 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.

Fan, Y., Gentry, M., and Li, T. (2011). A New Class of Asymptotically Efficient Estimators for Moment Condition Models. *Journal of Econometrics*, 162: 268–277.

Guay, A. and Pelgrin, F. (2008). Using implied probabilities to improve estimation with unconditional moment restrictions. Technical report, UQAM, CIRPEE, CIREQ, HEC, IEMS, University of Lausanne.

Guggenberger, P. (2008). Finite sample evidence suggesting a heavy tail problem of the generalized empirical likelihood estimator. *Econometric Reviews*, 26: 526–541.

Guggenberger, P. and Hahn, J. (2005). Finite sample properties of the 2-step empirical likelihood estimator. *Econometric Reviews*, 24: 247–263.

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

Hall, A. R. (2000). Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test. *Econometrica*, 68: 1517–1527.

Hall, P. and Horowitz, J. (1996). Bootstrap Critical Values for Tests Based on Generalized Method of Moment Estimators. *Econometrica*, 64: 891–916.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50: 1029–1054.

Horowitz, J. (1998). Bootstrap Methods for Covariance Structures. *The Journal of Human Resources*, 33: 39–61.

Imbens, G. W., Spady, R. H., and Johnson, P. (1998). Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica*, 66: 333–357.

Kitamura, Y. (2001). Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, 69: 1661–1672.

Kitamura, Y. (2006). Empirical Likelihood Methods in Econometrics: Theory and Practice. Technical Report Discussion paper No. 1569, Cowles Foundation for Research in Economics, Yale University.

Kleibergen, F. (2002). Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, 70: 1781–1803.

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

Mittelhammer, R. C., Judge, G. G., and Schoenberg, R. (2005). Empirical evidence concerning the finite sample performance of EL-type structural equation estimation and inference methods. In Andrews, D. W. K. and Stock, J. H., editors, *Honor of Thomas Rothenberg*, pages 282–305. Cambridge University Press.

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

Newey, W. K. and 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., editor, *Probability and Statistics, the Harald Cramer Volume*, pages 313–334. Almqvist and Wiksell, Uppsala.

Qin, J. and Lawless, J. (1994). Empirical Likelihood and General Estimating Equations. *Annals of Statistics*, 22: 300–325.

Schennach, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *The Annals of Statistics*, 35: 634–672.

Smith, R. J. (1997). Alternative semi-parametric likelihood approaches to generalized method of moments estimation. *Economic Journal*, 107: 503–519.

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.

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. (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.

Specifications (n = 100)				Score tests						
				2S-GMM	EEL Hybrid			EL Hybrid		
k	ρ	Moments			1	2	3	1	2	3
		Skew	Strength							
2	0	high	none	9.7	9.0	14.4	14.5	14.4	14.5	9.7
2	0	high	weak	9.2	8.4	13.9	13.4	13.9	13.4	8.9
2	0	high	strong	9.9	9.4	13.4	12.8	13.4	12.8	9.9
2	.5	med	none	9.3	6.6	16.2	13.4	16.2	13.4	6.8
2	.5	med	weak	6.4	7.9	13.2	13.4	13.2	13.4	7.4
2	.5	med	strong	7.1	7.9	13.9	13.5	13.9	13.5	7.9
2	.9	low	none	13.9	5.6	19.7	10.1	19.7	10.1	5.7
2	.9	low	weak	9.4	6.4	15.0	10.2	15.0	10.2	6.4
2	.9	low	strong	5.2	5.1	10.0	9.3	10.0	9.3	5.4
4	0	high	none	13.3	11.4	24.0	21.9	23.2	21.8	12.5
4	0	high	weak	14.3	11.9	23.0	20.7	22.2	20.1	13.3
4	0	high	strong	17.9	15.4	22.8	19.6	22.8	20.4	16.6
4	.5	med	none	17.9	8.5	31.1	20.6	32.2	21.2	9.4
4	.5	med	weak	7.6	10.2	21.6	19.9	24.0	21.7	10.2
4	.5	med	strong	7.8	10.4	19.9	20.0	20.1	19.6	10.2
4	.9	low	none	39.4	5.7	51.0	14.4	49.5	14.5	7.2
4	.9	low	weak	18.0	5.5	30.9	14.2	30.5	14.3	6.5
4	.9	low	strong	6.8	5.6	16.9	14.1	16.6	13.2	5.8
8	0	high	none	16.8	13.8	36.8	34.8	36.6	33.8	17.5
8	0	high	weak	26.5	19.6	39.2	33.9	39.4	34.1	23.2
8	0	high	strong	36.0	27.4	40.8	32.6	40.3	32.2	29.0
8	.5	med	none	35.6	12.6	55.0	37.1	56.6	36.0	13.4
8	.5	med	weak	9.2	14.2	36.9	35.6	39.2	37.1	17.0
8	.5	med	strong	10.9	14.9	34.9	35.0	35.6	35.1	16.7
8	.9	low	none	79.8	6.1	84.3	26.5	84.2	27.3	7.7
8	.9	low	weak	38.0	5.8	60.4	25.9	60.5	26.4	7.7
8	.9	low	strong	9.9	5.1	32.3	25.0	31.5	24.3	6.8
16	0	high	none	24.9	17.2	58.1	55.2	25.5	29.6	29.5
16	0	high	weak	55.6	30.9	63.9	54.8	36.2	47.4	33.2
16	0	high	strong	72.4	46.8	69.2	58.7	51.9	61.9	40.6
16	.5	med	none	63.6	17.7	66.0	57.3	21.7	78.7	31.6
16	.5	med	weak	13.7	24.1	57.2	58.4	27	34.8	32.7
16	.5	med	strong	19.3	22.8	58.4	58.1	28.3	24.9	32.2
16	.9	low	none	99.3	7.1	69.7	50.7	11.9	99.6	20.5
16	.9	low	weak	70.8	5.2	70.4	50.0	10.4	83.5	18.2
16	.9	low	strong	17.2	5.4	56.6	48.9	10	34	17.3

Table 1: DGP-I: Finite-sample rejection rate (in %) of the true parameter value by score tests for $H_0 : \theta = \theta_0$ with nominal level 5%. Number of Monte-Carlo trials is 5000.

Specifications (n = 1000)				Score tests						
				2S-GMM	EEL Hybrid			EL Hybrid		
k	ρ	Moments			1	2	3	1	2	3
		Skew	Strength							
2	0	high	none	5.4	5.1	6.9	6.5	5.9	6.4	6.6
2	0	high	weak	5.7	5.5	6.8	6.6	5.6	6.6	6.4
2	0	high	strong	5.5	5.5	5.6	5.5	5.1	5.8	5.6
2	.5	med	none	7.9	5.4	8.8	6.4	4.9	8.4	5.8
2	.5	med	weak	6.3	6.3	7.9	7.1	5.5	7.2	6.4
2	.5	med	strong	4.8	5.3	5.6	5.8	5.4	5.8	5.9
2	.9	low	none	13.5	5.1	14.3	5.6	5.0	13.9	5.4
2	.9	low	weak	8.6	5.1	9.4	5.5	4.9	8.7	5.3
2	.9	low	strong	4.6	4.7	5.2	5.1	4.7	4.9	5.1
4	0	high	none	6.1	5.8	7.9	7.7	6.1	8.2	7.9
4	0	high	weak	6.6	6.5	7.7	7.4	6.1	7.8	7.8
4	0	high	strong	7.1	7.0	6.9	6.9	6.9	6.8	6.6
4	.5	med	none	14.8	5.9	17.7	7.4	6.4	17.5	7.8
4	.5	med	weak	7.3	6.9	11.6	7.9	6.1	10.8	7.3
4	.5	med	strong	5.1	5.9	7.4	6.8	6.0	6.7	6.4
4	.9	low	none	35.5	4.7	37.0	5.4	4.8	37.3	5.5
4	.9	low	weak	16.3	5.1	18.5	5.8	4.8	18.2	5.5
4	.9	low	strong	6.5	5.3	7.7	6.0	5.6	7.9	6.3
8	0	high	none	6.3	6.4	10.1	10.3	6.5	11.0	10.4
8	0	high	weak	8.6	8.3	10.3	10.0	8.4	9.7	9.3
8	0	high	strong	11.0	10.5	8.6	8.3	10.8	8.9	8.6
8	.5	med	none	28.5	6.7	34.9	9.9	6.3	33.8	8.9
8	.5	med	weak	10.3	7.9	19.6	10.4	8.5	20.3	10.3
8	.5	med	strong	5.0	7.8	10.8	9.0	7.8	10.6	8.9
8	.9	low	none	75.0	5.1	77.0	6.6	4.8	76.7	6.2
8	.9	low	weak	37.8	6.0	42.1	7.4	5.6	40.8	7.1
8	.9	low	strong	9.7	5.5	12.6	7.0	5.6	12.9	7.2
16	0	high	none	9.1	8.6	18.5	17.9	9.3	8.5	8.7
16	0	high	weak	16.9	16.1	16.4	15.3	16.2	8.4	8.2
16	0	high	strong	24.3	22.8	15.0	13.1	24.2	8.3	7.8
16	.5	med	none	54.0	8.1	64.4	14.3	9.0	54.6	9.1
16	.5	med	weak	14.4	12.6	37.5	15.1	11.6	25.7	8.1
16	.5	med	strong	6.4	13.6	17.1	13.7	13.6	7.7	7.1
16	.9	low	none	98.5	5.0	98.8	8.0	6.2	98.5	6.1
16	.9	low	weak	70.0	5.6	76.2	8.0	5.7	71.2	5.4
16	.9	low	strong	14.2	4.8	22.0	8.0	5.7	17.1	5.9

Table 2: DGP-I: Finite-sample rejection rate (in %) of the true parameter value by score tests for $H_0 : \theta = \theta_0$ with nominal level 5%. Number of Monte-Carlo trials is 5000.

Specifications		Score tests						
		2S-GMM	EEL			EL		
n	k		Hybrid-1	Hybrid-2	Hybrid-3	Hybrid-1	Hybrid-2	Hybrid-3
100	3	15.8	18.5	25.5	20.5	21.9	15.4	13.6
100	10	31.8	25.8	54	42.5	41.3	35.3	25.3
200	3	13.2	16.2	20.2	15.9	18.5	12.3	10.5
200	10	28.3	39	46.1	32.9	46.7	23.1	19.1
500	3	9.8	12.4	14.4	11.8	13.2	9	7.7
500	10	21.9	38.7	34.3	23.6	41.4	16.6	13.6
1000	3	8.8	10.8	11.8	10.1	11.3	7.8	7
1000	10	18.9	33.9	27.9	19.2	35.4	12.6	10.9

Table 3: DGP-II: Finite-sample rejection rate (in %) of the true parameter value by score tests for $H_0 : \theta = \theta_0$ with nominal level 5%. Number of Monte-Carlo trials is 5000.

n	$\theta_{10} - \theta_1^0$	Score tests				Infeasible estimated Asym. Power
		2S-GMM	EEL Hybrid-2 (3)	EL Hybrid-2 (3)		
				(EL θ_2)	(2S-GMM θ_2)	
100	-1	99.7	97.5	-	99.5	99.9
100	-0.8	98.6	92.5	-	97.6	97.9
100	-0.6	93.4	78.4	-	91.2	85.1
100	-0.4	76.2	50.9	-	72.6	51.6
100	-0.2	42.1	20.3	-	39.1	17.0
100	0	10.6	7.8	11.9	13.0	5
100	0.2	6.6	21.8	-	26.1	17.0
100	0.4	34.4	57.4	-	68.4	51.6
100	0.6	74.4	76.4	-	94.7	85.1
100	0.8	92.3	62.5	-	99.8	97.9
100	1	96.2	37.8	-	100.0	99.9
1000	-0.3162	99.3	99.3	96.6	97.1	99.9
1000	-0.253	96.9	97.6	90.4	91.6	97.9
1000	-0.1897	86.9	91.2	74.9	76.4	85.1
1000	-0.1265	60.5	73.1	46.7	48.1	51.6
1000	-0.0632	25.6	39.5	18.0	18.9	17.0
1000	0	6.2	11.9	6.4	6.8	5
1000	0.0632	11.4	15.8	20.4	21.3	17.0
1000	0.1265	45.1	33.2	59.9	61.1	51.6
1000	0.1897	85.6	24.8	91.6	92.0	85.1
1000	0.253	98.5	8.9	99.5	99.5	97.9
1000	0.3162	100.0	1.5	100.0	100.0	99.9

Table 4: DGP-III: Finite-sample rejection rate (in %) of score tests for $H_{10} : \theta_1 = \theta_{10}$ with nominal level 5%. In columns 5 and 6 we plug-in $\widetilde{\theta}_{2n}(\theta_{10})$ obtained from EL and 2S-GMM respectively in the EL Hybrid-2 score statistic. Number of Monte-Carlo trials is 5000.

nominal level		Plug-in vs Projection-based score tests with 2S-GMM $(1-\tau)$ CI								
$\alpha = 5\%$		2S-GMM			EEL Hybrid-2 (3)			EL Hybrid-2 (3)		
n	$\theta_{10} - \theta_1^0$	plug-in	$\tau = 5\%$	$\tau = 1\%$	plug-in	$\tau = 5\%$	$\tau = 1\%$	plug-in	$\tau = 5\%$	$\tau = 1\%$
100	-1	99.7	99.3	98.4	97.5	96.3	95.2	99.5	99.4	98.9
100	-0.8	98.6	96.9	93.7	92.5	89.5	87.0	97.6	96.7	95.2
100	-0.6	93.4	87.7	79.2	78.4	72.3	67.3	91.2	88.7	83.5
100	-0.4	76.2	64.0	50.4	50.9	42.9	37.6	72.6	64.9	55.4
100	-0.2	42.1	28.6	18.1	20.3	15.2	12.2	39.1	29.3	21.3
100	0	10.6	6.1	2.9	7.8	5.6	4.6	13.0	7.7	5.4
100	0.2	6.6	6.1	6.0	21.8	18.6	17.9	26.1	20.5	20.1
100	0.4	34.4	34.4	34.4	57.4	50.8	48.5	68.4	63.8	63.6
100	0.6	74.4	74.4	74.4	76.4	66.5	61.6	94.7	93.4	93.4
100	0.8	92.3	92.3	92.3	62.5	49.0	43.1	99.8	99.8	99.8
100	1	96.2	96.2	96.2	37.8	23.2	18.5	100.0	100.0	100.0
1000	-0.3162	99.3	99.3	99.1	99.3	97.6	92.9	97.1	96.1	94.9
1000	-0.253	96.9	96.3	95.7	97.6	93.4	84.7	91.6	88.1	86.1
1000	-0.1897	86.9	84.9	83.1	91.2	81.5	68.4	76.4	71.1	67.1
1000	-0.1265	60.5	57.3	53.8	73.1	56.9	41.4	48.1	42.3	37.5
1000	-0.0632	25.6	23.2	20.6	39.5	24.7	14.8	18.9	14.9	12.4
1000	0	6.2	5.6	4.8	11.9	6.1	3.1	6.8	4.8	4.0
1000	0.0632	11.4	11.4	11.3	15.8	7.4	5.8	21.3	18.8	18.8
1000	0.1265	45.1	45.1	45.1	33.2	16.0	13.0	61.1	58.2	58.2
1000	0.1897	85.6	85.6	85.6	24.8	13.3	11.1	92.0	90.8	90.8
1000	0.253	98.5	98.5	98.5	8.9	5.1	4.3	99.5	99.4	99.4
1000	0.3162	100.0	100.0	100.0	1.5	1.0	0.8	100.0	100.0	100.0

Table 5: DGP-III: Finite-sample rejection rate (in %) of Projection-based score tests for $H_{10} : \theta_1 = \theta_{10}$ with nominal level $\alpha = 5\%$. Asymptotic size cannot exceed $\alpha + \tau$. Number of Monte-Carlo trials is 5000. (Since we could not reliably find the power when $n = 100$ by plugging-in the EL estimator of θ_2, the results for plug-in-EL are with the 2S-GMM estimator plugged-in.)

$\theta_{10}-\theta_1^0$	$k=2$				$k=4$				$k=8$				$k=16$			
	EL		EEL		EL		EEL		EL		EEL		EL		EEL	
	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3
-5	24	25	29	29	25	29	34	35	21	28	30	31	19	32	22	24
-4.5	23	24	28	28	25	29	34	34	21	28	29	30	19	32	21	23
-4	23	24	27	27	25	28	33	34	20	28	28	29	19	31	21	22
-3.5	22	23	25	25	24	27	32	32	20	27	27	28	18	31	21	22
-3	21	22	24	24	23	26	30	31	19	26	27	27	18	31	20	21
-2.5	20	20	22	22	22	25	28	29	19	25	25	26	18	31	18	20
-2	18	19	20	20	20	23	26	26	18	24	23	23	17	30	17	18
-1.5	16	16	17	18	17	20	22	22	16	22	20	21	17	28	15	15
-1	12	13	14	14	14	16	17	18	13	18	16	16	15	26	12	14
-0.5	8	9	8	8	9	11	9	9	10	14	10	10	13	23	7	8
0	5	5	5	5	6	8	4	4	7	10	5	5	11	19	5	5
0.5	18	19	24	24	20	23	26	26	17	23	24	25	17	29	17	18
1	61	61	75	75	39	44	50	51	26	36	37	38	21	38	23	25
1.5	63	65	76	76	37	43	44	45	23	32	30	30	18	34	17	18
2	56	57	69	69	38	44	52	53	25	34	37	39	22	37	24	26
2.5	50	52	63	63	39	45	53	54	26	35	42	43	21	37	29	31
3	47	48	59	59	39	44	54	54	27	35	43	44	21	37	30	32
3.5	44	45	55	55	38	43	53	53	27	35	43	44	21	37	32	33
4	42	44	53	53	37	42	52	53	27	35	44	45	21	37	31	33
4.5	41	42	51	51	36	41	51	52	27	35	44	45	21	37	31	33
5	40	41	49	49	36	41	51	51	27	35	44	44	21	37	31	32
-0.1	64	65	74	74	82	86	93	93	83	90	96	96	76	92	97	98
-0.09	55	57	65	65	75	79	86	86	76	84	92	92	70	88	94	94
-0.08	47	48	55	56	66	70	77	77	67	77	84	84	62	82	88	89
-0.07	38	40	45	45	55	60	66	66	57	67	74	75	53	74	79	80
-0.06	30	31	36	36	43	49	54	54	46	56	59	60	43	64	65	65
-0.05	23	24	26	26	33	38	41	41	34	44	47	47	34	53	49	51
-0.04	16	17	18	18	23	26	29	29	24	32	33	34	25	42	34	35
-0.03	12	12	12	13	15	18	18	19	17	23	22	22	18	31	20	22
-0.02	8	9	8	8	10	12	10	10	11	15	11	12	12	22	11	11
-0.01	6	6	6	6	6	8	5	5	8	12	6	6	10	17	5	6
0	5	5	5	5	6	7	4	4	6	10	4	4	8	15	3	4
0.01	6	6	5	5	6	8	5	5	8	11	5	6	9	17	5	6
0.02	8	9	6	6	10	12	9	10	11	15	11	11	12	22	10	11
0.03	12	13	13	13	15	18	18	18	16	23	20	20	17	30	19	21
0.04	17	18	23	23	23	27	29	30	25	32	33	33	25	41	33	35
0.05	24	26	31	32	34	38	42	43	34	43	47	48	33	52	48	50
0.06	34	35	45	45	45	50	57	57	45	56	62	62	43	63	65	66
0.07	44	46	56	56	56	62	70	70	57	68	75	75	53	73	78	79
0.08	55	57	67	68	69	73	81	81	68	78	85	86	62	82	88	89
0.09	66	68	80	80	79	82	91	91	77	86	93	94	70	89	95	96
0.1	75	77	87	88	86	89	96	96	85	92	98	98	77	93	98	98

Table 6: DGP-IV: Finite-sample rejection rate (in %) of the EEL and EL Hybrid-1 (respective 1st columns) and Hybrid-3 (respective 2nd columns) score tests (GS-05) for $H_{10}:\theta_1=\theta_{10}$ with nominal level 5%. Number of Monte-Carlo trials is 5000 and identification. In the upper panel the moment conditions are weak for θ_1 and the in the lower panel they are strong.

Specifications ($n=1000, k=4$)		EL Hybrid-3 tests using plug-in and projection for nuisance parameter θ_2			
		plug-in $\widetilde{\theta}_{2n}(\theta_{10})$ from			
Moments for θ_1	$\theta_{10}-\theta_1^0$	EL estimator	2S-GMM estimator	$(1-\tau)$ 2S-GMM confidence region for θ_2	
		$\alpha=5\%$	$\alpha=5\%$	$\alpha=4.5\%, \tau=.5\%$	$\alpha=5\%, \tau=5\%$
weak	-5	29	34	31	33
weak	-4.5	29	34	31	33
weak	-4	28	33	30	32
weak	-3.5	27	32	29	31
weak	-3	26	30	28	30
weak	-2.5	25	28	26	27
weak	-2	23	26	24	25
weak	-1.5	20	22	21	22
weak	-1	16	18	15	17
weak	-0.5	11	9	8	9
weak	0	8	4	3	3
weak	0.5	23	26	24	25
weak	1	44	50	43	46
weak	1.5	43	44	32	36
weak	2	44	52	43	47
weak	2.5	45	53	46	50
weak	3	44	53	47	51
weak	3.5	43	53	47	50
weak	4	42	52	46	49
weak	4.5	41	51	45	49
weak	5	41	51	45	48
strong	-0.1	86	89	87	89
strong	-0.09	79	82	80	82
strong	-0.08	70	73	70	72
strong	-0.07	60	61	58	60
strong	-0.06	49	50	47	49
strong	-0.05	38	37	35	37
strong	-0.04	26	26	24	26
strong	-0.03	18	17	16	17
strong	-0.02	12	11	9	10
strong	-0.01	8	7	5	6
strong	0	7	5	4	5
strong	0.01	8	7	6	6
strong	0.02	12	10	9	10
strong	0.03	18	17	15	17
strong	0.04	27	27	24	26
strong	0.05	38	39	37	39
strong	0.06	50	52	50	52
strong	0.07	62	66	63	65
strong	0.08	73	77	75	77
strong	0.09	82	86	85	86
strong	0.1	89	92	91	92

Table 7: DGP-IV: Finite-sample rejection rate (in %) of nominal-level-α EEL Hybrid-3 score test for $H_{10}: \theta_1 = \theta_{10}$ using as plug-in the estimator of θ_2 obtained by EL (in column 3) and 2S-GMM (in column 4), and the $(\tau+\alpha)$ projection-based EL Hybrid-3 score test using a 2S-GMM confidence region of θ_2. Number of Monte-Carlo trials is 5000.