A New Class of Tests for Overidentifying Restrictions in Moment Condition Models

Wang, Xuexin

The Wang Yanan Institute for Studies in Economics, Xiamen University, Fujian, China.

24 January 2016

A New Class of Tests for Overidentifying Restrictions in Moment Condition Models

Xuexin Wang*

Xiamen University

January 21, 2016

Abstract

In this paper, we propose a new class of tests for overidentifying restrictions in moment condition models. The tests in this new class are quite easy to compute. They avoid the complicated saddle point problem in generalized empirical likelihood (GEL) estimation, only a $\sqrt{n}$ consistent estimator, where n is the sample size, is needed. In addition to discussing their first-order properties, we establish that under some regularity conditions these tests share the same higher order properties as GEL overidentifying tests, given proper consistent estimators. Monte Carlo simulation study shows that the new class of tests of overidentifying restrictions has better finite sample performance than the two-step GMM overidentification test, and compares well to several potential alternatives in terms of overall performance.

JEL Classification: C12, C22

Keywords: Generalized Empirical Likelihood (GEL); Tests for Overidentifying Restrictions; $C(\alpha)$ Type Tests; High Order Equivalence;

*The Wang Yanan Institute for Studies in Economics, Xiamen University, Fujian, China. xuexinwang@xmu.edu.cn. Research support from National Natural Science Funds of China (Grant No.71301136) is gratefully acknowledged.

1 Introduction

The generalized method of moments (GMM), initially developed by Hansen (1982), provides a unifying econometric framework, nesting a lot of econometric methods such as maximum likelihood (ML), ordinary least squares and two-stage least squares in instrumental regressions. It generalizes the traditional method of moments in the sense that it allows the number of moments functions to be larger than the number of unknown parameters, which is prevalent in econometrics. Its applications include, but not limited to, rational expectation models, panel data models, continuous models and semiparametric models.

The importance and usefulness of GMM mandates accurate estimation and inference procedures. Hansen (1982) proposes a two-step GMM procedure. The basic idea is to minimize the criterion function of a quadratic form of sample average of the moment functions with an optimal weighting matrix. Since the optimal weighting matrix depends on the unknown parameters, Hansen (1982) suggests using initial, possibly inefficient, estimates to estimate this optimal weighting matrix. He also suggests a test for overidentifying restrictions, the famous J or Sargan test, based on the value of the quadratic criterion function evaluated at the two-step GMM estimator, and shows that the J test follows, under standard regularity conditions, a chi-squared distribution with degrees of freedom equal to the number of overidentifying restrictions asymptotically under the null hypothesis that the moment restrictions hold.

Hansen (1982)'s two-step GMM procedure is relatively easy to implement, hence quite popular in practice. However theoretical analysis and Monte Carlo evidences have shown that two-step GMM estimators may be badly biased in finite samples and the first-order asymptotic theory often provides poor approximation to the distribution of test statistics based on it. For example, Newey and Smith (2004) establish the high order properties of two-step GMM estimators theoretically, Hansen, Heaton, and Yaron (1996) find that the J test is too large, leading to overrejection of the asset-pricing model they study, when asymptotic critical values are used. See also other papers in Special

Issue of the Journal of Business and Economic Statistics (July 1996). Because of this, a lot of efforts have been done to improve the finite sample properties of two-step GMM estimators and test statistics. One approach is to employ Bootstrapping methods, see Hall and Horowitz (1996) and Brown and Newey (2002). Another approach is to employ alternative criterion functions to obtain parameter estimators and derive overidentification test statistics. These include estimators and test statistics based on the empirical likelihood (EL) of Owen (1988, 1990), Qin and Lawless (1994), and Imbens (1997), the continuous-updating GMM of Hansen et al. (1996), and the exponential tilting (ET) of Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998). Newey and Smith (2004) show that these estimators share a common structure, being members of a class of generalized empirical likelihood (GEL) estimators. They also show that GEL estimators may be less prone to bias than two-step GMM estimator.

Despite the nice theoretical properties of GEL, it is unclear whether the higher order advantages of GEL over GMM estimators translate into improved finite sample performance and it has been rarely applied in empirical applications. This may be attributed to the computational difficulty arising from the saddle point characterization of GEL. This saddle point problem can be solved through an inner-loop and out-loop optimization algorithm, as discussed in Kitamura (2007). While the inner-loop optimization with respect to the auxiliary parameters is usually a well defined convex optimization problem, the outer-loop optimization is generally complicated because of its highly nonlinear nature. Several papers have focused on how to overcome the computational burden to obtain GMM estimators as efficient as GEL estimators, see, for example, Antoine, Bonnal, and Renault (2007), and Fan, Gentry and Li (2011).

In this paper, we focus on the inference of GEL. As its name implies, GEL bears a lot of similarities with the classical maximum likelihood methods, allowing to construct the likelihood ratio type tests, Lagrange multiplier (LM) type tests and score type tests for overidentifying restrictions, see Imbens et al. (1998), Imbens (2002), and Smith (1997, 2011). Imbens et al. (1998) find that particular GEL tests for overidentifying restrictions, especially ET, possess actual sizes closer to nominal size than the J test, although still oversized in finite samples. However, to construct these tests, we have to firstly compute the GEL estimators, which is still an expensive or infeasible task in real applications. In terms of alleviating the computational burden of parameter estimation, we notice that $C(\alpha)$ tests for composite hypotheses are proposed by Neyman (1959) in ML. Smith (1987) extends this idea for implicit function restrictions in ML. While $C(\alpha)$ tests rely on the score of log-likelihood, Wooldridge (1990) develops, in the scenario of conditional moment tests, new statistics based on the score of the conditional moment restrictions. Wang (2015) proposes new conditional moment tests based on projections, generalizing Wooldridge's idea into GMM context. In the event of GEL, $C(\alpha)$ type tests for overidentifying restrictions should be very useful. Aiming at this, we propose such tests in this direction. Like traditional $C(\alpha)$ type tests, the tests in this new class are quite easy to compute. They avoid the complicated saddle point problem of GEL, only a $\sqrt{n}$ consistent estimator, where n is the sample size, is needed. In addition to discussing their first-order properties, we establish that under some regularity conditions these tests share the same higher order properties as GEL overidentification tests, given some proper consistent estimators. Monte Carlo simulation study shows that the new class of tests of overidentifying restrictions has better finite sample performance than the two-step GMM overidentification test, and compares well to several potential alternatives in terms of overall performance.

The organization of the paper is as follows. In section 2, we discuss the preliminaries and give a review on tests for overidentifying restrictions in GMM and GEL framework. We introduce the new class of tests for overidentifying restrictions, and discuss their asymptotic properties in section 3. Monte Carlo simulations are conducted in section 4. Section 5 concludes.

2 Preliminaries

The model we consider is the one with a finite number of moment restrictions. Following the setup of Smith and Newey (2004), let z_i $(i = 1, ..., n)$ be i.i.d. observations on a data vector z. Also, let β be a $p\times 1$ parameter vector and $g(z,\beta)$ be an $m\times 1$ vector of functions of the data observation z and the parameter β. We consider the overidentified case, e.g., $m>p$. The model has the true parameter $\beta_0\in B\in\mathbb{R}^p$ satisfying the moment conditions

$$E\left[g(z,\beta_0)\right]=0 \text{ a.s.}$$

The null hypothesis we are interested in is

$$H_0: E\left[g(z,\beta_0)\right]=0 \text{ a.s., for } \beta_0\in B. \tag{1}$$

The alternative hypothesis is

$$H_0: \Pr\left(E\left[g(z,\beta)\right]=0\right)<1 \text{ a.s., for all } \beta\in B.$$

Let $g_i(\beta)=g(z_i,\beta)$, $\hat{g}(\beta)=n^{-1}\sum\limits_{i=1}^{n}g_i(\beta)$, and $\hat{\Omega}(\beta)=n^{-1}\sum\limits_{i=1}^{n}g_i(\beta)\,g_i(\beta)'$; Let $\tilde{\beta}$ be some preliminary estimator, for example, $\tilde{\beta}=\arg\min_{\beta\in B}\ \hat{g}(\beta)'\,\hat{W}^{-1}\hat{g}(\beta)$, where B denotes the parameter space, and $\hat{W}$ is a proper weighting matrix, normally identity matrix. The two-step GMM estimator $\hat{\beta}_{GMM}$ is obtained by minimizing the following criterion function

$$\min_{\beta\in B}\hat{g}(\beta)'\,\hat{\Omega}\left(\tilde{\beta}\right)^{-1}\hat{g}(\beta)\,.$$

Let $G_i(\beta)=\partial g_i(\beta)/\partial\beta'$, $\hat{G}(\beta)=\frac{1}{n}\sum\limits_{i=1}^{n}\partial g_i(\beta)/\partial\beta'$. The first-order condition is

$$\hat{G}\left(\hat{\beta}_{GMM}\right)'\hat{\Omega}\left(\tilde{\beta}\right)^{-1}\hat{g}\left(\hat{\beta}_{GMM}\right)=0.$$

The corresponding J test is

$$\hat{S}_{GMM}=\hat{S}(\hat{\beta}_{GMM})=n\hat{g}\left(\hat{\beta}_{GMM}\right)'\hat{\Omega}\left(\tilde{\beta}\right)^{-1}\hat{g}\left(\hat{\beta}_{GMM}\right).$$

Hansen (1982) shows that J test follows, under standard regularity conditions, a chi-squared distribution with degrees of freedom equal to the number of overidentifying restrictions asymptotically under the null. However there are increasing simulation evidences indicating that the two-step GMM estimator may be severely biased and the J test tends to be oversized in small samples. Because of this, a number of alternative estimation and inference approaches have been proposed in various forms. Hansen et al. (1996) propose the continuous-updating estimator (CUE), which is defined as the solution to a minimization problem as follows

$$\hat{\beta}_{CUE} = \arg\min_{\beta \in B} \hat{g}\left(\beta\right)' \hat{\Omega}\left(\beta\right)^{-} \hat{g}\left(\beta\right),$$

where A^{-} denotes any generalized inverse of a matrix A, satisfying $AA^{-}A = A$. The difference between continuous-updating criterion function and two-step criterion function is that the weighting function is not fixed in the case of CUE. This makes the first order conditions for this minimization problem more complicated. The first-order condition right now is

$$\hat{C}\left(\hat{\beta}_{CUE}\right)' \hat{\Omega}\left(\hat{\beta}_{CUE}\right)^{-1} \hat{g}\left(\hat{\beta}_{CUE}\right) = 0,$$

where

$$\hat{C}\left(\beta\right) = \hat{G}\left(\beta\right) - \frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}\left(\beta\right)' \hat{\Omega}\left(\beta\right)^{-1} g_i\left(\beta\right)\right] G_i\left(\beta\right). \tag{2}$$

Given $\hat{\beta}_{CUE}$, the test for overidentifying restrictions is

$$\hat{S}_{CUE} = \hat{S}(\hat{\beta}_{CUE}) = n\hat{g}\left(\hat{\beta}_{CUE}\right)' \hat{\Omega}\left(\hat{\beta}_{CUE}\right)^{-1} \hat{g}\left(\hat{\beta}_{CUE}\right).$$

A major advantage of the CUE is that it has invariance properties. The two-step GMM estimator requires that the researcher make an initial choice about the weighting matrix used in the first step. This choice affects the numerical values of the final estimates, even if this difference is of sufficiently low order that it does not affect the large-sample asymptotic distribution. Hansen et al. (1996) find that $\hat{S}_{CUE}$ is more reliable than $\hat{S}_{GMM}$ in terms of the size properties, even though $\hat{\beta}_{CUE}$ tends to have heavy tails.

Other alternative estimators, especially ET and EL, have appealing information-theoretic interpretations in addition to being invariant to linear transformations of the moment functions. EL is a nonparametric method of inference based on a data-driven likelihood ratio function. Intuitively, given i.i.d data z_i $(i = 1, ..., n)$ only, the natural estimate the distribution of z_i is the empirical distribution, which puts weight $1/n$ on each of the n sample points. However in a GMM setting because of the moment restrictions $E[g(z, \beta_0)] = 0$, the empirical distribution function with weights $1/n$ does not satisfy this restriction. The idea behind EL is to modify the weights to ensure that the estimated distribution does satisfy the restrictions. The empirical likelihood estimator is obtained by minimizing the following problem,

$$\hat{\beta}_{EL} = \arg \min_{\beta \in B, \pi_1 \ldots \pi_n} -\sum_{i=1}^{n} \ln(\pi_i), \text{ subject to } \sum_{i=1}^{n} \pi_i g_i(\beta) = 0, \quad \sum_{i=1}^{n} \pi_i = 1. \tag{3}$$

Based on minimization of the Kullback-Leibler information criterion, Imbens et al. (1998) propose the exponential tilting estimator such that

$$\hat{\beta}_{ET} = \arg \min_{\beta \in B, \pi_1 \ldots \pi_n} -\sum_{i=1}^{n} \pi_i \ln(\pi_i), \text{ subject to } \sum_{i=1}^{n} \pi_i g_i(\beta) = 0, \quad \sum_{i=1}^{n} \pi_i = 1. \tag{4}$$

From a perspective of computation, the optimization problems of (3) and (4) are not attractive since they have a dimension $n + dim(\beta)$ which is larger than the sample size n. It is more convenient to rewrite them into a saddle point problem. To describe it, let $\rho(v)$ be a function of a scalar v that is concave in its domain, an open interval V containing zero. Let $\hat{\Lambda}_n(\beta) = \{\lambda : \lambda' g_i(\beta) \in V, i = 1, \cdots, n\}$,the GEL estimator $\hat{\beta}_{GEL}$ is the solution to a saddle point problem:

$$\hat{\beta}_{GEL} = \arg \min_{\beta \in B} \sup_{\lambda \in \hat{\Lambda}_n(\beta)} \sum_{i=1}^{n} \rho(\lambda' g_i(\beta)).$$

The EL estimator is a special case of GEL with $\rho(v) = \ln(1 - v)$, $V = (-\infty, 1)$. The ET estimator is a special case of GEL with $\rho(v) = -\exp(v)$. Newey and Smith (2004) also show that CUE is a member of GEL with $\rho(v) = -(v+1)^2/2$.

In contrast to the two-step GMM, computation of GEL is much more involved. Denote $\rho_j(v) = \partial^j \rho(v)/\partial v^j$, $\rho_j = \rho_j(0)$ $(j = 0, 1, 2, \cdots)$ and normalize that $\rho_1 = \rho_2 = -1$. For a given function $\rho(v)$, an associated GEL estimator $\hat{\beta}_{GEL}$, let

$$\hat{\pi}_{GEL,i} = \pi_i\left(\hat{\beta}_{GEL}, \hat{\lambda}_{GEL}\right) = \frac{\rho_1\left(\hat{\lambda}'_{GEL} g_i\left(\hat{\beta}_{GEL}\right)\right)}{\sum_{i=1}^n \rho_1\left(\hat{\lambda}'_{GEL} g_i\left(\hat{\beta}_{GEL}\right)\right)},$$

where

$$\hat{\lambda}_{GEL} = \lambda\left(\hat{\beta}_{GEL}\right) = \arg\max_{\lambda \in \hat{\Lambda}_n(\beta)} \sum_{i=1}^n \rho\left(\lambda' g_i\left(\hat{\beta}_{GEL}\right)\right)/n. \tag{5}$$

Define $k(v) = [\rho_1(v) + 1]/v$, $v \neq 0$, and $k(0) = -1$. Also, let $\hat{v}_{GEL,i} = \hat{\lambda}'_{GEL} g_i\left(\hat{\beta}_{GEL}\right)$, $\hat{k}_{GEL,i} = k(\hat{v}_{GEL,i})/\sum_{j=1}^n k(\hat{v}_{GEL,j})$. Theorem 2.3 in Newey and Smith (2004) show that the GEL's first-order conditions imply:

$$\left[\sum_{i=1}^n \hat{\pi}_{GEL,i} G_i\left(\hat{\beta}_{GEL}\right)\right]' \left[\sum_{i=1}^n \hat{k}_{GEL,i} g_i\left(\hat{\beta}_{GEL}\right) g_i\left(\hat{\beta}_{GEL}\right)'\right]^{-1} \hat{g}\left(\hat{\beta}_{GEL}\right) = 0. \tag{6}$$

From (6) we observe that, Instead of using the unweighted sample average to estimate the Jacobian of the moment conditions, GEL estimators employ an efficient estimator of the Jacobian of the moment conditions by using the implied probabilities $\hat{\pi}_{GEL,i}$. In addition, the EL estimator also makes use of an efficient estimator of the optimal weighting matrix.

Only in CUE case, $\hat{\lambda}_{GEL}$ and $\hat{\pi}_{GEL,i}$ have closed forms. In general EL and ET can be computed through a nested optimization algorithm, basing on (5) and (6). While the inner-loop optimization (5) with respect to the auxiliary parameters is usually a well defined convex optimization problem, the outer-loop optimization (6) is generally complicated by its highly nonlinear nature.

Associated with the empirical likelihood estimators are three tests for overidentiying restrictions that are similar to the classical trinity of the likelihood ratio, the score, and Lagrange multiplier tests. The likelihood-ratio type test is based on the value of the empirical likelihood function

$$\widehat{LR}_{GEL} = 2n\left[\sum_{i=1}^{n}\rho\left(\hat{\lambda}'_{GEL}g_i\left(\hat{\beta}_{GEL}\right)\right)/n-\rho_0\right].$$

The Lagrange-multiplier type test is

$$\widehat{LM}_{GEL} = n\hat{\lambda}'_{GEL}\left(\sum_{i=1}^{n}\hat{\pi}_{GEL,i}g_i\left(\hat{\beta}_{GEL}\right)g_i\left(\hat{\beta}_{GEL}\right)'\right)\hat{\lambda}_{GEL},$$

and the score type test[1] is

$$\hat{S}_{GEL} = n\hat{g}\left(\hat{\beta}_{GEL}\right)'\left(\sum_{i=1}^{n}\hat{\pi}_{GEL,i}g_i\left(\hat{\beta}_{GEL}\right)g_i\left(\hat{\beta}_{GEL}\right)'\right)^{-1}\hat{g}\left(\hat{\beta}_{GEL}\right).$$

All tests above follow a chi-squared distribution with $m-p$ degrees of freedom asymptotically under the null. Note that all tests require the calculation of GEL estimators in the first place. However, even with the rapid increase in computing power, it is still expensive or infeasible to compute them. Moreover, Monte Carlo simulation evidences have shown that the GEL estimators may suffer from "no moment" problem, see, for example, Hansen et al. (1996), Guggenberger (2008). When the focus is on tests for overidentifying restrictions, testing procedures circumventing the complicated estimation step should be useful. In the following, we will propose such tests.

3 The New Class of Tests for Overidentifying Restrictions

In this section, we adopt the same assumptions as in Newey and Smith (2004).

Assumption 1. *(a) $\beta_0 \in B$ is the unique solution to $E[g(z,\beta)] = 0$; (b) B is compact; (c) $g(z,\beta)$ is continuous at each $\beta \in B$ with probability 1; (d) $E\left[\sup_{\beta\in B}||g(z,\beta)||^{\alpha}\right] <$*

[1]The score type test is labeled as average moment test in Imbens et al. (1998), and bears lots of similarity with the J test.

∞ *for some* $\alpha > 2$; *(e)* Ω *is nonsingular, where* $\Omega = E\left[g_i(\beta_0)\, g_i(\beta_0)'\right]$; *(f)* $\rho(v)$ *is twice continuously differentiable in a neighborhood of zero.*

Assumption 2. *(a)* $\beta_0 \in int(B)$; *(b)* $g(z,\beta)$ *is continuously differentiable in a neighborhood* N *of* β_0 *and* $E\left[\sup_{\beta\in N} ||\partial g_i(\beta)/\partial\beta||\right] < \infty$; *(c)* $rank(G) = p$, *where* $G = E\left[\partial g_i(\beta_0)/\partial\beta\right]$.

Assumption 3. *Let* ∇^k *denote a vector of all distinct partial derivatives with respect to* β *of order* k. *There is* $b(z)$ *with* $E\left[b(z)^6\right] < \infty$ *such that for* $0 \le k \le 4$ *and all* z, $\nabla^k g(z,\beta)$ *exists on a neighborhood* N *of* β_0,$sup_{\beta\in N}\left\|\nabla^k g(z,\beta)\right\| \le b(z)$, *and for each* $\beta \in N$, $\|\nabla^4 g(z,\beta) - \nabla^4 g(z,\beta_0)\| \le b(z)\|\beta - \beta_0\|$. $\rho(v)$ *is four times continuously differentiable with Lipschitz fourth derivative in a neighborhood of zero.*

Suppose that one just gets some $\sqrt{n}$-consistent estimator $\hat{\beta}$, it may be obtained by the initial inefficient GMM estimation. Primitively, we present a test for overidentifying restrictions in GMM, which is denoted as $\widehat{TS}_{GMM}\left(\hat{\beta}\right)$, as follows:

$$\widehat{TS}_{GMM}\left(\hat{\beta}\right) = n\hat{g}\left(\hat{\beta}\right)' \hat{R}_{GMM}\left(\hat{\beta}\right)\hat{g}\left(\hat{\beta}\right),$$

where $\hat{R}_{GMM}(\beta) = \hat{\Omega}(\beta)^{-1} - \hat{\Omega}(\beta)^{-1}\hat{G}(\beta)\left[\hat{G}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{G}(\beta)\right]^{-1}\hat{G}(\beta)'\hat{\Omega}(\beta)^{-1}$. This test appears in Newey and McFadden (1994) Section 9.5. It is more convenient to rewrite $\widehat{TS}_{GMM}\left(\hat{\beta}\right)$ into the following

$$\widehat{TS}_{GMM}\left(\hat{\beta}\right) = n\bar{g}_{GMM}\left(\hat{\beta}\right)'\hat{\Omega}\left(\hat{\beta}\right)^{-1}\bar{g}_{GMM}\left(\hat{\beta}\right)',$$

where

$$\bar{g}_{GMM}\left(\hat{\beta}\right) = \hat{g}\left(\hat{\beta}\right) - \hat{G}\left(\hat{\beta}\right)\left[\hat{G}\left(\hat{\beta}\right)'\hat{\Omega}\left(\hat{\beta}\right)^{-1}\hat{G}\left(\hat{\beta}\right)\right]^{-1}\hat{G}\left(\hat{\beta}\right)'\hat{\Omega}\left(\hat{\beta}\right)^{-1}\hat{g}\left(\hat{\beta}\right).$$

This form bears a lot of similarities with the linearized classical test statistic proposed by Smith (1987) in ML. Note that the structure of $\bar{g}_{GMM}\left(\hat{\beta}\right)$ relies on the first-order condition of the two-step GMM objective function. Newey and McFadden (1994) derive the asymptotic first-order properties of $\widehat{TS}_{GMM}\left(\hat{\beta}\right)$ informally. Here we prove it robustly in the following theorem

Theorem 1. *Given a $\sqrt{n}$-consistent estimator $\hat{\beta}$, under Assumptions 1-2 and under the null hypothesis, the test statistic*

$$\widehat{TS}_{GMM}\left(\hat{\beta}\right) \xrightarrow{d} \chi^2_{m-p}.$$

Proof. See the Appendix. □

Extending this approach to continuous-updating GMM, we propose a new test of overidentifying restrictions as

$$\begin{aligned}\widehat{TS}_{CUE}\left(\hat{\beta}\right) &= n\hat{g}\left(\hat{\beta}\right)' \hat{R}_{CUE}\left(\hat{\beta}\right)\hat{g}\left(\hat{\beta}\right)\\ &= n\bar{g}_{CUE}\left(\hat{\beta}\right)'\hat{\Omega}\left(\hat{\beta}\right)^{-1}\bar{g}_{CUE}\left(\hat{\beta}\right)',\end{aligned}\tag{7}$$

where $\hat{R}_{CUE}(\beta) = \hat{\Omega}(\beta)^{-1} - \hat{\Omega}(\beta)^{-1}\hat{C}(\beta)\left[\hat{C}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{C}(\beta)\right]^{-1}\hat{C}(\beta)'\hat{\Omega}(\beta)^{-1}$, in which $\hat{C}(\beta)$ is defined as (2), and

$$\bar{g}_{CUE}\left(\hat{\beta}\right) = \hat{g}\left(\hat{\beta}\right) - \hat{C}(\beta)\left[\hat{C}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{C}(\beta)\right]^{-1}\hat{C}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{g}\left(\hat{\beta}\right).$$

The only difference between $\widehat{TS}_{GMM}\left(\hat{\beta}\right)$ and $\widehat{TS}_{CUE}\left(\hat{\beta}\right)$ is the estimator of $G(\beta)$ employed. We shall prove a similar result as Theorem 1 and discuss the higher order properties of $\widehat{TS}_{CUE}\left(\hat{\beta}\right)$ in GEL framework later on. Interestingly, Kleibergen (2005) proposes overidentification testing statistic $\widehat{TS}_{CUE}(\beta_0)$ in the case of weak identification. In the light of our general results below, no surprise that $\widehat{TS}_{CUE}(\beta_0)$ statistic holds in that case because the requirement of $\sqrt{n}(\beta_0 - \beta_0) = o_p(1)$ holds trivially. $\widehat{TS}_{CUE}\left(\hat{\beta}\right)$ should be quite useful since it avoids the complicated calculation of continuous-updating GMM estimators. Moreover, it has been reported that CUE suffers from the moment problem and exhibits wide dispersion, e.g, see Hansen et al. (1996). For this reason

Hansen argues "My own interest in the continuous-updating GMM estimator is not so much as a method for producing point estimates, but more as a method of making approximate inference." (Ghysels and Hall, 2002).

Right now, we are in a position to propose the new class of tests for overidentifying restrictioins in GEL framework. Given $\hat{\beta}$, we obtain $\lambda\left(\hat{\beta}\right)$ by maximizing the inner loop, that is

$$\hat{\lambda} \equiv \lambda\left(\hat{\beta}\right) = \arg\max_{\lambda \in \hat{\Lambda}_n\left(\hat{\beta}\right)} \sum_{i=1}^{n} \rho\left(\lambda' g_i\left(\hat{\beta}\right)\right)/n. \tag{8}$$

Let $\pi_i(\beta,\lambda) = \frac{\rho_1(\lambda' g_i(\beta))}{\sum_{i=1}^{n}\rho_1(\lambda' g_i(\beta))}$, $\hat{\pi}_i = \pi_i\left(\hat{\beta},\hat{\lambda}\right)$. Also, let $\hat{v}_i = \hat{\lambda}' g_i\left(\hat{\beta}\right)$, $\hat{k}_i = k(\hat{v}_i)/\sum_{j=1}^{T} k(\hat{v}_j)$, where $k(v) = [\rho_1(v)+1]/v$, $v \neq 0$, and $k(0) = -1$.

Define

$$\begin{aligned}
\bar{g}_{GEL}\left(\hat{\beta}\right) &= \hat{g}\left(\hat{\beta}\right) \\
&\quad - \sum_{i=1}^{n} \hat{\pi}_i G_i\left(\hat{\beta}\right)\left\{\left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right]'\left[\sum_{i=1}^{n}\hat{k}_i g_i\left(\hat{\beta}\right) g_i\left(\hat{\beta}\right)'\right]^{-1}\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right\}^{-1} \\
&\quad \times \left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right]'\left[\sum_{i=1}^{n}\hat{k}_i g_i\left(\hat{\beta}\right) g_i\left(\hat{\beta}\right)'\right]^{-1}\hat{g}\left(\hat{\beta}\right).
\end{aligned}$$

This transformation is based on the first-order condition of GEL. In terms of $\lambda\left(\hat{\beta}\right)$, we propose the following transformation:

$$\begin{aligned}
\bar{\lambda}_{GEL}\left(\hat{\beta}\right) &= \lambda\left(\hat{\beta}\right) \\
&\quad - \left[\frac{1}{n}\sum_{i=1}^{n} k(\hat{v}_i)\, g_i\left(\hat{\beta}\right) g_i\left(\hat{\beta}\right)'\right]^{-1}\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right) \\
&\quad \times \left\{\left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right]'\left[\frac{1}{n}\sum_{i=1}^{n} k(\hat{v}_i)\, g_i\left(\hat{\beta}\right) g_i\left(\hat{\beta}\right)'\right]^{-1}\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right\}^{-1} \\
&\quad \times \left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right]'\lambda\left(\hat{\beta}\right).
\end{aligned}$$

The new tests for overidentifying restrictions, which are denoted as $\widehat{TS}_{GEL}\left(\hat{\beta}\right)$ and

$\widehat{TLM}_{GEL}\left(\hat{\beta}\right)$ respectively, are

$$\widehat{TS}_{GEL}\left(\hat{\beta}\right) = n\bar{g}_{GEL}\left(\hat{\beta}\right)'\left[\sum_{i=1}^{n}\hat{\pi}_i g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]^{-1}\bar{g}_{GEL}\left(\hat{\beta}\right). \tag{9}$$

$$\widehat{TLM}_{GEL}\left(\hat{\beta}\right) = n\bar{\lambda}_{GEL}\left(\hat{\beta}\right)'\left[\sum_{i=1}^{n}\hat{\pi}_i g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]\bar{\lambda}_{GEL}\left(\hat{\beta}\right). \tag{10}$$

Comparing the tests for overidentifying restrictions based on GEL, the new class of tests only need solve (8), a simple convex optimization problem, circumventing the complicated outer-loop optimization problem. It is possible to follow alternative formulas proposed by Antoine et al. (2007) and Fan et al. (2011) to avoid solving (8). We do not exploit their approaches because of the unsophistication of the convex optimization problem.

In the case of CUE, $\hat{k}_i = 1/n$, $\lambda\left(\hat{\beta}\right) = -\hat{\Omega}\left(\hat{\beta}\right)^{-1}\hat{g}\left(\hat{\beta}\right)$, so $\widehat{TS}_{CUE}\left(\hat{\beta}\right) = \widehat{TLM}_{CUE}\left(\hat{\beta}\right)$.

When it comes to EL, $\hat{k}_i = \hat{\pi}_i$, we can rewrite (9) into

$$\begin{aligned}\widehat{TS}_{EL}\left(\hat{\beta}\right) &= n\bar{g}_{EL}\left(\hat{\beta}\right)'\left[\sum_{i=1}^{n}\hat{\pi}_i g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]^{-1}\bar{g}_{EL}\left(\hat{\beta}\right) && (11)\\ &= n\hat{g}\left(\hat{\beta}\right)'\hat{R}_{EL}\left(\hat{\beta}\right)\hat{g}\left(\hat{\beta}\right), && (12)\end{aligned}$$

where

$$\begin{aligned}\hat{R}_{EL}\left(\hat{\beta}\right) = &\left[\sum_{i=1}^{n}\hat{\pi}_i g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]^{-1}\\ &-\left[\sum_{i=1}^{n}\hat{\pi}_i g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]^{-1}\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\\ &\times\left\{\left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right]'\left[\sum_{i=1}^{n}\hat{\pi}_i g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]^{-1}\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right\}^{-1}\\ &\times\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\left[\sum_{i=1}^{n}\hat{\pi}_i g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]^{-1}.\end{aligned}$$

Different from $\widehat{TS}_{GMM}\left(\hat{\beta}\right)$, $\widehat{TS}_{EL}\left(\hat{\beta}\right)$ employs efficient estimates of Ω and G in this case.

In the theorem below, we characterize the asymptotic first-order properties of $\widehat{TS}_{GEL}\left(\hat{\beta}\right)$ and $\widehat{TLM}_{CUE}\left(\hat{\beta}\right)$.

Theorem 2. *Under Assumptions 1-2, and under the null hypothesis, if* $\hat{\beta} - \hat{\beta}_{GEL} = O_p\left(n^{-1/2}\right)$, *then the test statistics*

$$\widehat{TS}_{GEL}\left(\hat{\beta}\right) \xrightarrow{d} \chi^2_{m-p},$$

$$\widehat{TLM}_{GEL}\left(\hat{\beta}\right) \xrightarrow{d} \chi^2_{m-p}.$$

Proof. See the Appendix. □

The theorem shows that, given a $\sqrt{n}$ consistent estimator $\hat{\beta}$, $\widehat{TS}_{GEL}\left(\hat{\beta}\right)$ and $\widehat{TLM}_{GEL}\left(\hat{\beta}\right)$ are first-order equivalent to other overidentification tests. But one may expect that $\widehat{TS}_{GEL}\left(\hat{\beta}\right)$ and $\widehat{TLM}_{GEL}\left(\hat{\beta}\right)$ have better finite sample properties, since they take advantage of probability information implied by the moment restrictions.

Newey and Smith (2004) show that two-step GMM estimator and GEL estimator have the same first leading term in the stochastic expansions under some regularity conditions. In other words, there exists $\hat{\beta}$ such that $\hat{\beta} - \hat{\beta}_{GEL} = O_p\left(n^{-1}\right)$. Thus, with an asymptotically efficient estimator as our initial estimator, we will prove the higher-order equivalence between the new class of tests and GEL overidentification tests in the following theorem

Theorem 3. *Under Assumptions 1-3 and under the null hypothesis, If* $\hat{\beta} - \hat{\beta}_{GEL} = O_p\left(n^{-1}\right)$, *then*

$$\bar{g}_{GEL}\left(\hat{\beta}\right) = \hat{g}\left(\hat{\beta}_{GEL}\right) + O_p\left(n^{-3/2}\right),$$

$$\bar{\lambda}_{GEL}\left(\hat{\beta}\right) = \hat{\lambda}\left(\hat{\beta}_{GEL}\right) + O_p\left(n^{-3/2}\right),$$

$$\widehat{TS}_{GEL}\left(\hat{\beta}\right) = \hat{S}_{GEL} + O_p\left(n^{-1}\right),$$

and

$$\widehat{TLM}_{GEL}\left(\hat{\beta}\right) = \widehat{LM}_{GEL} + O_p\left(n^{-1}\right).$$

Proof. See the Appendix. □

This theorem shows that when $\hat{\beta}-\hat{\beta}_{GEL}=O_p\left(n^{-1}\right)$, $\widehat{TS}_{GEL}\left(\hat{\beta}\right)$ and $\widehat{TLM}_{GEL}\left(\hat{\beta}\right)$ are asymptotically second-order equivalent to $\hat{S}_{GEL}$ and $\widehat{LM}_{GEL}$ respectively. In a lot of cases, it is relatively easy to obtain asymptotically efficient estimators, note that the two-step GMM estimation does the job. In Imbens et al. (1998), in order to alleviate the computation burden, they start with $\hat{\beta}_{GMM}$ to obtain $\lambda\left(\hat{\beta}_{GMM}\right)$ by solving (8) in the case of ET, then construct a LM type test statistic basing on $\lambda\left(\hat{\beta}_{GMM}\right)$. However this statistic does not have this higher-order equivalence property.

In Fan et al. (2011) a new class of iterated GEL estimator is proposed such that $\hat{\beta}^j-\hat{\beta}_{GEL}=O_p\left(n^{-(j+1/)2}\right)$, if initial estimator $\hat{\beta}^0$ is a consistent estimator, $\hat{\beta}^j-\hat{\beta}_{GEL}=O_p\left(n^{-(j+2/)2}\right)$, if $\hat{\beta}^0-\hat{\beta}_{GEL}=O_p\left(n^{-1}\right)$, where j represents jth iteration. In this case, when $\hat{\beta}^j$ is employed, we have the following corollary:

Corollary 1. *Under Assumptions 1-3, if* $\hat{\beta}^j-\hat{\beta}_{GEL}=O_p\left(n^{-a}\right)$ *where* $a=(j+1/)2$ *or* $(j+2/)2$*, then*

$$\bar{g}_{GEL}\left(\hat{\beta}^j\right)=\hat{g}\left(\hat{\beta}_{GEL}\right)+O_p\left(n^{-a-1/2}\right),$$

$$\bar{\lambda}_{GEL}\left(\hat{\beta}^j\right)=\hat{\lambda}\left(\hat{\beta}_{GEL}\right)+O_p\left(n^{-a-1/2}\right),$$

$$\widehat{TS}_{GEL}\left(\hat{\beta}\right)=\hat{S}_{GEL}+O_p\left(n^{-a}\right),$$

and

$$\widehat{TLM}_{GEL}\left(\hat{\beta}\right)=\widehat{LM}_{GEL}+O_p\left(n^{-a}\right).$$

It is possible to construct tests for overidentifying restrictions by simply replacing $\hat{\beta}_{GEL}$ with $\hat{\beta}^j$ in $\hat{S}_{GEL}$ and $\widehat{LM}_{GEL}$, following Andrews (2002). However in this case it can be shown that $\hat{g}\left(\hat{\beta}^j\right)=\hat{g}\left(\hat{\beta}_{GEL}\right)+O_p\left(n^{-a}\right)$ and $\hat{\lambda}\left(\hat{\beta}^j\right)=\hat{\lambda}\left(\hat{\beta}_{GEL}\right)+O_p\left(n^{-a}\right)$, . In this sense, the $C\left(\alpha\right)$ type tests go one step further than Andrews' approach.

4 Monte Carlo Simulations

This section investigates the finite sample properties of $\widehat{TS}_{GEL}\left(\hat{\beta}\right)$ and $\widehat{TLM}_{GEL}\left(\hat{\beta}\right)$ proposed in previous sections. In particular, we examine their size properties, and assess their performance in comparison with tests $\widehat{S}_{GMM}$, $\widehat{LR}_{GEL}$, $\widehat{LM}_{GEL}$ and $\widehat{S}_{GEL}$.

4.1 Asset Pricing Model

We consider an extended version of an asset pricing model investigated by Hall and Horowitz (1996), Imbens et al. (1998). The parameter of interest is determined by the following moment conditions

$$Eg\left(X,\beta_0\right) = E\begin{pmatrix} r\left(X,\beta_0\right) \\ X_2 r\left(X,\beta_0\right) \\ \left(X_3-1\right) r\left(X,\beta_0\right) \\ \vdots \\ \left(X_m-1\right) r\left(X,\beta_0\right) \end{pmatrix} = 0,$$

where $X = \left(X_1, X_2, \cdots, X_m\right)$, $r\left(X,\beta\right) = \exp\left\{-0.72 - \beta\left(X_1+X_2\right) + 3X_2\right\} - 1$ and β is a scalar parameter. These restrictions are satisfied at $\beta_0 = 3$. Components of X are mutually independent. X_1, X_2 have a bivariate normal distribution with correlation coefficient zero, both means equal to zero and both variances equal to 0.16. $X_3, \cdots, X_m$ are independent and each follows a chi-squared distribution with 1 degree of freedom. We set $n = 200$, 400 and 800. The number of replications is $10,000$. The consistent estimator employed in $\widehat{TS}_{GEL}\left(\hat{\beta}\right)$ and $\widehat{TLM}_{GEL}\left(\hat{\beta}\right)$ is the two-step GMM estimator $\hat{\beta}_{GMM}$. In Imbens et al. (1998), Consistent estimators for the matrix Ω required in the computation of the $\widehat{LM}_{GEL}$ and $\hat{S}_{GEL}$ are obtained by using

$$\hat{\Omega}\left(\beta\right) = \sum_{i=1}^{n} \hat{\pi}\left(\beta,\lambda\right) g_i\left(\beta\right) g_i\left(\beta\right)',$$

or a robust estimate

$$\begin{aligned}\hat{\Omega}(\beta) &= \sum_{i=1}^{n} \hat{\pi}(\beta,\lambda)\, g_i(\beta)\, g_i(\beta)' \\ &\quad \times \left(n \sum_{i=1}^{n} \hat{\pi}^2(\beta,\lambda)\, g_i(\beta)\, g_i(\beta)'\right)^{-1} \sum_{i=1}^{n} \hat{\pi}(\beta,\lambda)\, g_i(\beta)\, g_i(\beta)'.\end{aligned}$$

In $\widehat{TS}_{GEL}\left(\hat{\beta}\right)$ and $\widehat{TLM}_{GEL}\left(\hat{\beta}\right)$, we also consider both estimates. It has been shown that the robust estimate only works well in the case of $\widehat{LM}_{ET}$ and $\widehat{TLM}_{ET}\left(\hat{\beta}\right)$, so we ignore the other results of the robust version of the tests in the tables. We use the Matlab package written by Evdokomiv and Kitamura (2011) to obtain the two-step GMM estimators and the CUE and EL overidentification test statistics, and modify their code to obtain the ET overidentification test statistics. As for the new class of tests, we rely on their inner-loop optimization code to obtain $\lambda\left(\hat{\beta}\right)$. The simulation results are reported in Tables 1 and 2. We summarize the simulation results in the following

1. In general, $\widehat{S}_{GMM}$ is heavily oversized, especially when $m = 3$. The increase of sample size does not change this pattern. So it is not reliable to use $\widehat{S}_{GMM}$ as a diagnostic tool in this example.

2. When $m = 2$, the superiority of the tests in the new class over $\widehat{S}_{GMM}$ in terms of the size property is not clear-cut. When $m = 3$, the tests in the new class, except for $\widehat{TS}_{ET}\left(\hat{\beta}_{GMM}\right)$, have better size properties than $\widehat{S}_{GMM}$. Among them, $\widehat{TLM}_{ET}\left(\hat{\beta}_{GMM}\right)$ with robust estimate of Ω performs best.

3. The size properties of $\widehat{TS}_{GEL}\left(\hat{\beta}_{GMM}\right)$ and $\widehat{TLM}_{GEL}\left(\hat{\beta}_{GMM}\right)$ are comparable to $\hat{S}_{GEL}$ and $\widehat{LM}_{GEL}$ respectively. In some cases $\widehat{TS}_{GEL}\left(\hat{\beta}_{GMM}\right)$ and $\widehat{TLM}_{GEL}\left(\hat{\beta}_{GMM}\right)$ even perform slightly better than $\hat{S}_{GEL}$ and $\widehat{LM}_{GEL}$. $\widehat{TLM}_{ET}\left(\hat{\beta}_{GMM}\right)$ with robust estimate of Ω has the best size properties among all the tests we consider.

All in all, the simulation results demonstrate that the new class of tests for overidentifying restrictions has better finite sample performance than the two-step GMM overidentification test, and compares well to several potential alternatives in terms of overall performance, which echoes the theoretical results we obtained in the previous section. Given the nice size properties and computational simplicity, the new class of tests should be quite useful when the GEL estimation is cumbersome.

4.2 Chi-squared Moments Model

As in Imbens et al. (1998), the moment vector is

$$E\left[g\left(X,\beta\right)\right] = E\begin{pmatrix} X-\beta \\ X^2-\beta^2-2\beta \end{pmatrix} = 0.$$

The distribution of X is chi-square with one degree of freedom, and $\beta_0 = 1$. Again, the consistent estimator employed in $\widehat{TS}_{GEL}\left(\hat{\beta}\right)$ and $\widehat{TLM}_{GEL}\left(\hat{\beta}\right)$ is the two-step GMM estimator $\hat{\beta}_{GMM}$. The number of replications is $10,000$. The sample size is 500 and $1,000$. The simulation results are reported in Table 3.

We summarize the results in the following

1. Among class of tests of $\widehat{LM}_{GEL}$, $\hat{S}_{GEL}$ and $\widehat{LR}_{GEL}$ (ET and EL), $\widehat{LM}_{ET}$ with the robust estimate of Ω performs best in most cases. $\widehat{LM}_{GEL}$, $\hat{S}_{GEL}$ and $\widehat{LR}_{GEL}$ all tend to be oversized, which is in accordance with the Monte Carlo evidences reported by Imbens et al. (1998) and Ramalho and Smith (2006).

2. The size properties of $\widehat{TS}_{GEL}\left(\hat{\beta}\right)$ and $\widehat{TLM}_{GEL}\left(\hat{\beta}\right)$ are identical to $\hat{S}_{GEL}$ and $\widehat{LM}_{GEL}$ respectively.

5 Conclusion

In this paper, we propose a new class of tests for overidentifying restrictions in moment condition models. These tests extend the idea of $C(\alpha)$ test of Neyman (1959) in ML to GEL framework. They are easy to compute, circumventing the complicated saddle point characterization in GEL estimation. It has be shown that these tests share the same higher order properties as GEL overidentification tests, given some proper consistent estimators. Monte carlo simulation has shown this new class of tests for overidentifying restrictions has better finite sample performance than the J test, and compares well to several potential alternatives in terms of overall performance. Given the nice finite sample properties and computational simplicity, the new class of tests should be quite useful when the GEL estimation is cumbersome and the focus is on inference.

Appendix

Proof of Theorem 1. Given $\hat{\beta}$ such that $\sqrt{n}\left(\hat{\beta}-\beta_0\right) = O_p(1)$, by first order Taylor expansion, $\sqrt{n}\hat{g}\left(\hat{\beta}\right) = \sqrt{n}\hat{g}\left(\beta_0\right) + \hat{G}\left(\bar{\beta}\right)\sqrt{n}\left(\tilde{\beta}-\beta_0\right)$, where $\tilde{\beta}$ lies between $\hat{\beta}$ and β_0. By Assumption 1, $\hat{\Omega}\left(\hat{\beta}\right) \xrightarrow{p} \Omega$, $\hat{\Omega}\left(\hat{\beta}\right)^{-1} \xrightarrow{p} \Omega^{-1}$. By Assumption 2, $\hat{G}\left(\bar{\beta}\right) = G + o_p(1)$, $\hat{G}\left(\hat{\beta}\right) = G + o_p(1)$. So by Slutsky Theorem

$$\begin{aligned}
\sqrt{n}\bar{g}_{GMM}\left(\hat{\beta}\right) &= \left(I - \hat{G}\left(\hat{\beta}\right)\left[\hat{G}\left(\hat{\beta}\right)'\hat{\Omega}\left(\hat{\beta}\right)^{-1}\hat{G}\left(\hat{\beta}\right)\right]^{-1}\hat{G}\left(\hat{\beta}\right)'\hat{\Omega}\left(\hat{\beta}\right)^{-1}\right) \\
&\quad \times \left(\sqrt{n}\hat{g}\left(\beta_0\right) + \hat{G}\left(\bar{\beta}\right)\sqrt{n}\left(\hat{\beta}-\beta_0\right)\right) \\
&= \sqrt{n}\hat{g}\left(\beta_0\right) - G\left[G'\Omega^{-1}G\right]^{-1}G'\Omega\sqrt{n}\hat{g}\left(\beta_0\right) + o_p(1).
\end{aligned}$$

Then

$$n\bar{g}'_{GMM}\left(\hat{\beta}\right)\hat{\Omega}\left(\hat{\beta}\right)^{-1}\bar{g}_{GMM}\left(\hat{\beta}\right) \xrightarrow{d} \chi^2_{m-p}.$$

$\square$

Proof of Theorem 2. Given that $\sqrt{n}\left(\hat{\beta}-\beta_0\right) = O_p(1)$, $\hat{g}\left(\hat{\beta}\right) = O_p\left(n^{-1/2}\right)$ and Assumption 1, based on Lemma A2 in Newey and Smith (2004), we get $\hat{\lambda} = O_p\left(n^{-1/2}\right)$. Lemma A1 in Newey and Smith (2004) implies that $\max_{1\le i\le n}\left|\hat{\lambda}'g_i\left(\hat{\beta}\right)\right| = o_p(1)$, then

$\rho_1\left(\hat{\lambda}'g_i\left(\hat{\beta}\right)\right)-\rho_1(0)=o_p(1)$. So we have

$$\hat{\pi}_i=\frac{\rho_1\left(\hat{\lambda}'g_i\left(\hat{\beta}\right)\right)}{\sum_{i=1}^n\rho_1\left(\hat{\lambda}'g_i\left(\hat{\beta}\right)\right)}=\frac{1}{n}\left[1+o_p(1)\right],\text{ uniformly in } i=1,\cdots,n.$$

So $\sum_{i=1}^n\hat{\pi}_iG_i\left(\hat{\beta}\right)=\hat{G}\left(\hat{\beta}\right)+o_p(1)$.

Similarly

$$k\left(\hat{\lambda}'g_i\left(\hat{\beta}\right)\right)=\frac{\rho_1\left(\hat{\lambda}'g_i\left(\hat{\beta}\right)\right)+1}{\hat{\lambda}'g_i\left(\hat{\beta}\right)}=\rho_2(0)+o_p(1),\text{ uniformly in } i=1,\cdots,n.$$

$$\hat{k}_i=\frac{k\left(\hat{\lambda}'g_i\left(\hat{\beta}\right)\right)}{\sum_{j=1}^n k\left(\hat{\lambda}'g_j\left(\hat{\beta}\right)\right)}=\frac{1}{n}\left[1+o_p(1)\right],\text{ uniformly in } i=1,\cdots,n.$$

So $\sum_{i=1}^n\hat{k}_ig_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'=\hat{\Omega}\left(\hat{\beta}\right)+o_p(1)$. Then we get to the conclusion that $\widehat{TS}_{GEL}\left(\hat{\beta}\right)\xrightarrow{d}\chi^2_{m-p}$ following the logic of proof of Theorem 1.

To prove $\widehat{TLM}_{GEL}\left(\hat{\beta}\right)\xrightarrow{d}\chi^2_{m-p}$, by the first order condition of the inner loop optimization

$$\sum_{i=1}^n\rho_1\left(\hat{\lambda}'g_i\left(\hat{\beta}\right)\right)g_i\left(\hat{\beta}\right)=0.$$

By the definition of $k(\hat{v}_i)$, we have

$$\sum_{i=1}^n k(\hat{v}_i)\,g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\hat{\lambda}-n\hat{g}\left(\hat{\beta}\right)=0.$$

Note that $\frac{1}{n}\sum_{i=1}^n k(\hat{v}_i)\,g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'=\hat{\Omega}\left(\hat{\beta}\right)+o_p(1)$. Then

$$\hat{\lambda}=\left(\frac{1}{n}\sum_{i=1}^n k(\hat{v}_i)\,g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right)^{-1}\hat{g}\left(\hat{\beta}\right).$$

So

$$\bar{\lambda}_{GEL}\left(\hat{\beta}\right)=\left(\frac{1}{n}\sum_{i=1}^n k(\hat{v}_i)\,g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right)^{-1}\left(\bar{g}_{GMM}\left(\hat{\beta}\right)+o_p(1)\right).$$

Then the conclusion follows. $\square$

In order to prove Theorem 3, we introduce the following Lemmas.

Lemma 1.

$$\hat{\lambda} = \lambda\left(\hat{\beta}\right) = \arg\max_{\lambda\in\hat{\Lambda}_n\left(\hat{\beta}\right)} \sum_{i=1}^{n} \rho\left(\lambda' g_i\left(\hat{\beta}\right)\right)/n.$$

$$\hat{\lambda}_{GEL} = \lambda\left(\hat{\beta}_{GEL}\right) = \arg\max_{\lambda\in\hat{\Lambda}_n\left(\hat{\beta}_{GEL}\right)} \sum_{i=1}^{n} \rho\left(\lambda' g_i\left(\hat{\beta}_{GEL}\right)\right)/n.$$

Then $\hat{\lambda} - \hat{\lambda}_{GEL} = O_p\left(\hat{\beta} - \hat{\beta}_{GEL}\right)$.

Proof. By the first order condition

$$\sum_{i=1}^{n} \rho_1\left(\hat{\lambda}' g_i\left(\hat{\beta}\right)\right) g_i\left(\hat{\beta}\right) = 0,$$

$$\sum_{i=1}^{n} \rho_1\left(\hat{\lambda}'_{GEL} g_i\left(\hat{\beta}_{GEL}\right)\right) g_i\left(\hat{\beta}_{GEL}\right) = 0.$$

By taking Taylor expansion around $\left(\hat{\beta}'_{GEL}, \hat{\lambda}'_{GEL}\right)'$ for $\sum_{i=1}^{n} \rho_1\left(\hat{\lambda}' g_i\left(\hat{\beta}\right)\right) g_i\left(\hat{\beta}\right)$, we obtain

$$\begin{aligned} 0 &= 0 + \frac{1}{n}\sum_{i=1}^{n} \rho_1\left(\tilde{\lambda}' g_i\left(\hat{\beta}\right)\right) G_i\left(\tilde{\beta}\right)\left(\hat{\beta} - \hat{\beta}_{GEL}\right) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \rho_2\left(\tilde{\lambda}' g_i\left(\hat{\beta}\right)\right) g_i\left(\tilde{\beta}\right) g_i\left(\hat{\beta}\right)'\left(\hat{\lambda} - \hat{\lambda}_{GEL}\right), \end{aligned}$$

where $\tilde{\beta}$ and $\tilde{\lambda}$ are values between $\hat{\beta}, \hat{\beta}_{GEL}$ and $\hat{\lambda}, \hat{\lambda}_{GEL}$ respectively that actually differ from row to row of the matrix $\rho_1\left(\tilde{\lambda}' g_i\left(\hat{\beta}\right)\right) G_i\left(\tilde{\beta}\right)$ and $\rho_2\left(\tilde{\lambda}' g_i\left(\hat{\beta}\right)\right) g_i\left(\tilde{\beta}\right) g_i\left(\hat{\beta}\right)'$. It follows as Lemma A1 in Newey and Smith (2004) that $max_{i\le n}\left|\tilde{\lambda}' g_i\left(\hat{\beta}\right)\right| \xrightarrow{p} 0$. Therefore, $max_{i\le n}\left|\rho_1\left(\tilde{\lambda}' g_i\left(\hat{\beta}\right)\right) + 1\right| \xrightarrow{p} 0$ and $max_{i\le n}\left|\rho_2\left(\tilde{\lambda}' g_i\left(\hat{\beta}\right)\right) + 1\right| \xrightarrow{p} 0$. It then follows from uniformly weak law of large numbers (UWL) that $\frac{1}{n}\sum_{i=1}^{n} \rho_1\left(\tilde{\lambda}' g_i\left(\hat{\beta}\right)\right) G_i\left(\tilde{\beta}\right) \xrightarrow{p} G$, and $\frac{1}{n}\sum_{i=1}^{n} \rho_2\left(\tilde{\lambda}' g_i\left(\hat{\beta}\right)\right) g_i\left(\tilde{\beta}\right) g_i\left(\hat{\beta}\right)' \xrightarrow{p} \Omega$. Then we get to conclusion. $\square$

Lemma 2. *If* $\hat{\beta} - \hat{\beta}_{GEL} = O_p\left(n^{-1}\right)$*, then*

$$\sum_{i=1}^{n} \hat{\pi}_i g_i\left(\hat{\beta}\right) g_i\left(\hat{\beta}\right)' - \sum_{i=1}^{n} \hat{\pi}_{GEL,i} g_i\left(\hat{\beta}_{GEL}\right) g_i\left(\hat{\beta}_{GEL}\right)' = O_p\left(n^{-1}\right).$$

Proof. Denote

$$A_1 = \sum_{i=1}^{n} \left(\hat{\pi}_i - \hat{\pi}_{GEL,i}\right) g_i\left(\hat{\beta}\right) g_i\left(\hat{\beta}\right)',$$

and

$$A_2 = \sum_{i=1}^{n} \hat{\pi}_{GEL,i} \left[g_i\left(\hat{\beta}_{GEL}\right) g_i\left(\hat{\beta}_{GEL}\right)' - g_i\left(\hat{\beta}\right) g_i\left(\hat{\beta}\right)' \right].$$

So

$$\sum_{i=1}^{n} \hat{\pi}_i g_i\left(\hat{\beta}\right) g_i\left(\hat{\beta}\right)' - \sum_{i=1}^{n} \hat{\pi}_{GEL,i} g_i\left(\hat{\beta}_{GEL}\right) g_i\left(\hat{\beta}_{GEL}\right)' = A_1 + A_2.$$

By Taylor Expansion

$$\begin{aligned} \hat{\pi}_i - \hat{\pi}_{GEL,i} &= \left(\frac{\rho_2\left(\tilde{\lambda}' g_i\left(\tilde{\beta}\right)\right)}{\sum_{i=1}^{n} \rho_1\left(\tilde{\lambda}' g_i\left(\tilde{\beta}\right)\right)} - \frac{\rho_1\left(\tilde{\lambda}' g_i\left(\tilde{\beta}\right)\right) \sum_{i=1}^{n} \rho_2\left(\tilde{\lambda}' g_i\left(\tilde{\beta}\right)\right)}{\left(\sum_{i=1}^{n} \rho_1\left(\tilde{\lambda}' g_i\left(\tilde{\beta}\right)\right)\right)^2} \right) \\ &\quad \times \left(\hat{\lambda}' g_i\left(\hat{\beta}\right) - \hat{\lambda}_{GEL}' g_i\left(\hat{\beta}_{GEL}\right)\right), \end{aligned}$$

where $\tilde{\beta}$ and $\tilde{\lambda}$ are values between $\hat{\beta}$, $\hat{\beta}_{GEL}$ and $\hat{\lambda}$, $\hat{\lambda}_{GEL}$. It again follows Lemma A1 in Newey and Smith (2004) $max_{i \leq n} \left|\tilde{\lambda}' g_i\left(\tilde{\beta}\right)\right| \xrightarrow{p} 0$, then

$$\frac{\rho_2\left(\tilde{\lambda}' g_i\left(\tilde{\beta}\right)\right)}{\sum_{i=1}^{n} \rho_1\left(\tilde{\lambda}' g_i\left(\tilde{\beta}\right)\right)} = \frac{1}{n}\left[1 + o_p(1)\right], \text{ uniformly in } i = 1, \cdots, n.$$

$$\frac{\rho_1\left(\tilde{\lambda}' g_i\left(\tilde{\beta}\right)\right) \sum_{i=1}^{n} \rho_2\left(\tilde{\lambda}' g_i\left(\tilde{\beta}\right)\right)}{\left(\sum_{i=1}^{n} \rho_1\left(\tilde{\lambda}' g_i\left(\tilde{\beta}\right)\right)\right)^2} = \frac{1}{n}\left[1 + o_p(1)\right], \text{uniformly in } i = 1, \cdots, n.$$

So

$$\hat{\pi}_i - \hat{\pi}_{GEL,i} = \frac{o_p(1)}{n} \left(\hat{\lambda}' g_i\left(\hat{\beta}\right) - \hat{\lambda}_{GEL}' g_i\left(\hat{\beta}_{GEL}\right)\right).$$

Take Taylor expansion on $g_i\left(\hat{\beta}_{GEL}\right)$ around $\hat{\beta}$, we have

$$\hat{\pi}_i - \hat{\pi}_{GEL,i} = \frac{o_p(1)}{n}\left(\left(\hat{\lambda} - \hat{\lambda}_{GEL}\right)' g_i\left(\hat{\beta}\right) - \hat{\lambda}'_{GEL} G_i\left(\breve{\beta}\right)\left(\hat{\beta} - \hat{\beta}_{GEL}\right)\right),$$

where $\breve{\beta}$ is a value between $\hat{\beta}$ and $\hat{\beta}_{GEL}$. When $\hat{\beta} - \hat{\beta}_{GEL} = O_p\left(n^{-1}\right)$, then $\hat{\lambda} - \hat{\lambda}_{GEL} = O_p\left(n^{-1}\right)$ by Lemma 1. So

$$A_1 = o_p\left(n^{-1}\right).$$

On the other hand,

$$\begin{aligned} vec\left[g_i\left(\hat{\beta}_{GEL}\right) g_i\left(\hat{\beta}_{GEL}\right)' - g_i\left(\hat{\beta}\right) g_i\left(\hat{\beta}\right)'\right] &= \left[G_i\left(\tilde{\beta}\right) \otimes g_i\left(\tilde{\beta}\right) + g_i\left(\tilde{\beta}\right) \otimes G_i\left(\tilde{\beta}\right)\right] \\ &\quad \times \left(\hat{\beta}_{GEL} - \hat{\beta}\right), \end{aligned}$$

where $\tilde{\beta}$ is between $\hat{\beta}_{GEL}$ and $\hat{\beta}$. So

$$\begin{aligned} vec(A_2) &= \left(E\left[G_i\left(\beta_0\right) \otimes g_i\left(\beta_0\right) + g_i\left(\beta_0\right) \otimes G_i\left(\beta_0\right)\right] + o_p(1)\right) \\ &\quad \times \left(\hat{\beta}_{GEL} - \hat{\beta}\right) \\ &= O_p\left(n^{-1}\right). \end{aligned}$$

Then we get to the result. □

Proof of Theorem 3. We only prove $\bar{g}_{GEL}\left(\hat{\beta}\right) - \hat{g}\left(\hat{\beta}_{GEL}\right) = O_p\left(n^{-3/2}\right)$, and $\widehat{TS}_{GEL}\left(\hat{\beta}\right) - \hat{S}_{GEL} = O_p\left(n^{-1}\right)$. Results about $\bar{\lambda}_{GEL}\left(\hat{\beta}\right)$ and $\widehat{TLM}_{GEL}\left(\hat{\beta}\right)$ can be proved following similar argument. By Taylor expansion

$$\hat{g}\left(\hat{\beta}\right) - \hat{g}\left(\hat{\beta}_{GEL}\right) = \hat{G}\left(\tilde{\beta}\right)\left(\hat{\beta} - \hat{\beta}_{GEL}\right),$$

where $\tilde{\beta}$ lies between $\hat{\beta}$ and $\hat{\beta}_{GEL}$. We use the result in the proof of Theorem 4.2 in Fan et al. (2011)

$$\sum_{i=1}^{n} \hat{\pi}_i G_i\left(\hat{\beta}\right) - \hat{G}\left(\hat{\beta}\right) = O_p\left(n^{-1/2}\right).$$

So

$$\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)-\hat{G}\left(\tilde{\beta}\right)=\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)-\hat{G}\left(\hat{\beta}\right)+\hat{G}\left(\hat{\beta}\right)-\hat{G}\left(\tilde{\beta}\right)=O_p\left(n^{-1/2}\right).$$

Then

$$\begin{aligned}\bar{g}_{GEL}\left(\hat{\beta}\right)-\hat{g}\left(\hat{\beta}_{GEL}\right) &= \left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)+O_p\left(n^{-1/2}\right)\right]\left(\hat{\beta}-\hat{\beta}_{GEL}\right)\\ &\quad -\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\left\{\left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right]'\left[\sum_{i=1}^{n}\hat{k}_i g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]^{-1}\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right\}^{-1}\\ &\quad \times\left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right]'\left[\sum_{i=1}^{n}\hat{k}_i g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]^{-1}\hat{g}\left(\hat{\beta}\right)\\ &= \left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right]\left[\hat{\beta}^1-\hat{\beta}_{GEL}\right]+O_p\left(n^{-3/2}\right),\end{aligned}$$

where

$$\begin{aligned}\hat{\beta}^1 &= \hat{\beta}-\left\{\left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right]'\left[\sum_{i=1}^{n}\hat{k}_i g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]^{-1}\left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right]\right\}^{-1}\\ &\quad\times\left[\sum_{i=1}^{n}\hat{\pi}_i G_i\left(\hat{\beta}\right)\right]'\left[\sum_{i=1}^{n}\hat{k}_i g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]^{-1}\hat{g}\left(\hat{\beta}\right).\end{aligned}$$

Specifcally, denote

$$f_n(\beta)=\left[\sum_{i=1}^{n}\pi_i(\beta,\lambda)G_i(\beta)\right]'\left[\sum_{i=1}^{n}k_i(\beta,\lambda)g_i\left(\hat{\beta}\right)g_i\left(\hat{\beta}\right)'\right]^{-1}\hat{g}\left(\hat{\beta}\right),$$

where $k_i(\beta,\lambda)=k(\lambda' g_i(\beta))/\sum_{j=1}^{T}k(\lambda' g_i(\beta))$. Define $F_n(\beta)=\partial f_n(\beta)/\partial\beta'$. Then

$$\left[\sum_{i=1}^{n}\pi_i(\beta,\lambda)G_i(\beta)\right]'\left[\sum_{i=1}^{n}k_i(\beta,\lambda)g_i(\beta)g_i(\beta)'\right]^{-1}\left[\sum_{i=1}^{n}\pi_i(\beta,\lambda)G_i(\beta)\right]=F_n(\beta)+O_p\left(n^{-1/2}\right).$$

Following Theorem 5 in Robinson (1988), we have $\hat{\beta}^1-\hat{\beta}_{GEL}=O_p\left(n^{-3/2}\right)$. Then

$$\bar{g}_{GEL}\left(\hat{\beta}\right) - \hat{g}\left(\hat{\beta}_{GEL}\right) = \left[\sum_{i=1}^{n} \hat{\pi}_i G_i\left(\hat{\beta}\right)\right]\left[\hat{\beta}^1 - \hat{\beta}_{GEL}\right] + O_p\left(n^{-3/2}\right) = O_p\left(n^{-3/2}\right).$$

So basing on this result and Lemma 2, we have

$$\begin{aligned}
\widehat{TS}_{GEL}\left(\hat{\beta}\right) - \hat{S}_{GEL} &= n(\hat{g}\left(\hat{\beta}_{GEL}\right) + O_p\left(n^{-3/2}\right))' \\
&\quad \times \left(\left(\sum_{i=1}^{n} \hat{\pi}_{GEL,i} g_i\left(\hat{\beta}_{GEL}\right) g_i\left(\hat{\beta}_{GEL}\right)'\right)^{-1} + O_p\left(n^{-1}\right)\right) \\
&\quad \times (\hat{g}\left(\hat{\beta}_{GEL}\right) + O_p\left(n^{-3/2}\right)) \\
&\quad - n\hat{g}\left(\hat{\beta}_{GEL}\right)'\left(\sum_{i=1}^{n} \hat{\pi}_{GEL,i} g_i\left(\hat{\beta}_{GEL}\right) g_i\left(\hat{\beta}_{GEL}\right)'\right)^{-1} \hat{g}\left(\hat{\beta}_{GEL}\right) \\
&= nO_p\left(n^{-1/2}\right) O_p(1)\, O_p\left(n^{-3/2}\right) + nO_p\left(n^{-1/2}\right) O_p\left(n^{-1}\right) O_p\left(n^{-1/2}\right) \\
&\quad + nO_p\left(n^{-1/2}\right) O_p\left(n^{-1}\right) O_p\left(n^{-3/2}\right) + nO_p\left(n^{-3/2}\right) O_p(1)\, O_p\left(n^{-1/2}\right) \\
&\quad + nO_p\left(n^{-3/2}\right) O_p(1)\, O_p\left(n^{-3/2}\right) + nO_p\left(n^{-3/2}\right) O_p\left(n^{-1}\right) O_p\left(n^{-1/2}\right) \\
&\quad + nO_p\left(n^{-3/2}\right) O_p\left(n^{-1}\right) O_p\left(n^{-3/2}\right) \\
&= O_p\left(n^{-1}\right).
\end{aligned}$$

□

References

[1] Andrews, D.W.K. (2002). Equivalence of the Higher Order Asymptotic Efficiency of k-step and Extremum Statistics, Econometric Theory, 18, 1040-1085.

[2] Antoine, B., H. Bonnal, and E. Renault (2007), On the Efficient Use of the Information Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood, Journal of Econometrics, 138, 461-487.

[3] Brown B., and K.W. Newey (2002), Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference. Journal of Business and Economic Statis-

tics, 20, 507-517.

[4] Evdokomiv K., and Y. Kitamura (2011), MATLAB/STATA codes for EL. http://kitamura.sites.yale.edu/matlabstata-codes-el.

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

[6] Ghysels, Eric and Alastair Hall, (2002). Interview with Lars Peter Hansen. Journal of Business and Economic Statistics Twentieth Anniversary Issue on the Generalized Method of Moments 20 (4): 442-447.

[7] Guggenberger, P. (2008), Finite Sample Evidence suggesting a Heavy Tail Problem the Generalized Empirical Likelihood Estimator, Econometric Reviews, 26, 526-541.

[8] Hall, P., Horowitz, J.L., (1996), Bootstrap critical values for tests based on generalized method of moments estimators. Econometrica 64, 891–916.

[9] Hansen, L.P. (1982), Large Sample Properties of Generalized Method of Moments Estimators, Econometrica, 50, 1029-1054.

[10] Hansen, L.P., J. Heaton, and A. Yaron (1996), finite-Sample Properties of Some Alternative GMM Estimators, Journal of Business and Economic Statistics, 14, 262-280.

[11] Imbens, G. (1997), One-Step Estimators for Over-Identified Generalized Method of Moments Models, Review of Economic Studies, 64, 359-383.

[12] Imbens, G. (2002), Generalized Method of Moments and Empirical Likelihood, Journal of Business and Economic Statistics, 20, 493-506.

[13] Imbens, G., R.H. Spady, and P. Johnson (1998), Information Theoretic Approaches to Inference in Moment Condition Models, Econometrica, 66, 333-357.

[14] Kttamura, Y.(2007). Empirical Likelihood Methods in Econometrics: Theory and Practice, in Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society, Vol. ed. by R. Blundell, W. Newey, and T. Persson. Cambridge, MA.

[15] Kitamura, Y. and M. Stutzer (1997), On Information-Theoretic Alternative to Generalized Method of Moments Estimation, Econometrica, 65, 861-874.

[16] Kleibergen, F. (2005), Testing Parameters in GMM without Assuming that They Are Identified, Econometrica, 73, 1103-1123.

[17] Newey, W.K., McFadden, D., (1994). Large Sample Estimation and Hypothesis Testing. In: Engle, McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. North-Holland Amsterdam, 2113–2245.

[18] Newey, W.K. and R.J. Smith (2004), Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators, Econometrica, 72, 219-255.

[19] Neyman, J. (1959). Optimal Asymptotic Tests of Composite Statistical Hypotheses., in Grenander, U. (ed.) Probability and Statistics (New York: Wiley).

[20] Owen, A. (1988), Empirical Likelihood Ratio Confidence Intervals for a Single Func-tional, Biometrika, 75, 237-249.

[21] Owen, A. (1990), Empirical Likelihood Ratio Confidence Regions, The Annals of Statistics, 18(1), 90–120.

[22] Owen, A. (2001), Empirical Likelihood. Chapman and Hall/CRC.

[23] Qin, J. and J. Lawless (1994), Empirical Likelihood and General Estimating Equations, Annals of Statistics, 22, 300-325.

[24] Ramalho, J.J.S. and R.J. Smith (2006), Goodness of Fit Tests for Moment Condition Models, Working Paper, UNIVERSIDADE DE EVORA.

[25] Robinson, P.M. (1988), The Stochastic Difference Between Econometric Statistics, Econometrica, 56, 531-548.

[26] Smith, R.J. (1987), Alternative asymptotically optimal tests and their application to dynamic specification. Review of Economic Studies 54 (1987): 665-680.

[27] Smith, R.J. (1997), Alternative Semi-parametric Likelihood Approaches to Generalized Method of Moments Estimation. Economic Journal 107, 503-519.

[28] Smith, R.J. (2011), GEL Criteria for Moment Condition Models. Econometric Theory, 27, 1192-1235.

[29] Wang, X (2015), A General Approach to Conditional Moment Specification Testing with Projections. Econometric Reviews, forthcoming, DOI:10.1080/07474938.2015.1032165.

[30] Wooldridge, J. (1990), A unified approach to robust, regression-based specification tests. Econometric Theory, 6, 17-43.

n	Nominal Size (%)	$\hat{S}_{GMM}$	$\hat{S}_{CUE}$	$\widehat{TS}_{CUE}$	$\hat{S}_{EL}$	$\widehat{TS}_{EL}$	$\hat{S}_{ET}$	$\widehat{TS}_{ET}$	$\widehat{LM}_{EL}$	$\widehat{TLM}_{EL}$	$\widehat{LM}_{ET}$	$\widehat{TLM}_{ET}$	$\widehat{LM}^{r}_{ET}$	$\widehat{TLM}^{r}_{ET}$	$\widehat{LR}_{EL}$	$\widehat{LR}_{ET}$
200	20.0	25.8	23.4	24.0	26.5	26.4	27.7	27.7	26.5	26.4	24.1	24.2	25.4	25.2	25.6	25.9
	15.0	20.3	18.0	19.0	21.3	21.3	22.8	22.6	21.3	21.3	19.2	19.3	19.6	19.4	21.1	20.7
	10.0	14.9	12.8	14.2	15.5	15.5	17.3	17.2	15.5	15.5	14.5	14.6	13.3	13.2	15.4	15.2
	5.0	8.6	7.5	8.9	9.7	9.6	11.2	10.9	9.7	9.6	9.3	9.4	7.1	7.0	9.1	9.1
	1.0	2.7	2.4	3.9	3.1	3.0	4.8	4.7	3.1	3.0	4.3	4.3	1.8	1.8	3.0	3.5
400	20.0	23.1	22.0	22.2	23.5	23.5	24.7	24.6	23.5	23.5	22.2	22.3	22.5	22.4	23.6	23.1
	15.0	17.6	16.7	17.0	18.3	18.4	19.5	19.5	18.3	17.4	17.2	17.3	17.0	16.7	18.1	17.9
	10.0	12.4	11.6	12.0	13.0	13.0	14.3	14.2	13.0	13.0	12.2	12.3	11.3	11.3	12.8	12.6
	5.0	7.2	6.8	8.3	7.2	7.2	8.6	8.4	7.2	7.2	7.5	7.6	5.7	5.7	7.2	8.5
	1.0	2.2	2.5	2.9	2.0	2.0	3.2	3.1	2.0	2.0	3.0	3.1	1.2	1.2	2.1	2.5
800	20.0	22.2	21.3	21.3	22.8	22.9	23.6	23.7	22.8	22.9	21.5	21.5	22.1	22.1	23.0	22.6
	15.0	16.8	16.3	16.4	17.8	17.8	18.8	18.7	17.8	17.8	16.6	16.6	16.8	16.7	17.7	17.2
	10.0	12.1	11.7	11.8	12.6	12.7	13.5	13.5	12.6	12.7	11.9	11.9	11.4	11.3	12.3	12.3
	5.0	7.0	6.8	7.0	7.2	7.2	8.3	8.3	7.2	7.2	7.2	7.2	5.8	5.8	7.0	7.1
	1.0	2.0	2.4	2.5	1.8	1.8	2.8	2.7	1.8	1.8	2.6	2.7	1.2	1.2	2.0	2.3

Table 1: Finite sample performances of the new class of overidentifying tests in the asset pricing model. $m = 2$. $\hat{\beta} = \hat{\beta}_{GMM}$. $\widehat{TS}_{CUE}$ denotes $\widehat{TS}_{CUE}(\hat{\beta})$. $\widehat{TS}_{EL}$ denotes $\widehat{TS}_{EL}(\hat{\beta})$. $\widehat{TS}_{ET}$ denotes $\widehat{TS}_{ET}(\hat{\beta})$. $\widehat{TLM}_{EL}$ denotes $\widehat{TLM}_{EL}(\hat{\beta})$. $\widehat{TLM}^{r}_{ET}$ denotes $\widehat{TLM}^{r}_{ET}(\hat{\beta})$ and $\widehat{LM}^{r}_{ET}$ denotes $\widehat{LM}^{r}_{ET}(\hat{\beta})$ in which the robust estimate of Ω is employed.

n	Nominal Size (%)	$\hat{S}_{GMM}$	$\hat{S}_{CUE}$	$\widehat{TS}_{CUE}$	$\hat{S}_{EL}$	$\widehat{TS}_{EL}$	$\hat{S}_{ET}$	$\widehat{TS}_{ET}$	$\widehat{LM}_{EL}$	$\widehat{TLM}_{EL}$	$\widehat{LM}_{ET}$	$\widehat{TLM}_{ET}$	$\widehat{LM}^{r}_{ET}$	$\widehat{TLM}^{r}_{ET}$	$\widehat{LR}_{EL}$	$\widehat{LR}_{ET}$
200	20.0	37.2	27.4	29.5	35.1	34.7	38.5	37.6	35.1	34.7	29.9	30.5	30.4	29.8	33.9	32.7
	15.0	32.0	21.8	24.4	29.5	29.1	33.7	32.6	29.5	29.1	25.0	25.5	24.3	23.5	28.3	27.2
	10.0	26.1	15.9	19.0	23.2	22.8	28.1	27.3	23.2	22.8	19.9	20.4	17.8	17.1	22.0	21.3
	5.0	18.9	9.5	13.0	15.8	15.2	21.4	19.4	15.8	15.2	13.8	14.2	10.5	9.9	14.5	14.2
	1.0	8.6	2.7	5.7	6.7	6.2	12.1	10.4	6.7	6.2	6.3	6.7	2.9	2.9	6.0	6.3
400	20.0	31.3	25.0	25.8	29.8	29.7	33.0	32.4	29.8	29.8	26.3	26.4	26.2	25.8	29.2	28.1
	15.0	26.2	20.2	21.0	24.5	24.4	28.0	27.3	24.5	24.5	21.6	21.8	20.6	20.1	23.4	22.9
	10.0	20.5	14.9	15.8	18.4	18.4	22.2	21.3	18.4	18.4	16.5	16.6	14.5	14.0	17.8	17.4
	5.0	13.8	9.1	10.2	12.2	12.0	15.8	14.9	12.2	12.0	10.8	10.9	8.4	7.9	11.0	11.0
	1.0	6.5	3.0	4.2	4.5	4.3	8.0	7.2	4.5	4.3	4.8	4.9	2.3	1.9	4.1	4.4
800	20.0	28.1	24.2	24.4	27.0	27.2	29.5	29.3	27.0	27.2	25.0	25.0	24.6	24.4	26.8	26.3
	15.0	22.8	19.1	19.3	22.0	22.1	24.6	24.3	22.0	22.1	20.0	20.0	19.3	19.1	21.4	20.8
	10.0	17.5	14.2	14.4	16.5	16.6	19.5	19.2	16.5	16.6	15.2	15.2	13.3	13.1	15.8	15.5
	5.0	11.6	8.8	9.1	9.7	9.8	12.8	12.4	9.7	9.8	9.6	9.6	7.1	6.9	9.3	9.6
	1.0	4.7	3.1	3.4	3.0	3.1	5.3	5.0	3.0	3.1	3.8	3.8	1.6	1.5	2.8	3.4

Table 2: Finite sample performances of the new class of overidentifying tests in the asset pricing model. $m = 3$. $\hat{\beta} = \hat{\beta}_{GMM}$. $\widehat{TS}_{CUE}$ denotes $\widehat{TS}_{CUE}(\hat{\beta})$. $\widehat{TS}_{EL}$ denotes $\widehat{TS}_{EL}(\hat{\beta})$. $\widehat{TS}_{ET}$ denotes $\widehat{TS}_{ET}(\hat{\beta})$. $\widehat{TLM}_{EL}$ denotes $\widehat{TLM}_{EL}(\hat{\beta})$. $\widehat{TLM}^{r}_{ET}$ denotes $\widehat{TLM}^{r}_{ET}(\hat{\beta})$ and $\widehat{LM}^{r}_{ET}$ denotes $\widehat{LM}^{r}_{ET}(\hat{\beta})$ in which the robust estimate of Ω is employed.

n	Nominal Size(%)	$\hat{S}_{GMM}$	$\hat{S}_{CUE}$	$\widehat{TS}_{CUE}$	$\hat{S}_{EL}$	$\widehat{TS}_{EL}$	$\hat{S}_{ET}$	$\widehat{TS}_{ET}$	$\widehat{LM}_{EL}$	$\widehat{TLM}_{EL}$	$\widehat{LM}_{ET}$	$\widehat{TLM}_{ET}$	$\widehat{LM}^{r}_{ET}$	$\widehat{TLM}^{r}_{ET}$	$\widehat{LR}_{EL}$	$\widehat{LR}_{ET}$
500	20.0	24.8	24.8	24.8	26.2	26.2	27.1	27.1	26.2	26.2	24.7	24.7	24.8	24.8	26.3	26.7
	15.0	20.9	20.9	20.9	22.0	22.1	22.5	22.5	22.0	22.1	20.8	20.8	20.2	20.6	22.1	21.2
	10.0	16.3	16.3	16.3	16.2	16.3	17.3	17.3	16.2	16.3	16.2	16.2	14.5	14.5	16.1	16.9
	5.0	11.9	11.9	11.9	10.1	10.2	11.4	11.6	10.1	10.2	11.9	11.9	7.5	7.8	10.4	11.1
	1.0	7.4	7.4	7.4	2.9	3.1	4.0	4.2	2.9	3.1	7.4	7.4	1.6	1.6	3.4	5.2
1000	20.0	23.7	23.7	23.7	23.4	23.4	24.6	24.6	24.6	23.4	23.6	23.6	22.3	22.3	24.6	23.7
	15.0	19.6	19.6	19.6	18.0	18.0	19.0	19.0	18.0	18.0	19.4	19.4	16.7	16.7	18.1	19.1
	10.0	13.5	13.5	13.5	12.6	12.6	13.5	13.5	12.6	12.6	13.2	13.2	11.3	11.3	13.2	14.0
	5.0	8.3	8.3	8.3	6.8	7.1	8.2	8.2	6.8	7.1	8.2	8.2	5.1	5.1	6.9	7.1
	1.0	3.7	3.7	3.7	2.0	2.0	2.8	2.8	2.0	2.0	3.6	3.6	1.0	1.0	2.0	2.4

Table 3: Finite sample performances of the new class of overidentifying tests of chi-squared moments model. $\hat{\beta} = \hat{\beta}_{GMM}$. $\widehat{TS}_{CUE}$ denotes $\widehat{TS}_{CUE}(\hat{\beta})$. $\widehat{TS}_{EL}$ denotes $\widehat{TS}_{EL}(\hat{\beta})$. $\widehat{TS}_{ET}$ denotes $\widehat{TS}_{ET}(\hat{\beta})$. $\widehat{TLM}_{EL}$ denotes $\widehat{TLM}_{EL}(\hat{\beta})$. $\widehat{TLM}^{r}_{ET}$ denotes $\widehat{TLM}^{r}_{ET}(\hat{\beta})$ and $\widehat{LM}^{r}_{ET}$ denotes $\widehat{LM}^{r}_{ET}(\hat{\beta})$ in which the robust estimate of Ω is employed.