Goodness of Fit Tests for Moment Condition Models

Joaquim J.S. Ramalho
Department of Economics
University of Evora

Richard J. Smith
*c*emmap
I.F.S. and U.C.L.
and
Department of Economics
University of Warwick

Abstract

This paper proposes novel methods for the construction of tests for models specified by unconditional moment restrictions. It exploits the classical-like nature of generalized empirical likelihood (GEL) to define Pearson-type statistics for over-identifying moment conditions and parametric constraints based on constrasts of GEL implied probabilities which are natural by-products of GEL estimation. As is increasingly recognized, GEL can possess both theoretical and empirical advantages over the more standard generalized method of moments (GMM). Monte Carlo evidence comparing GMM, GEL and Pearson-type statistics for over-identifying moment conditions indicates that the size properties of a particular Pearson-type statistic is competitive in most and an improvement over other statistics in many circumstances.

JEL Classification: C13, C30

Keywords: GMM, Generalized Empirical Likelihood, Overidentifying Moments, Parametric Restrictions, Pearson-Type Tests.

1 Introduction

This paper proposes novel methods for the construction of tests for models specified by unconditional moment restrictions. The generalized method of moments (GMM), Hansen (1982), is the conventional method of fit for such models. In view of increasing Monte Carlo evidence indicating that GMM estimators may be badly biased in finite samples and that the empirical and nominal size of associated tests may differ substantially, see, for example, the Special Issue of the *Journal of Business & Economic Statistics* (July 1996), a number of alternative estimators which are asymptotically first-order equivalent to efficient GMM have been suggested. These estimators include empirical likelihood (EL) [Qin and Lawless (1994), Imbens (1997), Owen (2001)], exponential tilting (ET) [Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998)] and the continuous updating estimator (CUE) [Hansen, Heaton and Yaron (1996)].

These estimators share a common structure, being members of a class of generalized empirical likelihood (GEL) estimators [Newey and Smith (2004) and Smith (1997, 2001)]. GEL estimation seems to possess many attractive theoretical features relative to GMM. Large sample analysis, Newey and Smith (2004), indicates that GEL estimators may be less prone to bias than those based on GMM. GEL also appears to have diverse advantages over GMM in finite samples. Imbens (1997) and Newey, Ramalho and Smith (2002) report promising Monte Carlo results concerning the small sample bias of GEL estimators, while Imbens, Spady and Johnson (1998) find that particular GEL tests of overidentifying moment conditions, although also oversized in finite samples, possess actual sizes closer to nominal size than Hansen's (1982) test.

GEL bears certain similarities to likelihood-based methods, allowing the construction of classical-type tests of hypotheses in the moment condition framework. These include overidentifying moment conditions, for which only Hansen's (1982) test is typically available in the GMM setting. This paper exploits the classical-like feature of GEL and proposes new specification tests for moment condition models similar in spirit to the standard Pearson tests for goodness of fit. In particular, a set of implied or empirical probabilities which incorporate the moment condition information are associated with each GEL estimator, which by reweighting the data impose exactly all moment conditions on the sample, rather than particular linear combinations as in the GMM case. See Newey and Smith (2004). Implied probabilities based on GMM may also be be constructed in a likewise fashion by utilising the GEL criterion function evaluated at an efficient GMM estimator as discussed in Brown and Newey (1992, 2003). The resultant GEL distribution function estimator formed from the implied probabilities is an efficient estimator of the distribution of the data, in particular, it dominates the empirical distribution function (EDF) implicitly used by GMM. Contrasts between GEL implied and EDF probabilities allow the construction of classical Pearson-type tests of over-identifying moment conditions. A similar approach can be used to construct tests for parametric restrictions based on contrasts of restricted and unrestricted GEL implied probabilities.

In a set of Monte Carlo experiments based on those considered in Imbens, Spady and Johnson (1998), we compare the finite sample size behaviour of Pearson-type statistics for over-identifying moment conditions with other existing GMM and GEL tests, such as Hansen's (1982) test and those proposed in Smith (1997).

This paper is organized as follows. Section 2 briefly reviews GMM and GEL estimation. Pearson-type tests for over-identifying moment conditions are presented in section 3 while parametric restrictions are considered in section 4. The Monte Carlo experiments are discussed in section 5. Section 6 concludes. Proofs of the results contained in the paper are provided in the Appendix.

2 The Model and Estimators

This section briefly reconsiders the model and estimators. The set-up considered and notation used here is similar to that in Newey and Smith (2004), which is henceforth abbreviated as NS.

Let z_i, $(i = 1, ..., n)$, denote independent and identically distributed observations on the k-vector z. Also, let $g(z, \beta)$ be an m-vector of known functions of the data observation z and the p-vector of parameters β, where $m \geq p$. The model has a true parameter β_0 satisfying the unconditional moment condition

$$E[g(z, \beta_0)] = 0, \tag{2.1}$$

where $E[.]$ denotes expectation taken with respect to the distribution of z.

Various methods of estimation have been proposed for models specified by moment conditions of the type (2.1). The standard method is two-step GMM estimation, see Hansen (1982). Let $g_i(\beta) \equiv g(z_i, \beta)$, $\hat{g}(\beta) \equiv \sum_{i=1}^n g_i(\beta)/n$ and $\hat{\Omega}(\beta) \equiv \sum_{i=1}^n g_i(\beta)g_i(\beta)'/n$ or the centred estimator $\sum_{i=1}^n [g_i(\beta) - \hat{g}(\beta)][g_i(\beta) - \hat{g}(\beta)]'/n$. Also, let $\tilde{\beta}$ be some preliminary estimator given by $\tilde{\beta} = \arg\min_{\beta \in \mathcal{B}} \hat{g}(\beta)'\hat{W}^{-1}\hat{g}(\beta)$ where $\mathcal{B}$ denotes the parameter space and $\hat{W}$ is a random matrix with properties to be specified below. The two-step efficient GMM estimator is defined by

$$\hat{\beta}_{GMM} = \arg\min_{\beta \in \mathcal{B}} \hat{g}(\beta)'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\beta). \tag{2.2}$$

Alternative estimation methods which share the first order asymptotic properties of two-step GMM are those in the generalized empirical likelihood (GEL) class, as in NS and Smith (1997, 2001). To describe them let $\rho(v)$ be a function of a scalar v that is concave on its domain, an open interval $\mathcal{V}$ containing zero with derivatives $\rho_j(v) = \partial^j \rho(v)/\partial v^j$ and $\rho_j = \rho_j(0)$, $(j = 0, 1, ...)$. Also let $\hat{\Lambda}_n(\beta) = \{\lambda : \lambda' g_i(\beta) \in \mathcal{V}, i = 1, ..., n\}$. The GEL estimator is the solution to a saddle point problem

$$\hat{\beta}_{GEL} = \arg\min_{\beta \in \mathcal{B}} \sup_{\lambda \in \hat{\Lambda}_n(\beta)} \hat{P}(\beta, \lambda), \tag{2.3}$$

where $\hat{P}(\beta, \lambda) = \sum_{i=1}^n \rho(\lambda' g_i(\beta))/n$. Each of the elements of the m-vector λ of auxiliary parameters is associated with an element of the moment indicator vector $g_i(\beta)$ and may be interpreted as Lagrange multipliers for the sample moment constraint $\sum_{i=1}^n \rho_1(\lambda' g_i(\beta))g_i(\beta) = 0$. We define the optimal auxiliary parameter estimator

$$\hat{\lambda} = \arg\max_{\lambda \in \hat{\Lambda}_n(\hat{\beta})} \hat{P}(\hat{\beta}, \lambda). \tag{2.4}$$

Let $\hat{\lambda}(\beta) = \arg\max_{\lambda \in \hat{\Lambda}_n(\beta)} \hat{P}(\beta, \lambda)$.

The GEL class includes as special cases the empirical likelihood (EL) estimator, $\rho(v) = \log(1 - v)$ and $\mathcal{V} = (-\infty, 1)$, (Qin and Lawless, 1994, Imbens, 1997, and Smith, 1997), and the exponential tilting (ET) estimator, $\rho(v) = -\exp(v)$, (Kitamura and Stutzer, 1997, Imbens, Spady and Johnson, 1998, and Smith, 1997). The continuous updating estimator (CUE) of Hansen, Heaton and Yaron (1996) $\hat{\beta}_{CUE} = \arg\min_{\beta \varepsilon \mathcal{B}} \hat{g}(\beta)'\hat{\Omega}(\beta)^{-}\hat{g}(\beta)$, where A^{-} denotes any generalized inverse of a matrix A satisfying $AA^{-}A = A$, is also a special case with $\rho(v)$ quadratic as are members of the Cressie and Read (1984) power divergence family of discrepancies, $\rho(v) = -(1+\gamma v)^{(\gamma+1)/\gamma}/(\gamma+1)$, see NS, Theorem 2.2.

We impose the following innocuous normalization on $\rho(v)$. We set $\rho_1 = \rho_2 = -1$. If $\rho_1 \neq 0$ and $\rho_2 < 0$, this normalization can always be imposed by replacing $\rho(v)$ by $[-\rho_2/\rho_1^2]\rho([\rho_1/\rho_2]v)$. It does not affect the estimator of β and renders the estimator for λ comparable for different choices of $\rho(v)$. It is satisfied by the $\rho(v)$ given above for CUE, EL, ET and Cressie and Read (1984) discrepancies.

In the following because of their first order asymptotic equivalence, the notation $\hat{\beta}$ is used to denote both efficient GMM and GEL estimators of β_0. Consistency of $\hat{\beta}$ is obtained under the following identification and regularity conditions; for GEL, see Theorem 3.1 of NS. Let $\Omega(\beta) \equiv E[g_i(\beta)g_i(\beta)']$ or in the centred case $E[g_i(\beta)g_i(\beta)'] - E[g_i(\beta)]E[g_i(\beta)]'$ and $\Omega \equiv \Omega(\beta_0)$.

Assumption 2.1 *There exists W such that $\hat{W} = W + o_p(1)$ and W is positive definite.*

This assumption is only required by GMM which together with the next assumption ensures the consistency of the preliminary estimator $\tilde{\beta}$.

Assumption 2.2 *(a) $\beta_0 \in \mathcal{B}$ is the unique solution to $E[g(z, \beta)] = 0$; (b) $\mathcal{B}$ is compact; (c) $g(z, \beta)$ is continuous at each $\beta \in \mathcal{B}$ with probability one; (d) $E\left[\sup_{\beta \in \mathcal{B}} \|g(z, \beta)\|^{\alpha}\right] < \infty$ for some $\alpha > 2$; (e) Ω is nonsingular; (f) $\rho(v)$ is twice continuously differentiable in a neighborhood of zero.*

The restriction on the parameter α may be set to the weak inequality $\alpha \geq 2$ for GMM. Assumption 2.2 also implies $\hat{g}(\hat{\beta}) = O_p(n^{-1/2})$, $\hat{\lambda}$ (2.4) exists w.p.a.1 and $\hat{\lambda} = O_p(n^{-1/2})$.

The following additional conditions are needed for asymptotic normality. Let $G(\beta) = E[\partial g_i(\beta)/\partial\beta]$ and $G = G(\beta_0)$.

Assumption 2.3 *(a) $\beta_0 \in int(\mathcal{B})$; (b) $g(z,\beta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of β_0 and $E[\sup_{\beta\in\mathcal{N}} \|\partial g_i(\beta)/\partial\beta'\|] < \infty$; (c) $rank(G) = p$.*

Let $\Sigma = (G'\Omega^{-1}G)^{-1}$, $H = \Sigma G'\Omega^{-1}$, and $P = \Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1}$. If Assumptions 2.1-2.3 hold,

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Sigma),$$
$$n^{1/2}\hat{\lambda} \xrightarrow{d} N(0, P),$$

and are asymptotically independent. Moreover, defining the normalised and centred optimised GEL criterion as $GELR_n = 2n[\hat{P}(\hat{\beta},\hat{\lambda}) - \rho_0]$, we have

$$GELR_n \xrightarrow{d} \chi^2(m-p).$$

See Theorem 3.2 of NS.

3 Goodness of Fit Tests for Over-Identifying Moment Conditions

In the GMM and GEL frameworks there are several ways of assessing the validity of the over-identifying moment conditions (2.1). Classical-like GEL statistics, suggested by Smith (1997, 2001), also see Imbens, Spady and Johnson (1998) and Kitamura and Stutzer (1997), are the GEL criterion function statistic given above

$$GELR_n = 2n[\hat{P}(\hat{\beta},\hat{\lambda}) - \rho_0], \tag{3.1}$$

the Lagrange multiplier form

$$LM_n = n\hat{\lambda}'\hat{\Omega}(\hat{\beta})\hat{\lambda}, \tag{3.2}$$

and the score statistic

$$S_n = n\hat{g}(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\hat{g}(\hat{\beta}). \tag{3.3}$$

The last statistic is of course identical in form to Hansen's (1982) GMM test statistic for over-identifying moment restrictions. Given the asymptotic equivalence between GMM and GEL estimators, these statistics may also be equivalently evaluated at an efficient GMM estimator defining $\hat{\lambda}$ as in (2.4) above. If Assumption 2.2 is satisfied the matrix $\hat{\Omega}(\beta)$ evaluated at a consistent estimator for β_0 is a consistent estimator for Ω. Consequently, $GELR_n$, LM_n and S_n are asymptotically equivalent and thus from above possess a chi-square limiting distribution with $m-p$ degrees of freedom.

This section considers alternative statistics for testing the moment conditions (2.1) based on implied probabilities $\hat{\pi}_i$ (3.4), $(i=1,...,n)$, and an associated GEL distribution function estimator $\hat{\mu}_n(\cdot)$ (3.5) defined below.

3.1 Implied Probabilities

Implied or empirical probabilities for the observations which incorporate the moment restrictions (2.1) may be associated with each GMM and GEL estimator. These probabilities form the basis for the statistics developed below so we briefly describe them here. For a given function $\rho(v)$, an associated efficient GMM or GEL estimator $\hat{\beta}$ and $\hat{g}_i \equiv g_i(\hat{\beta})$, they are given by

$$\hat{\pi}_i = \rho_1(\hat{\lambda}'\hat{g}_i)/\sum_{j=1}^{n}\rho_1(\hat{\lambda}'\hat{g}_j), (i=1,...,n), \tag{3.4}$$

where $\hat{\lambda}$ is defined in (2.4). The empirical probabilities $\hat{\pi}_i$, $(i=1,...,n)$, sum to one by construction and are positive when $\hat{\lambda}'\hat{g}_i$ is small uniformly in i as is the case with probability approaching 1, see Lemma A1 of NS. Moreover, they impose the sample moment condition $\sum_{i=1}^{n}\pi_i(\beta,\lambda)g_i(\beta)=0$, where $\pi_i(\beta,\lambda)=\rho_1(\lambda' g_i(\beta))/\sum_{j=1}^{n}\rho_1(\lambda' g_j(\beta))$, $(i=1,...,n)$, when the first-order conditions for λ hold, mirroring the population moment condition (2.1). For EL the implied probabilities were given by Owen (1988), for ET by Kitamura and Stutzer (1997), for quadratic $\rho(v)$ by Back and Brown (1993), and for the general case by Brown and Newey (1992). Also see Brown and Newey (2003), NS and Smith (1997, 2001).

For any function $a(z,\beta)$ and efficient GMM or GEL estimator $\hat{\beta}$ the implied probabilities can be used to form an efficient estimator $\sum_{i=1}^{n}\hat{\pi}_i a(z_i,\hat{\beta})$ of the expectation $E[a(z,\beta_0)]$ as in Brown and Newey (1998). Of particular interest here is the cumulative distribution function $\mu(z)=\mathcal{P}\{z_i\leq z\}$ of the observation vector z which may also be written in expectation form as $\mu(z)=E[1(z_i\leq z)]$, where $1(.)$ denotes the indicator function, $1(z_i\leq z)=1$ if $z_i\leq z$ and 0 otherwise. The efficient estimator for the observation distribution function $\mu(\cdot)$ obtained from the implied probabilities $\hat{\pi}_i$, $(i=1,...,n)$, defined in (3.4), is therefore given by

$$\hat{\mu}_n(z)=\sum_{i=1}^{n}\hat{\pi}_i 1(z_i\leq z). \tag{3.5}$$

In particular, $\hat{\mu}_n(z)$ is a more efficient estimator for $\mu(z)$ than the empirical distribution function (EDF)

$$\mu_n(z)=\sum_{i=1}^{n}1(z_i\leq z)/n. \tag{3.6}$$

It is well known that when z is univariate and continuous the empirical process $n^{1/2}[\mu_n(z)-\mu(z)]$ weakly converges to a Brownian bridge, a Gaussian process with mean zero and covariance function $\mu(z_1)\wedge\mu(z_2)-\mu(z_1)\mu(z_2)$, see, for example, Durbin (1973) and Shorack and Wellner (1986). We need to develop a similar result for the normalised contrast $n^{1/2}[\hat{\mu}_n(z)-\mu_n(z)]$ between the GEL distribution function estimator and the EDF to obtain a particular form of Pearson-type test statistic for the over-identifying moment restrictions (2.1). Let Z denote the sample space of z and also let

$$n^{1/2}[\hat{\mu}_n(z)-\mu_n(z)]\equiv\hat{\Lambda}_n(z), z\in Z.$$

Lemma 3.1 *If Assumptions 2.1-2.3 are satisfied then $\hat{\Lambda}_n\Rightarrow\hat{\Lambda}$ where $\hat{\Lambda}$ is a Gaussian process on Z with zero mean and covariance function $E[\hat{\Lambda}(z_1)\hat{\Lambda}(z_2)]=b(z_1)'Pb(z_2)$ where $b(z)=E[1(z_i\leq z)g_i(\beta_0)]$.*

3.2 Pearson-Type Tests

Suppose that the sample z_i, $(i=1,...,n)$, is drawn from a discrete distribution with support $(z^1,...,z^s)$ and that the distinct value z^j arises $n_j\geq 1$ times. In a parametric context, we may wish to test whether a given distribution function $\mu(z^j) = \mathcal{P}\{z = z^j\}$, $(j = 1, ..., s)$, correctly characterizes the distribution of z. To this end, two versions of the Pearson statistic are usually applied, *viz.* $\sum_{j=1}^{s}(n\mu(z^j) - n_j)^2/n_j$ and $\sum_{j=1}^{s}(n\mu(z^j) - n_j)^2/n\mu(z^j)$, where n_j and $n\mu(z^j)$ are, respectively, the actual and expected numbers of observations of the distinct value z^j, $(j = 1, ..., s)$, under the assumed distribution $\mu(\cdot)$. For the latter statistic it is assumed that $\mu(z^j) > 0$ for all $j = 1, ..., s$. If the distribution $\mu(\cdot)$ is correctly specified, then differences between the observed and expected numbers of outcomes arise solely because of random fluctuations. Both statistics are asymptotically equivalent and have a limiting chi-square distribution with $s - 1$ degrees of freedom.

In the GEL framework, we can dispense with the assumption of a discrete distribution and instead think in terms of probabilities associated with individual observations; see *inter alia* Owen (2001). In other words, we proceed as if a single data point was observed in each cell of a n-cell contingency table. That is, GEL versions of the above statistics may be obtained directly by setting $s = n$, $n_j = 1$, $z^j = z_j$ and $\mu(z^j) = \hat{\pi}_j$, $(j = 1, ..., n)$. The consequent versions of the standard Pearson statistics to test the moment restrictions (2.1) are based on the normalised contrasts $n\hat{\pi}_i - 1$, $(i = 1, ..., n)$, comparing predicted probabilities from the GEL distribution function $\hat{\mu}_n(\cdot)$ and those from the unrestricted EDF $\mu_n(\cdot)$; *viz.*

$$P_n^a = \sum_{i=1}^{n} (n\hat{\pi}_i - 1)^2 \tag{3.7}$$

and

$$P_n^b = \sum_{i=1}^{n} \frac{(n\hat{\pi}_i - 1)^2}{n\hat{\pi}_i}. \tag{3.8}$$

Theorem 3.1 *If Assumptions 2.1-2.3 are satisfied then P_n^a and P_n^b are asymptotically equivalent to $GELR_n$, LM_n and S_n. Therefore P_n^a, $P_n^b \xrightarrow{d} \chi^2_{m-p}$.*

Therefore, an α asymptotic level test of the over-identifying moment restrictions (2.1) has critical region $\{P_n \geq \chi^2_{m-p}(\alpha)\}$ where P_n is P_n^a or P_n^b and $\chi^2_{m-p}(\alpha)$ denotes the $1 - \alpha$ quantile from the chi-square distribution with $m - p$ degrees of freedom.

Alternative forms of Pearson-type tests for the over-identifying moment conditions (2.1) may be based on a discretization of the distribution of z obtained by employing a finite partition of the sample space Z. These statistics are similar in spirit to those discussed by Andrews (1988a, 1988b) but are adapted for the moment condition setting considered here. As shown in Lemma 3.1 above, the distribution function $\mu(\cdot)$ of the data observation vector z is consistently estimated under (2.1) by both the moment restricted estimator $\hat{\mu}_n(\cdot)$ of (3.5) and the EDF $\mu_n(\cdot)$ of (3.6). Test statistics for the validity of the over-identifying moment conditions (2.1) proposed below exploit this result and are based on quadratic forms suitably defined in terms of the contrast $\hat{\mu}_n(\cdot) - \mu_n(\cdot)$.

Let the sample space Z of z be partitioned into the subsets Z_j, $(j = 1, 2, ...)$. Consider the (arbitrary) finite collection of subsets Z_j, $(j = 1, ..., s)$, whose union may not equal Z, that is, $\cup_{j=1}^{s} Z_j \subset Z$. We impose the order condition $s \geq m$ and require $\mu(Z_j) > 0$, $(j = 1, ..., s)$. Define

$$\hat{\mu}_n(Z_j) = \sum_{i=1}^{n} \hat{\pi}_i 1\left(z_i \in Z_j\right) \tag{3.9}$$

and

$$\mu_n(Z_j) = \sum_{i=1}^{n} 1\left(z_i \in Z_j\right)/n. \tag{3.10}$$

Because the choice of the collection $\{Z_j\}_{j=1}^{s}$ is arbitrary, an advantage of this approach is that these subsets Z_j, $(j = 1, ..., s)$, may be chosen judiciously by the researcher to explore the validity of the moment restrictions (2.1). Andrews (1988a, 1988b) provides an extensive discussion and references for such choices in a fully parametric setting. However, unlike there, we restrict ourselves to consideration only of a non-stochastic partition Z_j, $(j = 1, 2, ...)$, for ease of exposition. This assumption may be relaxed though but at the expense of some additional complexity by adopting the approach used in Andrews (1988b). This would permit a random partition which would weakly converge to one with the properties ascribed below for Z_j, $(j = 1, 2, ...)$. See Andrews (1988b, Assumption RC1, p.1425, and Section 3.1, pp.1427-1431).

Let $\hat{\mu}_n^s = (\hat{\mu}_n(Z_1), ..., \hat{\mu}_n(Z_s))'$ and $\mu_n^s = (\mu_n(Z_1), ..., \mu_n(Z_s))'$. Also let $B_s = (b(Z_1), ..., b(Z_s))$

where $b(Z_j) = E[1(z \in Z_j)g(z, \beta_0)]$, $(j = 1, ..., s)$. The test statistics defined below are based on the normalised contrast $\hat{\mu}_n^s - \mu_n^s$ from (3.9) and (3.10). It follows immediately from Lemma 3.1 that $n^{1/2}(\hat{\mu}_n^s - \mu_n^s) \xrightarrow{d} N(0, B_s' P B_s)$. Now if B_s is full row rank m then $B_s'(B_s B_s')^{-1}\Omega(B_s B_s')^{-1}B_s$ is a g-inverse for $B_s' P B_s$. Therefore, we consider the statistic[1]

$$P_n^{alt} = n(\hat{\mu}_n^s - \mu_n^s)'\hat{B}_s'(\hat{B}_s\hat{B}_s')^{-1}\hat{\Omega}(\hat{B}_s\hat{B}_s')^{-1}\hat{B}_s(\hat{\mu}_n^s - \mu_n^s), \tag{3.11}$$

where $\hat{B}_s = (\hat{b}(Z_1), ..., \hat{b}(Z_s))$, $\hat{b}(Z_j) = \sum_{i=1}^n 1(z \in Z_j)\hat{g}_i/n$ or $\sum_{i=1}^n \hat{\pi}_i 1(z \in Z_j)\hat{g}_i$, $(j = 1, ..., s)$, and $\hat{\Omega} = \sum_{i=1}^n \hat{g}_i\hat{g}_i'/n$, $\sum_{i=1}^n [\hat{g}_i - \hat{g}][\hat{g}_i - \hat{g}]'/n$, $\hat{g} = \hat{g}(\hat{\beta})$, or $\sum_{i=1}^n \hat{\pi}_i\hat{g}_i\hat{g}_i'$.

Theorem 3.2 *If Assumptions 2.1-2.3 are satisfied and* $rk(B_s) = m$ *then the statistic* P_n^{alt} *is asymptotically equivalent to* $GELR_n$, LM_n, S_n *and* P_n^a, P_n^b. *Therefore* $P_n^{alt} \xrightarrow{d} \chi^2_{m-p}$.

An α asymptotic level test of the over-identifying moment restrictions (2.1) has critical region $\{P_n^{alt} \geq \chi^2_{m-p}(\alpha)\}$. If in addition $s = m$ then B_s is nonsingular so that $B_s^{-1}\Omega B_s'^{-1}$ is a g-inverse for $B_s' P B_s$.

Corollary 3.1 *If Assumptions 2.1-2.3 are satisfied,* $rk(B_s) = m$ *and* $s = m$ *then the statistic* $P_n^{alt} = n(\hat{\mu}_n^s - \mu_n^s)'\hat{B}_s^{-1}\hat{\Omega}\hat{B}_s'^{-1}(\hat{\mu}_n^s - \mu_n^s)$ *is asymptotically equivalent to* $GELR_n$, LM_n, S_n *and* P_n^a, P_n^b. *Therefore* $P_n^{alt} \xrightarrow{d} \chi^2_{m-p}$.

Limiting distributional and asymptotic equivalence results between P_n^a, P_n^b and P_n^{alt} similar to those described above may be shown under the local alternatives $H_n : E_n[g(z_i, \beta_0)] = n^{-1/2}\eta + o(n^{-1/2})$, $(i = 1, ..., n)$, $n = 1, 2, ...$ Then, $n^{1/2}\hat{g}(\beta_0) \xrightarrow{d} N(\eta, \Omega)$ under H_n and consistency of the GEL and auxiliary parameter estimators $\hat{\beta}$ and $\hat{\lambda}$ for β_0 and 0 still obtains. Moreover, the expansions $n^{1/2}(\hat{\beta} - \beta_0) = -\Sigma G'\Omega^{-1}n^{1/2}\hat{g}(\beta_0) + o_p(1)$ and $n^{1/2}\hat{\lambda} = -Pn^{1/2}\hat{g}(\beta_0) + o_p(1)$ remain valid under H_n. Therefore, the statistics P_n^a, P_n^b and P_n^{alt} are asymptotically equivalent to $GELR_n$, LM_n, S_n and converge in distribution to a

[1] More generally the limiting distribution of the statistic $n(\hat{\mu}_n^s - \mu_n^s)'\hat{\Xi}^-(\hat{\mu}_n^s - \mu_n^s)$, where $\hat{\Xi}^-$ denotes a consistent estimator for a g-inverse of $B_s' P B_s$, is that of a chi-square random variable with $rk(B_s' P B_s)$ degrees of freedom.

non-central chi-square random variable with $m-p$ degrees of freedom and non-centrality parameter $\eta' P\eta$.

We conclude this section by briefly considering the consistency of the tests P_n^a, P_n^b and P_n^{alt}. As detailed in section 2, the GEL criterion is optimised with respect to λ such that $\lambda' g(z_i, \beta) \in \mathcal{V}$, $(i = 1, ..., n)$. Therefore, because $\mathcal{V}$ is bounded, $\rho(\beta, \lambda) = E[\rho(\lambda' g(z, \beta))|\lambda' g(z, \beta) \in \mathcal{V}]$ exists and so by a uniform weak law of large numbers $\hat{P}(\beta, \lambda) \xrightarrow{p} \rho(\beta, \lambda)$ uniformly $\beta \in \mathcal{B}$ and λ with $\rho(\beta, \lambda)$ continuous in $\beta \in \mathcal{B}$ and λ. Let $\lambda(\beta) = \arg\max_\lambda \rho(\beta, \lambda)$, $\beta \in \mathcal{B}$. For GMM $\hat{g}(\beta)'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\beta) \xrightarrow{p} g(\beta)'\Omega(\beta_{**})^{-1}g(\beta)$ uniformly $\beta \in \mathcal{B}$ where $\tilde{\beta} \xrightarrow{p} \beta_{**}$.

Assumption 3.1 *(a) no $\beta \in \mathcal{B}$ exists such that $E[g(z, \beta)] = 0$; (b) $\mathcal{B}$ is compact; (c) $g(z, \beta)$ is continuous at each $\beta \in \mathcal{B}$ with probability one; (d) $E\left[\sup_{\beta\in\mathcal{B}} \|g(z, \beta)\|^\alpha\right] < \infty$ for some $\alpha > 2$; (e) $\Omega(\beta_{**})$ is nonsingular; (f) $\rho(v)$ is twice continuously differentiable on $\mathcal{V}$; (g) $\lambda(\beta)$ is the unique maximiser of $\rho(\beta, \lambda)$ and is continuous in $\beta \in \mathcal{B}$; (h) β_* is the unique minimiser in $\mathcal{B}$ of $\rho(\beta, \lambda(\beta))$ or $g(\beta)'\Omega(\beta_{**})^{-1}g(\beta)$.*

Assumptions 3.1(g)(h) are convenient high level assumptions made to simplify the exposition. Uniqueness of $\lambda(\beta)$ is required for the consistency of $\hat{\lambda}(\beta) = \arg\max_{\lambda\in\hat{\Lambda}_n(\beta)} \hat{P}(\beta, \lambda)$ for $\lambda(\beta)$. Continuity of $\lambda(\beta)$ and uniqueness of β_* guarantee consistency of the GEL estimator $\hat{\beta}$ for β_*. Now, $E[\rho_1(\lambda(\beta)' g(z, \beta))g(z, \beta)|\lambda(\beta)' g(z, \beta) \in \mathcal{V}] = 0$ from the first order conditions determining $\hat{\lambda}(\beta)$. Therefore, $\lambda(\beta) \neq 0$ for all $\beta \in \mathcal{B}$ otherwise a contradiction with Assumption 3.1(a) would result. In particular, $\lambda_* \equiv \lambda(\beta_*)$ is non-zero. We are now able to establish the consistency of tests based on the statistics P_n^a and P_n^b.

Theorem 3.3 *If Assumptions 2.1 and 3.1 are satisfied then P_n^a, $P_n^b \xrightarrow{p} \infty$.*

For the consistency of P_n^{alt} we require additional assumptions as in Andrews (1988b, Section 4.2). Let $b_*(Z_j) = E[1(z \in Z_j)g(z, \beta_*)|\lambda_*' g(z, \beta_*) \in \mathcal{V}]$ or $E[\rho_1(\lambda_*' g(z, \beta_*))1(z \in Z_j)g(z, \beta_*)|\lambda_*' g(z, \beta_*) \in \mathcal{V}]/\rho_1^*$, $(j = 1, ..., s)$, and $B_{s*} = (b_*(Z_1), ..., b_*(Z_s))$, where $\rho_1^* = E[\rho_1(\lambda_*' g(z, \beta_*))|\lambda_*' g(z, \beta_*) \in \mathcal{V}]$. Also let $\Omega_* = E[g(z, \beta_*)g(z, \beta_*)'|\lambda_*' g(z, \beta_*) \in \mathcal{V}]$, $E[(g(z, \beta_*) - g_*)(g(z, \beta_*) - g_*)'|\lambda_*' g(z, \beta_*) \in \mathcal{V}]$, $g_* = E[g(z, \beta_*)|\lambda_*' g(z, \beta_*) \in \mathcal{V}]$, or

$E[\rho_1(\lambda_*'g(z,\beta_*))g(z,\beta_*)g(z,\beta_*)'|\lambda_*'g(z,\beta_*) \in \mathcal{V}]/\rho_1^*$. Then $\hat{B}_s \xrightarrow{p} B_{s*}$ and $\hat{\Omega} \xrightarrow{p} \Omega_*$. Define $\delta_{*j} = E[(\rho_1(\lambda_*'g(z,\beta_*)) - \rho_1^*)1\,(z \in Z_j)\,|\lambda_*'g(z,\beta_*) \in \mathcal{V}]/\rho_1^*$, $(j = 1,...,s)$, and $\delta_* = (\delta_{*1},...,\delta_{*s})'$.

Theorem 3.4 *If Assumptions 2.1 and 3.1 are satisfied, $rk(B_{s*}) = m$ and $\Omega_*(B_{s*}B_{s*}')^{-1}B_{s*}\delta_* \neq 0$, then $P_n^{alt} \xrightarrow{p} \infty$.*

The condition $\Omega_*(B_{s*}B_{s*}')^{-1}B_{s*}\delta_* \neq 0$ is critical for test consistency and requires that δ_* does not lie in the null space of $\Omega_*(B_{s*}B_{s*}')^{-1}B_{s*}$. If $rk(\Omega_*) = m$, then this condition may be abbreviated to $B_{s*}\delta_* \neq 0$. If $s = m$ as in Corollary 3.1 and $\hat{B}_s^{-1}\hat{\Omega}\hat{B}_s'^{-1}$ replaces $\hat{B}_s'(\hat{B}_s\hat{B}_s')^{-1}\hat{\Omega}(\hat{B}_s\hat{B}_s')^{-1}\hat{B}_s$ in the definition of P_n^{alt} (3.11), the consistency condition $\Omega_*B_{s*}'^{-1}\delta_* \neq 0$ [or $\delta_* \neq 0$ if $rk(\Omega_*) = m$] should be substituted for $\Omega_*(B_{s*}B_{s*}')^{-1}B_{s*}\delta_* \neq 0$ of Theorem 3.4.

4 Goodness of Fit Tests for Parametric Restrictions

This section adapts the goodness of fit statistics of the previous section to test the parametric restrictions defined by the null hypothesis

$$H_0 : r\,(\beta_0) = 0, \tag{4.1}$$

where $r(.)$ is a r-vector of functions.

The following assumptions modify Assumptions 2.2 and 2.3 appropriately for the results of this section and are adapted from Smith (2001).

Assumption 4.1 *(a) $\beta_0 \in \mathcal{B}$ is the unique solution to $E[g(z,\beta)] = 0$ and $r(\beta) = 0$; (b) $\mathcal{B}$ is compact; (c) $g(z,\beta)$ and $r(\beta)$ are continuous at each $\beta \in \mathcal{B}$ with probability one; (d) $E\{\sup_{\beta\in\mathcal{B}} \|g(z,\beta)\|^\alpha\} < \infty$ for some $\alpha > 2$; (e) Ω is nonsingular; (f) $\rho(v)$ is twice continuously differentiable in a neighborhood of zero.*

Let $R(\beta) = \partial r(\beta)/\partial\beta'$ and $R = R(\beta_0)$.

Assumption 4.2 *(a) $\beta_0 \in int(\mathcal{B})$; (b) $g(z,\beta)$ is differentiable in a neighborhood $\mathcal{N}$ of β_0 and $E[\sup_{\beta\in\mathcal{N}} \|\partial g(z,\beta)/\partial\beta'\|] < \infty$; (c) $r(\beta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of β_0 and $\sup_{\beta\in\mathcal{N}} \|R(\beta)\| < \infty$; (d) $rank(G) = p$ and $rank(R) = r$.*

4.1 Restricted GEL Estimation

The GEL framework is easily adapted to deal with parametric constraints expressed in contraint equation form. We redefine the GEL criterion function as

$$\tilde{P}(\beta, \lambda, \eta) = \sum_{i=1}^{n} \rho(\lambda' g_i(\beta) + \eta' r(\beta))/n. \tag{4.2}$$

The first order conditions corresponding to η are $\sum_{i=1}^{n} \rho_1(\lambda' g_i(\beta) + \eta' r(\beta)) r(\beta) = 0$ which imply that the constraints $r(\beta) = 0$ of (4.1) are imposed. Therefore, this formulation (4.2) of the optimisation problem is equivalent to that based on the GEL criterion $\hat{P}(\beta, \lambda)$ subject to $r(\beta) = 0$. The corresponding GEL, auxiliary parameter and Lagrange multiplier estimators are denoted by $\tilde{\beta}$, $\tilde{\lambda}$ and $\tilde{\eta}$ respectively.

Let $\mathcal{B}^r = \{\beta : r(\beta) = 0, \beta \in \mathcal{B}\}$. Then, defining the solution $\tilde{\lambda}(\beta) = \arg\max_{\lambda \in \hat{\Lambda}_n(\beta)} \hat{P}(\beta, \lambda)$, $\beta \in \mathcal{B}^r$, we have $\tilde{\lambda}(\beta) = \hat{\lambda}(\beta)$ for $\beta \in \mathcal{B}^r$, where $\hat{\lambda}(\beta)$ is defined below (2.4). Therefore, also let $\tilde{\beta} = \arg\min_{\beta \in \mathcal{B}^r} \hat{P}(\beta, \hat{\lambda}(\beta))$ and $\tilde{\lambda} = \arg\max_{\lambda \in \hat{\Lambda}_n(\tilde{\beta})} \hat{P}(\tilde{\beta}, \lambda)$.[2]

For completeness, we detail the limiting properties of the GEL, auxiliary parameter and Lagrange multiplier estimators in the following result.

Proposition 4.1 *If Assumptions 4.1 and 4.2 are satisfied, then* $\tilde{\beta} \xrightarrow{p} \beta_0$, $\tilde{\lambda} \xrightarrow{p} 0$ *and* $\tilde{\eta} \xrightarrow{p} 0$ *and*

$$n^{1/2}(\tilde{\beta} - \beta_0) \xrightarrow{d} N(0, K),$$

$$n^{1/2} \begin{pmatrix} \tilde{\lambda} \\ \tilde{\eta} \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega^{-1} - \Omega^{-1} G K G' \Omega^{-1} & -\Omega^{-1} G \Sigma R' (R \Sigma R')^{-1} \\ -(R \Sigma R')^{-1} R \Sigma G' \Omega^{-1} & (R \Sigma R')^{-1} - I_r \end{pmatrix} \right),$$

where $K \equiv \Sigma - \Sigma R' (R \Sigma R')^{-1} R \Sigma$. *Moreover, the restricted GEL estimator* $\tilde{\beta}$ *and auxiliary parameter and Lagrange multiplier estimators* $(\tilde{\lambda}, \tilde{\eta})$ *are asymptotically uncorrelated.*

An efficient restricted GMM estimator for β_0 and Lagrange multiplier estimator associated with the constraints $r(\beta_0) = 0$ may also be defined straightforwardly from (2.2) and

[2] Let the Lagrange multiplier estimator $\tilde{\eta} = \tilde{\eta}(\tilde{\beta}, \tilde{\lambda})$. Also let $\pi_i(\beta, \lambda) = \rho_1(\lambda' g_i(\beta)) / \sum_{j=1}^{n} \rho_1(\lambda' g_j(\beta))$ as in (3.4). Then, from the Proof of Proposition 4.1, (A.2), $\tilde{\eta}(\beta, \lambda)$ satisfies $\tilde{\eta}(\beta, \lambda) = -(\sum_{i=1}^{n} \pi_i(\beta, \lambda)(R(\beta) Q R(\beta)')^{-1} R(\beta) Q G_i(\beta)')\lambda$ with probability approaching one where Q is an (arbitrary) nonsingular matrix. Hence, the auxiliary parameter estimator $\tilde{\lambda}(\beta)$ satisfies $(\sum_{i=1}^{n} \pi_i(\beta, \tilde{\lambda}(\beta))[I_p - R(\beta)'(R(\beta) Q R(\beta)')^{-1} R(\beta) Q] G_i(\beta)')\tilde{\lambda}(\beta) = 0$ with probability approaching one.

(4.1). Under Assumptions 2.1, 4.1 and 4.2 they are asymptotically equivalent to the GEL estimators $\tilde{\beta}$ and $\tilde{\eta}$ given above. An auxiliary parameter estimator based on an efficient restricted GMM estimator which is asymptotically equivalent to the GEL estimator $\tilde{\lambda}$ may then be obtained in a similar fashion to $\hat{\lambda}$. We therefore adopt the common notation $\tilde{\beta}$ for both restricted efficient GMM and GEL estimators.

4.2 Implied Probabilities

Let $\tilde{g}_i \equiv g_i(\tilde{\beta})$, $(i = 1, ..., n)$. As the restricted GMM or GEL estimator $\tilde{\beta}$ satisfies the constraints (4.1), we define the constrained implied probabilities as

$$\tilde{\pi}_i = \frac{\rho_1(\tilde{\lambda}'\tilde{g}_i)}{\sum_{j=1}^{n} \rho_1(\tilde{\lambda}'\tilde{g}_j)}, (i = 1, ..., n)\,. \tag{4.3}$$

The efficient estimator of the observation distribution function $\mu(\cdot)$ incorporating both constraint (4.1) and moment restriction (2.1) information is given by

$$\tilde{\mu}_n(z) = \sum_{i=1}^{n} \tilde{\pi}_i 1(z_i \leq z). \tag{4.4}$$

Both the EDF $\mu_n(z)$ and the unconstrained GMM or GEL estimator $\hat{\mu}_n(z)$ remain consistent estimators of the observation distribution $\mu(z)$, whether or not the null hypthesis $H_0 : r(\beta_0) = 0$ is true. Therefore, similar to the previous section alternative statistics appropriate for testing the restrictions (4.1) may be based on contrasts of the restricted and unrestricted implied probabilities $\tilde{\pi}_i$ and $\hat{\pi}_i$, $(i = 1, ..., n)$, (4.3) and (3.4), and the GEL distribution function estimators $\tilde{\mu}_n(\cdot)$ and $\hat{\mu}_n(\cdot)$, (3.5) and (4.4). Let

$$\begin{aligned} n^{1/2}[\tilde{\mu}_n(z) - \mu_n(z)] &\equiv \tilde{\Lambda}_n(z), \\ n^{1/2}[\tilde{\mu}_n(z) - \hat{\mu}_n(z)] &\equiv \Delta_n(z), z \in Z. \end{aligned}$$

Lemma 4.1 *If Assumptions 2.1, 4.1 and 4.2 are satisfied then* $\tilde{\Lambda}_n \Rightarrow \tilde{\Lambda}$ *and* $\Delta_n \Rightarrow \Delta$ *where* $\hat{\Lambda}$ *and* Δ *are Gaussian processes on* Z *both with zero mean and respective covariance functions* $E[\tilde{\Lambda}(z_1)\tilde{\Lambda}(z_2)] = b(z_1)'(\Omega^{-1} - \Omega^{-1}GKG'\Omega^{-1})b(z_2)$ *and* $E[\Delta(z_1)\Delta(z_2)] = b(z_1)'\Omega^{-1}G\Sigma R'(R\Sigma R')^{-1}R\Sigma G\Omega^{-1}b(z_2)$ *where* $b(z) = E[1(z_i \leq z)g_i(\beta_0)]$.

4.3 Pearson-Type Tests

The statistics suggested below for testing the parametric restrictions (4.1) are based on the contrasts $n\tilde{\pi}_i - n\hat{\pi}_i$, $(i = 1, ..., n)$, and adapt the statistics P_n^a (3.7) and P_n^b (3.8) to this context. Therefore, replacing the (implicit) unrestricted EDF divisor unity in P_n^a by $n\hat{\pi}_i$ and the restricted divisor $n\hat{\pi}_i$ in P_n^b by $n\tilde{\pi}_i$,

$$P_n^{a,r} = \sum_{i=1}^{n} \frac{(n\tilde{\pi}_i - n\hat{\pi}_i)^2}{n\hat{\pi}_i} \tag{4.5}$$

and

$$P_n^{b,r} = \sum_{i=1}^{n} \frac{(n\tilde{\pi}_i - n\hat{\pi}_i)^2}{n\tilde{\pi}_i}. \tag{4.6}$$

Of course, the EDF divisor unity can also be employed; *viz.*

$$P_n^{c,r} = \sum_{i=1}^{n} (n\tilde{\pi}_i - n\hat{\pi}_i)^2. \tag{4.7}$$

In the Appendix we show that these three statistics are asymptotically equivalent to the Wald statistic

$$W_n = nr(\hat{\beta})'(\hat{R}\hat{\Sigma}\hat{R}')^{-1}r(\hat{\beta}) \tag{4.8}$$

for testing the parametric restrictions $H_0 : r(\beta_0) = 0$ of (4.1), where $\hat{R} = R(\hat{\beta})$, $\hat{\Sigma} = (\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}$, $\hat{G} = \sum_{i=1}^{n} G_i(\hat{\beta})/n$ or $\sum_{i=1}^{n} \hat{\pi}_i G_i(\hat{\beta})$ and $\hat{\Omega}$ is defined above Theorem 3.2. Therefore:[3]

Theorem 4.1 *If Assumptions 2.1, 4.1 and 4.2 are satisfied, the GEL Pearson-type statistics $P_n^{a,r}$, $P_n^{b,r}$ and $P_n^{c,r}$ are asymptotically equivalent to W_n. Therefore $P_n^{a,r}$, $P_n^{b,r}$, $P_n^{c,r} \xrightarrow{d} \chi_r^2$.*

[3]Lemma 4.1 may be exploited to provide a test of the joint hypothesis given by the contraints (4.1) and moment restrictions (2.1). Pearson-type statistics are defined similarly to P_n^a (3.7) and P_n^b (3.8) as $\sum_{i=1}^{n}(n\tilde{\pi}_i - 1)^2/n\tilde{\pi}_i$ and $\sum_{i=1}^{n}(n\tilde{\pi}_i - 1)^2$. Under Assumptions 2.1, 4.1 and 4.2, these statistics are asymptotically equivalent to the corresponding GMM and GEL statistics and have a limiting chi-square distribution with $m - p + r$ degrees of freedom.

As in section 3.2 consider the partition Z_j, $(j = 1, 2, ...)$, of the sample space Z of z and the (arbitrary) finite collection of subsets Z_j, $(j = 1, ..., s)$, whose union may not equal Z, that is, $\cup_{j=1}^{s} Z_j \subset Z$. We impose the order condition $s \geq m$ and require $\mu(Z_j) > 0$, $(j = 1, ..., s)$. Define the distribution function estimator

$$\tilde{\mu}_n(Z_j) = \sum_{i=1}^{n} \tilde{\pi}_i 1\left(z_i \in Z_j\right), j = 1, ..., s. \tag{4.9}$$

Let $\tilde{\mu}_n^s = (\tilde{\mu}_n(Z_1), ..., \tilde{\mu}_n(Z_s))'$. Also let $B_s = (b(Z_1), ..., b(Z_s))$ where $b(Z_j) = E[1(z \in Z_j)g(z, \beta_0)]$, $(j = 1, ..., s)$. The test statistics defined below are based on the normalised contrast $\tilde{\mu}_n^s - \hat{\mu}_n^s$ from (4.9) and (3.9). It follows immediately from Lemma 4.1 that $n^{1/2}(\tilde{\mu}_n^s - \hat{\mu}_n^s) \xrightarrow{d} N(0, B_s'\Omega^{-1}G\Sigma R'(R\Sigma R')^{-1}R\Sigma G'\Omega^{-1}B_s)$. Now if B_s is full row rank m then $B_s'(B_sB_s')^{-1}G\Sigma G'(B_sB_s')^{-1}B_s$ is a g-inverse for $B_s'\Omega^{-1}G\Sigma R'(R\Sigma R')^{-1}R\Sigma G'\Omega^{-1}B_s$. A test for the restrictions (4.1) may be based on the alternative statistic

$$P_n^{a,alt,r} = n(\tilde{\mu}_n^s - \hat{\mu}_n^s)'\hat{B}_s'(\hat{B}_s\hat{B}_s')^{-1}\hat{G}\hat{\Sigma}\hat{G}'(\hat{B}_s\hat{B}_s')^{-1}\hat{B}_s(\tilde{\mu}_n^s - \hat{\mu}_n^s), \tag{4.10}$$

where $\hat{B}_s$, $\hat{G}$ and $\hat{\Sigma}$ are defined above Theorems 3.2 and 4.1.[4,5] The statistic $P_n^{a,alt,r}$ of (4.10) may be further simplified using Lemma 4.1 by noting that $G'(\Omega^{-1} - \Omega^{-1}GKG'\Omega^{-1}) = R'(R\Sigma R')^{-1}R\Sigma G'\Omega^{-1}$ yielding the statistic

$$P_n^{b,alt,r} = n(\tilde{\mu}_n^s - \mu_n^s)'\hat{B}_s'(\hat{B}_s\hat{B}_s')^{-1}\hat{G}\hat{\Sigma}\hat{G}'(\hat{B}_s\hat{B}_s')^{-1}\hat{B}_s(\tilde{\mu}_n^s - \mu_n^s). \tag{4.11}$$

Theorem 4.2 *If Assumptions 2.1, 4.1 and 4.2 are satisfied and $rk(B_s) = m$ then the GEL Pearson-type test statistics $P_n^{alt,r}$ and $P_n^{b,alt,r}$ are asymptotically equivalent to $P_n^{a,r}$, $P_n^{b,r}$ and $P_n^{c,r}$. Therefore, $P_n^{alt,r}$, $P_n^{b,alt,r} \xrightarrow{d} \chi_r^2$.*

[4]More generally the limiting distribution of the statistic $n(\tilde{\mu}_n^s - \hat{\mu}_n^s)'\hat{\Xi}^-(\tilde{\mu}_n^s - \hat{\mu}_n^s)$, where $\hat{\Xi}^-$ denotes a consistent estimator for a g-inverse of $B_s'\Omega^{-1}G\Sigma R'(R\Sigma R')^{-1}R\Sigma G'\Omega^{-1}B_s$, is that of a chi-square random variable with $rk(B_s'\Omega^{-1}G\Sigma R'(R\Sigma R')^{-1}R\Sigma G'\Omega^{-1}B_s)$ degrees of freedom.

[5]By a proof similar to those of Lemmas 3.1 and 4.1 $n^{1/2}(\tilde{\mu}_n^s - \mu_n^s) \xrightarrow{d} N(0, B_s'(\Omega^{-1} - \Omega^{-1}GKG'\Omega^{-1})B_s)$. If B_s is full row rank then $B_s'(B_sB_s')^{-1}\Omega(B_sB_s')^{-1}B_s$ is a g-inverse for $B_s'(\Omega^{-1} - \Omega^{-1}GKG'\Omega^{-1})B_s$. Therefore a test for the joint hypothesis given by the contraints (4.1) and moment restrictions (2.1) is given by a Pearson-type statistic defined similarly to P_n^{alt} (3.11), that is, $n(\tilde{\mu}_n^s - \mu_n^s)'\hat{B}_s'(\hat{B}_s\hat{B}_s')^{-1}\hat{\Omega}(\hat{B}_s\hat{B}_s')^{-1}\hat{B}_s(\tilde{\mu}_n^s - \mu_n^s)$. Under Assumptions 2.1, 4.1 and 4.2, this statistic is asymptotically equivalent to the corresponding GMM and GEL statistics and Pearson-type statistics defined in fn. 3 and has a limiting chi-square distribution with $m - p + r$ degrees of freedom.

If in addition $s = m$ then B_s is nonsingular so that $B_s^{-1} G \Sigma G' B_s'^{-1}$ is a g-inverse for $B_s' \Omega^{-1} G \Sigma R' (R \Sigma R')^{-1} R \Sigma G' \Omega^{-1} B_s$.

Corollary 4.1 *If Assumptions 2.1, 4.1 and 4.2 are satisfied, $rk(B_s) = m$ and $s = m$ then the statistics $P_n^{a,alt,r} = n(\tilde{\mu}_n^s - \hat{\mu}_n^s)' \hat{B}_s^{-1} \hat{G} \hat{\Sigma} \hat{G}' \hat{B}_s'^{-1} (\tilde{\mu}_n^s - \hat{\mu}_n^s)$ and $P_n^{b,alt,r} = n(\tilde{\mu}_n^s - \mu_n^s)' \hat{B}_s^{-1} \hat{G} \hat{\Sigma} \hat{G}' \hat{B}_s'^{-1} (\tilde{\mu}_n^s - \mu_n^s)$ are asymptotically equivalent to $P_n^{a,r}$, $P_n^{b,r}$ and $P_n^{c,r}$. Therefore $P_n^{alt,r}$, $P_n^{b,alt,r} \xrightarrow{d} \chi_r^2$.*

Consider the local alternatives to the constraints (4.1) $H_n : r(\beta_0) = n^{-1/2}\xi + o(n^{-1/2})$, $(i = 1, ..., n)$, $n = 1, 2, ...$ As above, $n^{1/2} \hat{g}(\beta_0) \xrightarrow{d} N(0, \Omega)$ remains valid under H_n. Consistency of the restricted GEL and auxiliary parameter estimators $\tilde{\beta}$, $\tilde{\lambda}$ and Lagrange multiplier estimator $\tilde{\eta}$ for β_0, 0 and 0 still obtains. The expansions $n^{1/2}(\tilde{\beta} - \beta_0) = -\Sigma R'(R\Sigma R')^{-1}\xi - KG'\Omega^{-1} n^{1/2} \hat{g}(\beta_0) + o_p(1)$ and $n^{1/2}\tilde{\lambda} = -\Omega^{-1} G \Sigma R'(R\Sigma R')^{-1}\xi - (\Omega^{-1} - \Omega^{-1} G K G' \Omega^{-1}) n^{1/2} \hat{g}(\beta_0) + o_p(1)$ and $n^{1/2}\tilde{\eta} = (R\Sigma R')^{-1}\xi + (R\Sigma R')^{-1} R \Sigma G' \Omega^{-1} n^{1/2} \hat{g}(\beta_0) + o_p(1)$ become appropriate under H_n. Hence, the statistics $P_n^{a,r}$, $P_n^{b,r}$, $P_n^{c,r}$ and $P_n^{a,alt,r}$, $P_n^{b,alt,r}$ remain asymptotically equivalent to W_n and other GMM or GEL statistics for testing the constraints $r(\beta_0) = 0$ (4.1). Therefore, $P_n^{a,r}$, $P_n^{b,r}$, $P_n^{c,r}$ and $P_n^{a,alt,r}$, $P_n^{b,alt,r}$ converge in distribution to a non-central chi-square random variable with r degrees of freedom and non-centrality parameter $\xi'(R\Sigma R')^{-1}\xi$ under H_n.

When considering the consistency of the tests using the statistics $P_n^{a,r}$, $P_n^{b,r}$, $P_n^{c,r}$ and $P_n^{alt,r}$, $P_n^{b,alt,r}$, we firstly need to examine the limiting behaviour of the restricted GMM or GEL estimator $\tilde{\beta}$ and associated auxiliary parameter and Lagrange multiplier estimators $\tilde{\lambda}$ and $\tilde{\eta}$ when $r(\beta_0) \neq 0$. Because the hypothesis $r(\beta) = 0$ is imposed, $\tilde{P}(\beta, \lambda, \eta) = \hat{P}(\beta, \lambda)$, $\beta \in \mathcal{B}^r$. Therefore, $\tilde{P}(\beta, \lambda, \eta) \xrightarrow{p} \rho(\beta, \lambda) = E[\rho(\lambda' g(z, \beta)) | \lambda' g(z, \beta) \in \mathcal{V}]$ uniformly $\beta \in \mathcal{B}^r$ and λ with $\rho(\beta, \lambda)$ continuous in β and λ. As in section 3 let $\lambda(\beta) = \arg\max_\lambda \rho(\beta, \lambda)$. For GMM, as $\tilde{\beta} \xrightarrow{p} \beta_0$, $\hat{g}(\beta)' \hat{\Omega}(\tilde{\beta})^{-1} \hat{g}(\beta) \xrightarrow{p} g(\beta)' \Omega(\beta_0)^{-1} g(\beta)$ uniformly $\beta \in \mathcal{B}^r$.

We modify Assumption 3.1 appropriately.

Assumption 4.3 *(a) $r(\beta_0) \neq 0$; (b) $r(\beta)$ is continuous at each $\beta \in \mathcal{B}^r$; (c) $\lambda(\beta)$ is the* *unique maximiser of $\rho(\beta,\lambda)$ and is continuous in $\beta \in \mathcal{B}^r$; (d) β_* is the unique minimiser in $\mathcal{B}^r$ of $\rho(\beta,\lambda(\beta))$ or $g(\beta)'\Omega(\beta_0)^{-1}g(\beta)$.*

The consistency of tests based on the statistics $P_n^{a,r}$, $P_n^{b,r}$ and $P_n^{c,r}$ now follows.

Theorem 4.3 *If Assumptions 2.1-2.3 and 4.3 are satisfied then $P_n^{a,r}$, $P_n^{b,r}$, $P_n^{c,r} \xrightarrow{p} \infty$.*

Under Assumptions 2.1, 2.2 and 2.3, $\hat{G} \xrightarrow{p} G$, $\hat{\Omega} \xrightarrow{p} \Omega$ and $\hat{B}_s \xrightarrow{p} B_s$. Let $\lambda_* \equiv \lambda(\beta_*)$. Recall that $\delta_{*j} = E[(\rho_1(\lambda_*'g(z,\beta_*)) - \rho_1^*)1(z \in Z_j)|\lambda_*'g(z,\beta_*) \in \mathcal{V}]/\rho_1^*$, $(j = 1, ..., s)$, where $\rho_1^* = E[\rho_1(\lambda_*'g(z,\beta_*))|\lambda_*'g(z,\beta_*) \in \mathcal{V}]$, and $\delta_* = (\delta_{*1}, ..., \delta_{*s})'$.

Theorem 4.4 *If Assumptions 2.1-2.3 and 4.3 are satisfied, $rk(B_s) = m$ and $G'(B_sB_s')^{-1}B_s\delta_* \neq 0$, then $P_n^{a,alt,r}$, $P_n^{b,alt,r} \xrightarrow{p} \infty$.*

If $s = m$ as in Corollary 4.1 and thus $\hat{B}_s^{-1}\hat{G}\hat{\Sigma}\hat{G}'\hat{B}_s'^{-1}$ replaces $\hat{B}_s'(\hat{B}_s\hat{B}_s')^{-1}\hat{G}\hat{\Sigma}\hat{G}'(\hat{B}_s\hat{B}_s')^{-1}$ in the definition of $P_n^{a,alt,r}$ (4.10) and $P_n^{b,alt,r}$ (4.11), the consistency condition of Theorem 4.4 becomes $G'B_s'^{-1}\delta_* \neq 0$.

Alternatively, B_s, G and Σ may be estimated consistently under $H_0 : r(\beta_0) = 0$ (4.1) using the restricted estimator $\tilde{\beta}$ and implied probabilities $\tilde{\pi}_i$, $(i = 1, ..., n)$, that is, by $\tilde{B}_s = (\tilde{b}(Z_1), ..., \tilde{b}(Z_s))$, $\tilde{b}(Z_j) = \sum_{i=1}^n 1(z \in Z_j)\tilde{g}_i/n$ or $\sum_{i=1}^n \tilde{\pi}_i 1(z \in Z_j)\tilde{g}_i$, $(j = 1, ..., s)$, $\tilde{\Sigma} = (\tilde{G}'\tilde{\Omega}^{-1}\tilde{G})^{-1}$, $\tilde{G} = \sum_{i=1}^n G_i(\tilde{\beta})/n$ or $\sum_{i=1}^n \tilde{\pi}_i G_i(\tilde{\beta})$, $\tilde{\Omega} = \sum_{i=1}^n \tilde{g}_i\tilde{g}_i'/n$, $\sum_{i=1}^n [\tilde{g}_i - \tilde{g}][\tilde{g}_i - \tilde{g}]'/n$, $\tilde{g} = \hat{g}(\tilde{\beta})$, or $\sum_{i=1}^n \tilde{\pi}_i\tilde{g}_i\tilde{g}_i'$. No alteration is necessary to either the conclusions stated in Theorem 4.2 and Corollary 4.1 or the following discussion regarding the limiting behaviour of the Pearson-type statistics $P_n^{a,alt,r}$ and $P_n^{b,alt,r}$ under local alternatives. Some modification, however, is required for test consistency. Let $b_*(Z_j) = E[1(z \in Z_j)g(z,\beta_*)|\lambda_*'g(z,\beta_*) \in \mathcal{V}]$ or $E[\rho_1(\lambda_*'g(z,\beta_*))1(z \in Z_j)g(z,\beta_*)|\lambda_*'g(z,\beta_*) \in \mathcal{V}]/\rho_1^*$, $(j = 1, ..., s)$, $B_{s*} = (b_*(Z_1), ..., b_*(Z_s))$, $G_* = E[\partial g(z,\beta_*)/\partial\beta'|\lambda_*'g(z,\beta_*) \in \mathcal{V}]$ or $E[\rho_1(\lambda_*'g(z,\beta_*))\partial g(z,\beta_*)/\partial\beta'|\lambda_*'g(z,\beta_*) \in \mathcal{V}]/\rho_1^*$ and $\Omega_* = E[g(z,\beta_*)g(z,\beta_*)'|\lambda_*'g(z,\beta_*) \in \mathcal{V}]$, $E[(g(z,\beta_*) - g_*)(g(z,\beta_*) - g_*)'|\lambda_*'g(z,\beta_*) \in \mathcal{V}]$, $g_* = E[g(z,\beta_*)|\lambda_*'g(z,\beta_*) \in \mathcal{V}]$, or $E[\rho_1(\lambda_*'g(z,\beta_*))g(z,\beta_*)g(z,\beta_*)'|\lambda_*'g(z,\beta_*) \in \mathcal{V}]/\rho_1^*$. Then $\tilde{G} \xrightarrow{p} G_*$, $\tilde{\Omega} \xrightarrow{p} \Omega_*$ and $\tilde{B}_s \xrightarrow{p} B_{s*}$. The hypotheses of Theorem 3.4 require the additional conditions $rk(\Omega_*) = m$,

$rk(G_*) = p$, $rk(B_{s*}) = m$ and $G'_*(B_{s*}B'_{s*})^{-1}B_{s*}\delta_* \neq 0$. Hence, $P_n^{a,alt,r}$, $P_n^{b,alt,r} \xrightarrow{p} \infty$. If $s = m$ as in Corollary 4.1 and $\tilde{B}_s^{-1}\tilde{G}\tilde{\Sigma}\tilde{G}'\tilde{B}_s'^{-1}$ replaces $\tilde{B}'_s(\tilde{B}_s\tilde{B}'_s)^{-1}\tilde{G}\tilde{\Sigma}\tilde{G}'(\tilde{B}_s\tilde{B}'_s)^{-1}$ then the test consistency condition is $G'_*B_{s*}'^{-1}\delta_* \neq 0$ substituting for $G'_*(B_{s*}B'_{s*})^{-1}B_{s*}\delta_* \neq 0$.

5 Simulation Evidence: Finite Sample Properties of Tests of Over-Identifying Moment Conditions

This section investigates the finite sample properties of some of the Pearson-type tests proposed in previous sections. In particular, we examine the size properties of the P_n^a (3.7), P_n^b (3.8) and P_n^{alt} (3.11) test statistics for overidentifying moment restrictions. We assess their performance in comparison with tests based on the GEL criterion function: $GELR_n$ (3.1), Lagrange multiplier LM_n (3.2) and score S_n (3.3) statistics.

5.1 Experimental Designs

The simulation study in Imbens, Spady and Johnson (1998) forms the basis for our comparison of the finite sample properties of the aforementioned tests. In particular, we use their first two experimental designs for our investigation. The first design is a simplified version of an asset-pricing model, characterized by the moment indicators

$$g(z,\beta) = \begin{pmatrix} \exp\left[-0.72 - \beta(z_1 + z_2) + 3z_2\right] - 1 \\ z_2(\exp[-0.72 - \beta(z_1 + z_2) + 3z_2] - 1) \end{pmatrix}, \tag{5.1}$$

after partitioning $z = (z_1, z_2)'$, where z_1 and z_2 are generated independently from a $N(0, 0.16)$ distribution and the true value $\beta_0 = 3$. The second experiment is based on the moment indicator

$$g(z,\beta) = \begin{pmatrix} z - \beta \\ z^2 - \beta^2 - 2\beta \end{pmatrix}, \tag{5.2}$$

where z has a chi-square distribution with one degree of freedom and $\beta_0 = 1$. We considered samples of size $n = 100$, 200, 500 and 1000 observations, each experiment being replicated 10000 times.

Tests evaluated at GEL estimators ($GELR_n$, LM_n, P_n^a, P_n^b and P_n^{alt}) use either ET or EL estimation. Consistent estimators for the matrices G and Ω required in the computation of the LM_n and P_n^{alt} statistics were obtained in three different ways:

- $gel(n)$: sample means, for example:

$$\hat{\Omega} = \sum_{i=1}^{n} g_i(\hat{\beta})g_i(\hat{\beta})'/n; \tag{5.3}$$

- $gel(s)$: GEL implied probabilities $\hat{\pi}_i$, $(i = 1, ..., n)$, for example:

$$\hat{\Omega} = \sum_{i=1}^{n} \hat{\pi}_i g_i(\hat{\beta})g_i(\hat{\beta})'; \tag{5.4}$$

- $gel(r)$: G as in $gel\,(s)$ with Ω estimated robustly by:

$$\hat{\Omega} = \sum_{i=1}^{n} \hat{\pi}_i g_i(\hat{\beta})g_i(\hat{\beta})' \left(n \sum_{i=1}^{n} \hat{\pi}_i^2 g_i(\hat{\beta})g_i(\hat{\beta})' \right)^{-1} \sum_{i=1}^{n} \hat{\pi}_i g_i(\hat{\beta})g_i(\hat{\beta})'. \tag{5.5}$$

These estimators for the variance matrix Ω were also used in the computation of the GEL score statistic S_n. Additionally, S_n was also evaluated at two-step (S_n^{2s}), iterated (S_n^i) and continuous updating (S_n^{cue}) GMM estimators. In these cases, however, only the consistent estimator for Ω based on sample means was used; see Hansen, Heaton and Yaron (1996).

In their Monte Carlo simulation study, Imbens, Spady and Johnson (1998) analyzed the finite sample behaviour of a test based on the following statistics: S_n^{2s}, S_n^i, S_n^{cue}, $S_n^{et(s)}$, $LM_n^{et(s)}$, $LM_n^{et(r)}$, $GELR_n^{et}$ and $GELR_n^{el}$. We replicate their results for the two experimental designs described above and examine whether their conclusions remain valid when other estimators are employed to evaluate the LM_n and S_n statistics. In particular, we study the effects of using EL instead of ET estimation [$S_n^{el(s)}$, $LM_n^{el(s)}$ and $LM_n^{el(r)}$]. We confirm their conjecture that robust estimation of Ω results in a deterioration in the performance of the score statistic S_n [$S_n^{et(r)}$ and $S_n^{el(r)}$] for reasons explained below. We also investigate the consequences of using the sample mean estimator for Ω when GEL estimation is utilized [$S_n^{et(n)}$, $S_n^{el(n)}$, $LM_n^{et(n)}$ and $LM_n^{el(n)}$].

The implementation of P_n^{alt} examined here used the complete partition of the sample space Z, that is, the partition of Z consists of s subsets. To examine the sensitivity of P_n^{alt} to s, we considered two values for s, $s = 8$ and 16. The definition of each subset constituting the partition of Z was such that in each Monte Carlo sample each subset contained approximately $(100/s)\,\%$ of the observations.

5.2 Results

Tables 1 and 2, for the asset-pricing model, and 3 and 4, for the chi-square moments case, report the estimated size of each test at seven different levels of significance 0.200, 0.100, 0.050, 0.025, 0.010, 0.005 and 0.001. For each significance level, sample size and model considered, the actual size closest to the nominal size is underlined.

Tables 1, 2, 3 and 4 about here

The results displayed in Tables 1 and 3 conform with those presented by Imbens, Spady and Johnson (1998) for the tests analyzed in their paper.[6] They show that all these tests are significantly oversized in almost all cases, even when $n = 1000$, particularly for the chi-square moments model. The statistic $LM_n^{et(r)}$ registers the best behaviour in most experiments, the only exceptions being for the largest nominal sizes, where S_n^{cue}, in the first model, and $LM_n^{el(r)}$, in both models, achieve superior performances. The size behaviour of the S_n statistic evaluated at the two-step GMM estimator, which is most commonly used to assess overidentifying moment condition models, is generally disastrous in these experiments. In particular, it is the worst of all versions [S_n^{2s}, S_n^i, S_n^{cue}, $S_n^{et(n)}$ and $S_n^{el(n)}$] using the sample mean estimator for Ω in the asset-pricing model. The $GELR_n$ tests also produced very modest results, with the EL version performing substantially better than that using ET, particularly for the chi-square moments model and for the smallest nominal sizes.

As noted by Imbens, Spady and Johnson (1998), estimation of the variance matrix Ω exerts a decisive influence on the performance of the tests. However, the extraordinary benefits from the use of robust estimation reported there for the Lagrange multiplier statistic $LM_n^{et(r)}$ do not extend to all tests, not even to $LM_n^{el(r)}$ for the smallest nominal sizes considered. The size behaviour of the score statistic S_n also deteriorates considerably. Although a theoretical analysis of the effects of using robust estimation is beyond

[6]The following correspondence holds between the notation used here and that utilized by Imbens, Spady and Johnson (1998): $S_n^{2s} = T_{g1}^{AM}$, $S_n^i = T_{g2}^{AM}$, $S_n^{cue} = T_{g3}^{AM}$, $S_n^{et(s)} = T_{et}^{AM}$, $LM_n^{et(s)} = T_{et(s)}^{LM}$, $LM_n^{et(r)} = T_{et(r)}^{LM}$, $GELR_n^{et} = T_{klic(et)}^{CF}$ and $GELR_n^{el} = T_{lr(el)}^{CF}$.

the scope of this paper, it is clear that LM_n and S_n are affected in an opposite manner because an estimator for Ω appears as an inverse in the latter statistic.

Estimated sizes for the Pearson-type statistics are reported in Tables 2 and 4. The P_n^a and P_n^b statistics perform very modestly, being substantially oversized in all cases. Their size behaviour does not differ much from that described above for the other tests.[7] In contradistinction, however, P_n^{alt} is more promising. Whichever number of classes s is chosen, the general effects of evaluation at different estimators are similar in all cases. Analogously to LM_n, the least number of rejections of the null hypothesis occurs when robust estimation of Ω is employed. This is unsurprising since Ω appears in the expressions for both tests in a similar manner. Overall, robust $et(r)$ and $el(r)$ versions of P_n^{alt} record most of the best size properties.

Figure 1 about here

Figure 1 displays QQ-plots comparing the six versions of P_n^{alt} for $s = 8$. Vertical coordinates are Monte Carlo estimates of quantiles of the finite sample distribution of those statistics and horizontal coordinates are quantiles of a chi-square variable with one degree of freedom. The vertical solid line marks the asymptotic critical value for a nominal size of 0.05. Clearly, the best performances are obtained by $P_n^{alt,et(r)}$ and $P_n^{alt,el(r)}$. Note that for $n \geq 500$ (first model) or $n = 1000$ (second model) the estimated and asymptotic quantiles of these statistics are very close while other versions of P_n^{alt} are still significantly oversized. It is also worthy of notice how, for small sample sizes, all three EL versions of P_n^{alt} tend to reject significantly less than the corresponding ET variants.

Figure 2 about here

The size performance of P_n^{alt} did not appear to be affected significantly by s for different sample sizes. This was particularly evident for the asset-pricing model case.

[7] The estimated sizes for the EL version of P_n^b test are numerically equal to those calculated for $S_n^{el(s)}$ and $LM_n^{el(s)}$. This is due to the particular form assumed by the EL implied probabilities (3.4): $\hat{\pi}_i = n^{-1}(1 + \hat{\lambda}' g_i(\hat{\beta}))^{-1}$, $(i = 1, ..., n)$. For example, as $\hat{\lambda}' g_i(\hat{\beta}) = n\hat{\pi}_i - 1$ and $\hat{\Omega} = \sum_{i=1}^n \hat{\pi}_i g_i(\hat{\beta}) g_i(\hat{\beta})'$, $LM_n^{el(s)} = n\hat{\lambda}'\hat{\Omega}\hat{\lambda} = \sum_{i=1}^n (\hat{\lambda}' g_i(\hat{\beta}))^2/(1 + \hat{\lambda}' g_i(\hat{\beta})) = P_n^b$.

For the chi-squared moment model the differences between $s = 8$ and $s = 16$ cases were more important but were attenuated by increasing sample size. Figure 2 illustrates this situation for $P_n^{alt,et(r)}$ displaying QQ-plots for both values of s.

Figure 3 about here

Figure 3 compares the robust forms of LM_n and P_n^{alt} for $s = 8$ evaluated at ET and EL estimators. Of the statistics considered by Imbens, Spady and Johnson (1998) and here $LM_n^{et(r)}$ registered the best behaviour. The statistic P_n^{alt} clearly performs better for both models with estimated and asymptotic quantiles being closer in most cases. Furthermore, while P_n^{alt} is relatively indifferent to the use of ET or EL estimation, at least for the larger sample sizes, EL estimation does not work well for LM_n, even for $n = 1000$.

6 Conclusions

This paper develops new Pearson-type statistics appropriate for testing over-identifying moment conditions and parametric restrictions. The Pearson-type statistic contructed using a partition of the sample space performed very well in Monte Carlo simulation experiments comparing tests for over-identifying moment conditions. The size behaviour for this statistic based on robust estimation of the moment indicator variance matrix appears to be superior to that of alternative competitor tests. Moreover, this statistic seems to be insensitive to the number of classes comprising the partition of the sample space.

Appendix: Proofs

Throughout the Appendix, with probability approaching one will be abbreviated as w.p.a.1, UWL will denote a uniform weak law of large numbers such as Lemma 2.4 of Newey and McFadden (1994), CS Cauchy-Schwartz and CLT will refer to the Lindeberg-Lévy central limit theorem.

Lemma A.1 *If Assumptions 2.1, 2.2 and 2.3 are satisfied, then* $n\hat{\pi}_i = 1 + o_p(1)$ *and*

$$n^{1/2}\left(\hat{\pi}_i - \frac{1}{n}\right) = \frac{1}{n}\hat{g}_i' n^{1/2}\hat{\lambda}(1 + o_p(1)) + O_p(n^{-3/2}),$$

uniformly $(i = 1, ..., n)$.

Proof: Let $b_i = \sup_{\beta\in\mathcal{B}} \|g_i(\beta)\|$. From the Proof of Lemma A1 and Theorem 3.1 in NS, as $\max_{1\leq i\leq n} b_i = O_p(n^{\frac{1}{\alpha}})$ and $\hat{\lambda} = O_p(n^{-1/2})$, $\sup_{\beta\in\mathcal{B},1\leq i\leq n}\left|\hat{\lambda}' g_i(\beta)\right| = O_p(n^{-(\frac{1}{2}-\frac{1}{\alpha})})$. A first order order Taylor expansion for $\rho_1(\hat{\lambda}'\hat{g}_i)$ yields

$$\rho_1(\hat{\lambda}'\hat{g}_i) = -1 + \rho_2(\dot{\lambda}'\hat{g}_i)\hat{\lambda}'\hat{g}_i,$$

where $\dot{\lambda}$ is on the line joining $\hat{\lambda}$ and 0. Now, $\max_{1\leq i\leq n}\left|\rho_2(\dot{\lambda}'\hat{g}_i) + 1\right| \xrightarrow{p} 0$ as $\sup_{\beta\in\mathcal{B},1\leq i\leq n}\left|\dot{\lambda}' g_i(\beta)\right| \xrightarrow{p} 0$ and so $\rho_2(\dot{\lambda}'\hat{g}_i)\hat{\lambda}'\hat{g}_i = -\hat{\lambda}'\hat{g}_i(1 + o_p(1))$ uniformly $(i = 1, ..., n)$. Therefore,

$$\rho_1(\hat{\lambda}'\hat{g}_i) = -1 - \hat{\lambda}'\hat{g}_i(1 + o_p(1)), \tag{A.1}$$

uniformly $(i = 1, ..., n)$. Similarly,

$$\begin{aligned}\frac{1}{\sum_{j=1}^{n}\rho_1(\hat{\lambda}'\hat{g}_j)} &= -\frac{1}{n} - \frac{1}{n}\left(\sum_{j=1}^{n}\rho_2(\dot{\lambda}'\hat{g}_j)\hat{g}_j'/n\right)\hat{\lambda} \\ &= -\frac{1}{n}(1 + O_p(n^{-1})),\end{aligned}$$

as $\sum_{j=1}^{n}\hat{g}_j/n = O_p(n^{-1/2})$ by Theorem 3.1 of NS. Combining eqs. (A.1) and (A.2)

$$\hat{\pi}_i = \frac{1}{n}(1 + \hat{\lambda}'\hat{g}_i(1 + o_p(1)))(1 + O_p(n^{-1}))$$

and, therefore, from Lemma A1 of NS,

$$\begin{aligned} n\hat{\pi}_i - 1 &= \hat{\lambda}'\hat{g}_i(1 + o_p(1)) + O_p(n^{-1}) \\ &= o_p(1)\end{aligned}$$

uniformly $(i = 1, ..., n)$. Similarly

$$n^{1/2}\left(\hat{\pi}_i - \frac{1}{n}\right) = \frac{1}{n}\hat{g}_i' n^{1/2}\hat{\lambda}(1 + o_p(1)) + O_p(n^{-3/2}),$$

uniformly $(i=1,...,n)$. ■

Proof of Lemma 3.1: By Lemma A.1 and noting from Theorem 3.2 of NS that $n^{1/2}\hat{\lambda} = -Pn^{1/2}\hat{g}(\beta_0) + o_p(1)$,

$$\begin{aligned}
n^{1/2}[\hat{\mu}_n(z) - \mu_n(z)] &= n^{1/2}\sum_{i=1}^{n}\left(\hat{\pi}_i - \frac{1}{n}\right)1(z_i \leq z) \\
&= \sum_{i=1}^{n}[n^{-1}\hat{g}_i' n^{1/2}\hat{\lambda}(1+o_p(1)) + O_p(n^{-3/2})]1(z_i \leq z) \\
&= \left(\sum_{i=1}^{n}1(z_i \leq z)\hat{g}_i'/n\right)n^{1/2}\hat{\lambda}(1+o_p(1)) + O_p(n^{-1/2}) \\
&= [b(z) + O_p(n^{-1/2})]' n^{1/2}\hat{\lambda} + o_p(1) \\
&\Rightarrow \hat{\Lambda}(z)
\end{aligned}$$

where $\hat{\Lambda}$ is Gaussian stochastic process on $\mathcal{R}^k$ with mean zero and covariance function $E[\hat{\Lambda}(z_1)\hat{\Lambda}(z_2)] = b(z_1)'Pb(z_2)$. ■

Proof of Theorem 3.1: Our method of proof is to demonstrate that the statistics P_n^a (3.7) and P_n^b (3.8) are asymptotically equivalent to the Lagrange multiplier test LM_n (3.2) for the over-identifying moment conditions (2.1). Using Lemma A.1

$$(n\hat{\pi}_i - 1)^2 = (\hat{\lambda}'\hat{g}_i(1+o_p(1)) + O_p(n^{-1}))^2,$$

uniformly $(i=1,...,n)$. Summing over $i=1,...,n$,

$$\begin{aligned}
\sum_{i=1}^{n}(n\hat{\pi}_i - 1)^2 &= n\hat{\lambda}'(\sum_{i=1}^{n}\hat{g}_i\hat{g}_i'/n)\hat{\lambda}(1+o_p(1)) + n^{1/2}\hat{\lambda}'(\sum_{i=1}^{n}\hat{g}_i/n^{1/2})(1+o_p(1))O_p(n^{-1}) \\
&\quad + O_p(n^{-1}) \\
&= n\hat{\lambda}'(\sum_{i=1}^{n}\hat{g}_i\hat{g}_i'/n)\hat{\lambda} + o_p(1) \\
&= LM_n + o_p(1).
\end{aligned}$$

From Lemma A.1,

$$\sum_{i=1}^{n}(n\hat{\pi}_i - 1)^2 = \sum_{i=1}^{n}\frac{(n\hat{\pi}_i - 1)^2}{n\hat{\pi}_i} + o_p(1).$$

■

Proof of Theorem 3.2: From a UWL, the matrix estimators $\hat{B}_s$, $\hat{G}$ and $\hat{\Omega}$ are consistent estimators for their population counterparts B_s, G and Ω. From the Proof of Lemma 3.1, $n^{1/2}(\hat{\mu}_n^s - \mu_n^s) = B_s' n^{1/2}\hat{\lambda} + o_p(1) = -B_s' P n^{1/2}\hat{g}(\beta_0) + o_p(1)$ and thus

$$n^{1/2}(\hat{\mu}_n^s - \mu_n^s) \xrightarrow{d} N(0, B_s' P B_s).$$

If $rk(B_s) = m$ then $B_s'(B_s B_s')^{-1}\Omega(B_s B_s')^{-1}B_s$ is a g-inverse for $B_s' P B_s$ as $P\Omega P = P$. Therefore,

$$\begin{aligned} P_n^{alt} &= n(\hat{\mu}_n^s - \mu_n^s)' B_s'(B_s B_s')^{-1}\Omega(B_s B_s')^{-1}B_s(\hat{\mu}_n^s - \mu_n^s) + o_p(1) \\ &= n\hat{g}(\beta_0)' P\Omega P \hat{g}(\beta_0) + o_p(1) \\ &= LM_n + o_p(1), \end{aligned}$$

as $P\Omega P = P$. ■

Proof of Theorem 3.3: From Assumption 3.1, it follows by standard consistency results for concave objective functions (e.g. Newey and McFadden, 1994, Theorem 2.7) that $\hat{\lambda}(\beta) = \arg\max_{\lambda \in \hat{\Lambda}_n(\beta)} \hat{P}(\beta, \lambda)$ exists w.p.a.1 and $\hat{\lambda}(\beta) \xrightarrow{p} \lambda(\beta)$. By a UWL $\sup_{\beta \in \mathcal{B}} \left\| \hat{P}(\beta, \hat{\lambda}(\beta)) - \rho(\beta, \lambda(\beta)) \right\| \xrightarrow{p} 0$. Therefore, the GEL estimator $\hat{\beta} \xrightarrow{p} \beta_*$ using e.g. Theorem 2.1 of Newey and McFadden (1994). As $\mathcal{V}$ is bounded, $\sum_{i=1}^n \rho_1(\lambda' g_i(\beta))/n \xrightarrow{p} E[\rho_1(\lambda' g(z, \beta))|\lambda' g(z, \beta) \in \mathcal{V}]$ and $\sum_{i=1}^n \rho_1(\lambda' g_i(\beta))^2/n \xrightarrow{p} E[\rho_1(\lambda' g(z, \beta))^2|\lambda' g(z, \beta) \in \mathcal{V}]$ uniformly β and λ. Therefore, by a UWL, $\sum_{i=1}^n \rho(\hat{\lambda}'\hat{g}_i)/n \xrightarrow{p} E[\rho_1(\lambda_*' g(z, \beta_*))|\lambda_*' g(z, \beta_*) \in \mathcal{V}]$ and $\sum_{i=1}^n \rho(\hat{\lambda}'\hat{g}_i)^2/n \xrightarrow{p} E[\rho_1(\lambda_*' g(z, \beta_*))^2|\lambda_*' g(z, \beta_*) \in \mathcal{V}]$. Consider the statistic P_n^a.

$$\begin{aligned} n^{-1}P_n^a &= \sum_{i=1}^n (n\hat{\pi}_i - 1)^2/n \\ &= \frac{\sum_{i=1}^n \rho_1(\hat{\lambda}'\hat{g}_i)^2/n}{(\sum_{j=1}^n \rho_1(\hat{\lambda}'\hat{g}_j)/n)^2} - 1 \\ &\xrightarrow{p} \frac{var[\rho_1(\lambda_*' g(z, \beta_*))|\lambda_*' g(z, \beta_*) \in \mathcal{V}]}{E[\rho_1(\lambda_*' g(z, \beta_*))|\lambda_*' g(z, \beta_*) \in \mathcal{V}]^2} > 0. \end{aligned}$$

Therefore, the conclusion follows as $P_n^a \xrightarrow{p} \infty$. Similarly, for P_n^b,

$$\begin{aligned} n^{-1}P_n^b &= n^{-1}\sum_{i=1}^n \frac{(n\hat{\pi}_i - 1)^2}{n\hat{\pi}_i} = \sum_{i=1}^n \rho_1(\hat{\lambda}'\hat{g}_i)/n \sum_{i=1}^n \frac{1}{n\rho_1(\hat{\lambda}'\hat{g}_j)} - 1 \\ &\xrightarrow{p} E[\rho_1(\lambda_*' g(z, \beta_*))^2|\lambda_*' g(z, \beta_*) \in \mathcal{V}] E[\rho_1(\lambda_*' g(z, \beta_*))^{-1}|\lambda_*' g(z, \beta_*) \in \mathcal{V}] - 1 > 0 \end{aligned}$$

by CS so $P_n^b \xrightarrow{p} \infty$. ■

Proof of Theorem 3.4: Follows immediately as $\hat{\mu}_n^s - \mu_n^s \xrightarrow{p} \delta_*$. ■

Proof of Proposition 4.1: The first order conditions determining the GEL and auxiliary parameter estimators $\tilde{\beta}$ and $\tilde{\lambda}$ and Lagrange multiplier estimator $\tilde{\eta}$ are

$$\sum_{i=1}^{n} \rho_1(\tilde{\lambda}'\tilde{g}_i + \tilde{\eta}' r(\tilde{\beta})) \begin{pmatrix} g_i(\tilde{\beta}) \\ G_i(\tilde{\beta})'\tilde{\lambda} + R(\tilde{\beta})'\tilde{\eta} \\ r(\tilde{\beta}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{A.2}$$

It is immediate from eq. (A.2) that the constrained GEL estimator $\tilde{\beta}$ satisfies the parametric constraints; *viz.* $r(\tilde{\beta}) = 0$. Hence, a similar proof to that for Theorem 3.1 of NS establishes that, if Assumption 4.1 holds, $\tilde{\beta} \xrightarrow{p} \beta_0$ and $\tilde{\lambda} \xrightarrow{p} 0$. Therefore, from (A.2), as $\max_{1 \le i \le n} \left| \rho_1(\tilde{\lambda}' g_i(\tilde{\beta})) + 1 \right| \xrightarrow{p} 0$ as in Lemma A1 of NS, using a UWL $\tilde{\eta} \xrightarrow{p} 0$ by Assumption 4.2 (c)(d). Arguments like those in the proof of Theorem 3.2 of NS give

$$\begin{aligned} n^{1/2}\hat{g}(\beta_0) + \Omega n^{1/2}\tilde{\lambda} + G n^{1/2}(\tilde{\beta} - \beta_0) &= o_p(1), \\ G' n^{1/2}\tilde{\lambda} + R' n^{1/2}\tilde{\eta} &= o_p(1), \qquad \text{(A.3)} \\ R n^{1/2}(\tilde{\beta} - \beta_0) &= o_p(1). \qquad \text{(A.4)} \end{aligned}$$

From eq. (A.3),

$$n^{1/2}\tilde{\eta} = -(R\Sigma R')^{-1} R\Sigma G' n^{1/2}\tilde{\lambda} + o_p(1) \tag{A.5}$$

and, thus, substituting back,

$$K G' n^{1/2}\tilde{\lambda} = o_p(1). \tag{A.6}$$

Therefore, premultiplying eq. (A.3) by $KG'\Omega^{-1}$ and using (A.6),

$$K G' \Omega^{-1} n^{1/2}\hat{g}(\beta_0) + K\Sigma^{-1} n^{1/2}(\tilde{\beta} - \beta_0) = o_p(1).$$

Hence, from eq. (A.4),

$$n^{1/2}(\tilde{\beta} - \beta_0) = -K G' \Omega^{-1} n^{1/2}\hat{g}(\beta_0) + o_p(1). \tag{A.7}$$

Substituting (A.7) back into eq. (A.3),

$$n^{1/2}\tilde{\lambda} = -(\Omega^{-1} - \Omega^{-1} G K G' \Omega^{-1}) n^{1/2}\hat{g}(\beta_0) + o_p(1), \tag{A.8}$$

and, thus, from eq. (A.5),

$$n^{1/2}\tilde{\eta} = (R\Sigma R')^{-1}R\Sigma G'\Omega^{-1}n^{1/2}\hat{g}(\beta_0) + o_p(1), \tag{A.9}$$

as $RK = 0$. The result follows immediately from eqs. (A.7)-(A.9) as $n^{1/2}\hat{g}(\beta_0) \xrightarrow{d} N(0,\Omega)$ by a CLT. ■

Lemma A.2 *If Assumptions 4.1 and 4.2 are satisfied, then* $n\tilde{\pi}_i = 1 + o_p(1)$,

$$n^{1/2}\left(\tilde{\pi}_i - \frac{1}{n}\right) = \frac{1}{n}\tilde{g}_i' n^{1/2}\tilde{\lambda}(1 + o_p(1)) + O_p(n^{-3/2}),$$

and

$$n^{1/2}\left(\tilde{\pi}_i - \hat{\pi}_i\right) = \frac{1}{n}\hat{g}_i' n^{1/2}(\hat{\lambda} - \tilde{\lambda})(1 + o_p(1)) + O_p(n^{-3/2}),$$

uniformly $(i = 1, ..., n)$.

Proof: The first and second conclusions follow by a similar argument to that of Lemma A.1. Therefore,

$$\begin{aligned} n^{1/2}(\tilde{\pi}_i - \hat{\pi}_i) &= (\frac{1}{n}\tilde{g}_i' n^{1/2}\tilde{\lambda} - \frac{1}{n}\hat{g}_i' n^{1/2}\hat{\lambda})(1 + o_p(1)) + O_p(n^{-3/2}) \\ &= \frac{1}{n}\hat{g}_i' n^{1/2}(\hat{\lambda} - \tilde{\lambda})(1 + o_p(1)) + O_p(n^{-3/2}) \end{aligned}$$

uniformly $(i = 1, ..., n)$ as $G_i(\beta) = O_p(1)$, $\tilde{\beta} - \hat{\beta} = O_p(n^{-1/2})$ and $\tilde{\lambda} = O_p(n^{-1/2})$. ■

Proof of Lemma 4.1: From Lemma A.2 and similarly to the Proof of Lemma 3.1,

$$\begin{aligned} n^{1/2}[\tilde{\mu}_n(z) - \mu_n(z)] &= n^{1/2}\sum_{i=1}^{n}\left(\tilde{\pi}_i - \frac{1}{n}\right)1(z_i \le z) \\ &= \sum_{i=1}^{n}[n^{-1}\tilde{g}_i' n^{1/2}\tilde{\lambda}(1 + o_p(1)) + O_p(n^{-3/2})]1(z_i \le z) \\ &= \left(\sum_{i=1}^{n}1(z_i \le z)\tilde{g}_i'/n\right)n^{1/2}\tilde{\lambda}(1 + o_p(1)) + O_p(n^{-1/2}) \\ &= [b(z) + O_p(n^{-1/2})]'n^{1/2}\tilde{\lambda} + o_p(1) \\ &\Rightarrow \tilde{\Lambda}(z) \end{aligned}$$

where $\tilde{\Lambda}$ is Gaussian stochastic process on $\mathcal{R}^k$ with mean zero and covariance function $E[\tilde{\Lambda}(z_1)\tilde{\Lambda}(z_2)] = b(z_1)'(\Omega^{-1} - \Omega^{-1}GKG'\Omega^{-1})b(z_2)$ using eq. (A.8). From eq. (A.10) and Lemma A.2

$$\begin{aligned}
n^{1/2}[\tilde{\mu}_n(z) - \hat{\mu}_n(z)] &= n^{1/2}\sum_{i=1}^{n}(\tilde{\pi}_i - \hat{\pi}_i)\,1(z_i \leq z) \\
&= \sum_{i=1}^{n}[(\frac{1}{n}\tilde{g}_i' n^{1/2}\tilde{\lambda} - \frac{1}{n}\hat{g}_i' n^{1/2}\hat{\lambda})(1 + o_p(1)) + O_p(n^{-3/2})]1(z_i \leq z) \\
&= \left(\sum_{i=1}^{n} 1(z_i \leq z)\hat{g}_i'/n + O_p(n^{-1/2})\right) n^{1/2}(\tilde{\lambda} - \hat{\lambda})(1 + o_p(1)) + O_p(n^{-1/2}) \\
&= [b(z) + O_p(n^{-1/2})]' n^{1/2}(\tilde{\lambda} - \hat{\lambda}) + o_p(1) \\
&\Rightarrow \Delta(z)
\end{aligned}$$

where Δ is Gaussian stochastic process on $\mathcal{R}^k$ with mean zero and covariance function $E[\Delta(z_1)\Delta(z_2)] = b(z_1)'\Omega^{-1}G\Sigma R'(R\Sigma R')^{-1}R\Sigma G\Omega^{-1}b(z_2)$ as

$$n^{1/2}(\tilde{\lambda} - \hat{\lambda}) = -\Omega^{-1}G\Sigma R'(R\Sigma R')^{-1}R\Sigma G\Omega^{-1}n^{1/2}\hat{g}(\beta_0) + o_p(1)$$

using eq. (A.8) and $n^{1/2}\hat{\lambda} = -Pn^{1/2}\hat{g}(\beta_0) + o_p(1)$ from the Proof of Theorem 3.2 in NS.

Proof of Theorem 4.1: From Lemma A.2, it follows immediately that

$$(n\tilde{\pi}_i - n\hat{\pi}_i)^2 = ((\hat{\lambda} - \tilde{\lambda})'\hat{g}_i(1 + o_p(1)) + O_p(n^{-1}))^2,$$

uniformly ($i = 1, ..., n$). Summing over $i = 1, ..., n$,

$$\begin{aligned}
\sum_{i=1}^{n}(n\tilde{\pi}_i - n\hat{\pi}_i)^2 &= n(\hat{\lambda} - \tilde{\lambda})'(\sum_{i=1}^{n}\hat{g}_i\hat{g}_i'/n)(\hat{\lambda} - \tilde{\lambda})(1 + o_p(1)) \\
&\quad + n^{1/2}(\hat{\lambda} - \tilde{\lambda})'(\sum_{i=1}^{n}\hat{g}_i/n^{1/2})(1 + o_p(1))O_p(n^{-1}) + O_p(n^{-1}) \\
&= n(\hat{\lambda} - \tilde{\lambda})'(\sum_{i=1}^{n}\hat{g}_i\hat{g}_i'/n)(\hat{\lambda} - \tilde{\lambda}) + o_p(1) \\
&= n(\hat{\lambda} - \tilde{\lambda})'\Omega(\hat{\lambda} - \tilde{\lambda}) + o_p(1) \\
&= n\hat{g}(\beta_0)'\Omega^{-1}G\Sigma R(R\Sigma R')^{-1}R\Sigma G\Omega^{-1}\hat{g}(\beta_0) + o_p(1) \\
&= nr(\hat{\beta})'(\hat{R}\hat{\Sigma}\hat{R}')^{-1}r(\hat{\beta}) + o_p(1),
\end{aligned}$$

the first term of which is the Wald test statistic for $r(\beta_0) = 0$ which has a limiting chi-square distribution with r degrees of freedom. See Newey and West (1987) and Smith (2001, section 5). From Lemmas A.1 and A.2

$$\begin{aligned}\sum_{i=1}^{n}(n\tilde{\pi}_i - n\hat{\pi}_i)^2 &= \sum_{i=1}^{n}\frac{(n\hat{\pi}_i - n\tilde{\pi}_i)^2}{n\hat{\pi}_i} + o_p(1) \\ &= \sum_{i=1}^{n}\frac{(n\hat{\pi}_i - n\tilde{\pi}_i)^2}{n\tilde{\pi}_i} + o_p(1).\end{aligned}$$

■

Proof of Theorem 4.2: From Lemma 4.1, as $n^{1/2}(\tilde{\mu}_n^s - \hat{\mu}_n^s) = -B_s' n^{1/2}(\tilde{\lambda}-\hat{\lambda}) + o_p(1)$,

$$\begin{aligned}P_n^{a,alt,r} &= n(\tilde{\lambda}-\hat{\lambda})'G\Sigma G'(\tilde{\lambda}-\hat{\lambda}) + o_p(1) \\ &= n\hat{g}(\beta_0)'\Omega^{-1}G\Sigma R(R\Sigma R')^{-1}R\Sigma G\Omega^{-1}\hat{g}(\beta_0) + o_p(1) \\ &= nr(\hat{\beta})'(\hat{R}\hat{\Sigma}\hat{R}')^{-1}r(\hat{\beta}) + o_p(1),\end{aligned}$$

which from the Proof of Theorem 4.1 is asymptotically equivalent to $P_n^{a,r}$, $P_n^{b,r}$ and $P_n^{c,r}$. Similarly, from Lemma 4.1, $n^{1/2}(\tilde{\mu}_n^s - \hat{\mu}_n^s) = B_s' n^{1/2}\hat{\lambda} + o_p(1)$. Therefore, from the Proof of Proposition 4.1, as $n^{1/2}\hat{\lambda} = -(\Omega^{-1} - \Omega^{-1}GKG'\Omega^{-1})n^{1/2}\hat{g}(\beta_0) + o_p(1)$ and $G'(\Omega^{-1} - \Omega^{-1}GKG'\Omega^{-1}) = R'(R\Sigma R')^{-1}R\Sigma G'\Omega^{-1}$,

$$\begin{aligned}P_n^{b,alt,r} &= n\hat{\lambda}'G\Sigma G'\hat{\lambda} + o_p(1) \\ &= n\hat{g}(\beta_0)'\Omega^{-1}G\Sigma R(R\Sigma R')^{-1}R\Sigma G\Omega^{-1}\hat{g}(\beta_0) + o_p(1).\end{aligned}$$

■

Proof of Theorem 4.3: The proof is very similar in outline to that of Theorem 3.3. Firstly, $\tilde{\lambda}(\beta) = \arg\max_{\lambda\in\tilde{\Lambda}_n(\beta)} \hat{P}(\beta,\lambda)$, $\beta \in \mathcal{B}^r$, exists w.p.a.1 and thus $\tilde{\lambda}(\beta) \xrightarrow{p} \lambda(\beta)$, $\beta \in \mathcal{B}^r$. Secondly, the restricted GEL estimator $\tilde{\beta} \xrightarrow{p} \beta_*$, $\beta_* \in \mathcal{B}^r$. As in the Proof of Theorem 3.3, $\sum_{i=1}^n \rho_1(\lambda' g_i(\beta))/n \xrightarrow{p} E[\rho_1(\lambda' g(z,\beta))|\lambda' g(z,\beta) \in \mathcal{V}]$ and $\sum_{i=1}^n \rho_1(\lambda' g_i(\beta))^2/n \xrightarrow{p} E[\rho_1(\lambda' g(z,\beta))^2|\lambda' g(z,\beta) \in \mathcal{V}]$ uniformly $\beta \in \mathcal{B}^r$ and λ. Therefore, by a UWL, $\sum_{i=1}^n \rho(\tilde{\lambda}' g_i(\tilde{\beta}))/n \xrightarrow{p} E[\rho_1(\lambda_*' g(z,\beta_*))|\lambda_*' g(z,\beta_*) \in \mathcal{V}]$ and $\sum_{i=1}^n \rho(\tilde{\lambda}' g_i(\tilde{\beta}))^2/n \xrightarrow{p} E[\rho_1(\lambda_*' g(z,\beta_*))^2|\lambda_*' g(z,\beta_*) \in \mathcal{V}]$.

Consider the statistic $P_n^{c,r}$.

$$\begin{aligned}
n^{-1}P_n^{c,r} &= \sum_{i=1}^{n}\left(n\tilde{\pi}_i - n\hat{\pi}_i\right)^2/n \\
&= \frac{\sum_{i=1}^n \rho_1(\tilde{\lambda}'g_i(\tilde{\beta}))^2/n}{(\sum_{j=1}^n \rho_1(\tilde{\lambda}'g_j(\tilde{\beta}))/n)^2} - 2\frac{\sum_{i=1}^n \rho_1(\tilde{\lambda}'g_i(\tilde{\beta}))\rho_1(\hat{\lambda}'g_i(\hat{\beta}))/n}{(\sum_{j=1}^n \rho_1(\tilde{\lambda}'g_j(\tilde{\beta}))/n)(\sum_{j=1}^n \rho_1(\hat{\lambda}'g_j(\hat{\beta}))/n)} \\
&\quad + \frac{\sum_{i=1}^n \rho_1(\hat{\lambda}'g_i(\hat{\beta}))^2/n}{(\sum_{j=1}^n \rho_1(\hat{\lambda}'g_j(\hat{\beta}))/n)^2} \\
&= \left(\frac{\sum_{i=1}^n \rho_1(\tilde{\lambda}'g_i(\tilde{\beta}))^2/n}{(\sum_{j=1}^n \rho_1(\tilde{\lambda}'g_j(\tilde{\beta}))/n)^2} - 1\right) + o_p(1) \\
&\xrightarrow{p} \frac{var[\rho_1(\lambda_*'g(z,\beta_*))|\lambda_*'g(z,\beta_*)\in\mathcal{V}]}{E[\rho_1(\lambda_*'g(z,\beta_*))|\lambda_*'g(z,\beta_*)\in\mathcal{V}]^2} > 0.
\end{aligned}$$

The third equality follows as $\rho_1(\hat{\lambda}'g(z_i,\hat{\beta})) = -1 + o_p(1)$, uniformly $(i = 1, ..., n)$, from Lemma A1 in NS, $\sum_{j=1}^n \rho_1(\hat{\lambda}'g(z_j,\hat{\beta}))^2/n \xrightarrow{p} 1$ and $\sum_{j=1}^n \rho_1(\hat{\lambda}'g(z_j,\hat{\beta}))/n \xrightarrow{p} -1$. The conclusion $P_n^{c,r} \xrightarrow{p} \infty$ is then immediate. Similarly, for $P_n^{a,r}$,

$$\begin{aligned}
n^{-1}P_n^{a,r} &= \sum_{i=1}^{n}\frac{\left(n\tilde{\pi}_i - n\hat{\pi}_i\right)^2}{\hat{\pi}_i} \\
&= \frac{(\sum_{i=1}^n \rho_1(\tilde{\lambda}'\tilde{g}_i)^2/n\rho_1(\hat{\lambda}'\hat{g}_i))(\sum_{j=1}^n \rho_1(\hat{\lambda}'\hat{g}_j)/n)}{(\sum_{j=1}^n \rho_1(\tilde{\lambda}'\tilde{g}_j)/n)^2} - 1 \\
&= \left(\frac{(\sum_{i=1}^n \rho_1(\tilde{\lambda}'\tilde{g}_i)^2/n)}{(\sum_{j=1}^n \rho_1(\tilde{\lambda}'\tilde{g}_j)/n)^2} - 1\right) + o_p(1) \\
&= n^{-1}P_n^{c,r} + o_p(1).
\end{aligned}$$

For $P_n^{b,r}$,

$$\begin{aligned}
n^{-1}P_n^{b,r} &= \sum_{i=1}^{n}\frac{\left(n\tilde{\pi}_i - n\hat{\pi}_i\right)^2}{\tilde{\pi}_i} \\
&= \frac{(\sum_{i=1}^n \rho_1(\hat{\lambda}'\hat{g}_i)^2/n\rho_1(\tilde{\lambda}'\tilde{g}_i))(\sum_{j=1}^n \rho_1(\tilde{\lambda}'\tilde{g}_j)/n)}{(\sum_{j=1}^n \rho_1(\hat{\lambda}'\hat{g}_j)/n)^2} - 1
\end{aligned}$$

$$\begin{aligned}
&\xrightarrow{p} E[\rho_1(\lambda_*'g(z,\beta_*))|\lambda_*'g(z,\beta_*) \in \mathcal{V}] \\
&\times E[\rho_1(\lambda_*'g(z,\beta_*))^{-1}|\lambda_*'g(z,\beta_*) \in \mathcal{V}] - 1 > 0
\end{aligned}$$

by CS so $P_n^{b,r} \xrightarrow{p} \infty$. ■

Proof of Theorem 4.4: Follows immediately as $\tilde{\mu}_n^s - \hat{\mu}_n^s \xrightarrow{p} \delta_*$ and $\tilde{\mu}_n^s - \mu_n^s \xrightarrow{p} \delta_*$. ■

References

Andrews, D. W. K. (1988a): "Chi-Square Diagnostic Tests for Econometric Models: Introduction and Applications", *Journal of Econometrics*, 37, 135-156.

Andrews, D. W. K. (1988b): "Chi-Square Diagnostic Tests for Econometric Models: Theory", *Econometrica*, 56, 1419-1453.

Back, K. and D. P. Brown (1993): "Implied probabilities in GMM estimators", *Econometrica*, 61(4), 971-975.

Brown, B.W. and W.K. Newey (1992): "Bootstrapping for GMM", *mimeo*, M.I.T.

Brown, B.W. and W.K. Newey (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica* 66, 453-464.

Brown, B.W. and W.K. Newey (2003): "Generalised Method of Moments, Efficient Bootstrapping, and Improved Inference," *Journal of Business and Economic Statistics*, 20, 507-517.

Cressie, N., and T. Read (1984): "Multinomial Goodness-of-Fit Tests", *Journal of the Royal Statistical Society Series B*, 46, 440-464.

Durbin, J. (1973): *Distribution Theory for Tests Based on the Sample Distribution Function.* CBMS-NSF Regional Conference Series in Applied Mathematics No.9. SIAM: Philadelphia.

Hansen, L. P. (1982): "Large sample properties of generalised method of moments estimators", *Econometrica*, 50(4), 1029-1054.

Hansen, L. P., J. Heaton and A. Yaron (1996): "Finite-sample properties of some alternative GMM estimators", *Journal of Business & Economic Statistics*, 14(3), 262-280.

Imbens, G. W. (1997): "One-step estimators for over-identified generalised method of moments models", *Review of Economic Studies*, 64, 359-383.

Imbens, G. W., R. H. Spady and P. Johnson (1998): "Information theoretic approaches to inference in moment condition models", *Econometrica* 66, 333-357.

Kitamura, Y. and M. Stutzer (1997): "An information-theoretic alternative to generalised method of moments estimation", *Econometrica*, 65(4), 861-874.

Newey, W.K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in Engle, R. and D. McFadden, eds., *Handbook of Econometrics, Vol. 4,* New York: North Holland.

Newey, W. K., J. J. S. Ramalho, and R. J. Smith, (2002): "Asymptotic Bias for GMM and GEL Estimators with Estimated Nuisance Parameters". Forthcoming in *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, eds. D.W.K. Andrews and J.H. Stock. Cambridge University Press: Cambridge..

Newey, W. K. and R. J. Smith (2001): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators", *Econometrica*, 72, 219-255.

Owen, A. (2001): *Empirical Likelihood.* New York: Chapman and Hall.

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

Rao, C. R. and S. K. Mitra (1971): *Generalized Inverse of Matrices and Its Applications.* Wiley: New York.

Shorack, G.R., and J.A. Wellner (1986): *Empirical Processes with Applications in Statistics.* Wiley: New York.

Smith, R. J. (1997), "Alternative Semi-Parametric Likelihood Approaches to Generalised Method of Moments Estimation", *Economic Journal*, 107, 503-519.

Smith, R. J. (2001): "GEL Methods for Moment Condition Models", working paper, Department of Economics, University of Bristol.

Table 1: Tests of Over-Identifying Moment Conditions: Asset-Pricing Model
Empirical Size: S_n, LM_n and $GELR_n$

n	Nominal Size	S_n									LM_n						$GELR_n$	
		2s	*i*	*cue*	*et(n)*	*et(s)*	*et(r)*	*el(n)*	*el(s)*	*el(r)*	*et(n)*	*et(s)*	*et(r)*	*el(n)*	*el(s)*	*el(r)*	*et*	*el*
100	20.0	26.7	26.1	24.0	25.5	29.8	29.2	25.9	28.3	28.5	29.9	25.3	26.9	28.1	28.3	24.8	27.2	27.9
	10.0	17.5	16.7	12.8	15.9	19.6	20.1	16.4	17.6	19.3	19.8	16.1	14.8	18.7	17.6	15.3	16.9	17.0
	5.0	12.2	11.3	7.2	10.6	13.7	14.6	11.2	11.2	14.0	13.8	11.0	8.3	13.6	11.2	10.4	11.0	11.1
	2.5	9.5	8.5	4.5	7.8	9.9	11.5	8.3	7.4	10.8	10.3	8.3	4.6	10.4	7.4	7.6	7.7	7.3
	1.0	6.9	5.9	2.5	5.2	6.9	8.7	5.7	4.3	8.1	7.3	5.8	2.2	7.6	4.3	5.3	5.0	4.1
	0.5	5.7	4.4	1.6	3.7	5.4	7.2	4.2	2.9	7.0	5.6	4.4	1.3	6.1	2.9	4.0	3.6	2.8
	0.1	3.9	2.3	0.7	1.7	3.3	4.7	2.1	1.1	4.6	3.5	2.4	0.4	4.0	1.1	2.7	1.8	1.2
200	20.0	25.3	25.1	24.2	24.8	28.0	27.7	25.0	26.8	27.2	28.1	24.7	25.7	26.0	26.8	23.6	26.0	26.8
	10.0	15.0	14.7	13.3	14.4	17.3	18.0	14.7	15.7	17.3	17.4	14.5	13.6	16.5	15.7	13.9	15.2	15.5
	5.0	9.9	9.5	7.6	9.2	11.2	12.5	9.5	9.4	11.9	11.3	9.5	7.1	11.2	9.4	8.9	9.3	9.1
	2.5	6.8	6.6	4.5	6.2	7.8	9.1	6.5	5.8	8.5	7.8	6.5	3.8	8.1	5.8	6.6	5.8	5.6
	1.0	4.6	4.3	2.4	4.0	5.0	6.2	4.3	2.9	6.0	5.0	4.3	1.5	5.9	2.9	4.6	3.6	3.0
	0.5	3.5	3.1	1.4	2.9	3.5	4.9	3.1	1.9	4.8	3.6	3.1	0.9	4.8	1.9	3.6	2.5	1.9
	0.1	1.9	1.5	0.4	1.3	2.0	3.0	1.5	0.7	3.0	2.1	1.6	<u>0.2</u>	3.2	0.7	2.1	1.1	0.7
500	20.0	23.1	23.0	22.7	22.8	25.6	25.4	22.9	24.7	25.1	25.7	22.9	23.7	24.2	24.7	22.1	24.0	24.4
	10.0	13.1	13.0	12.4	12.8	15.0	15.6	13.0	13.5	15.0	15.0	13.0	12.0	14.4	13.5	12.4	13.4	13.6
	5.0	8.0	7.9	7.3	7.6	9.3	10.1	7.8	7.9	9.7	9.4	7.8	6.3	9.3	7.9	7.9	7.8	7.7
	2.5	5.0	4.9	4.1	4.7	6.2	6.9	4.9	4.6	6.6	6.2	4.9	3.4	6.6	4.6	5.4	4.7	4.6
	1.0	3.0	3.0	2.3	2.9	3.7	4.4	3.0	2.4	4.2	3.7	3.0	1.3	4.5	2.4	3.7	2.5	2.2
	0.5	2.1	2.1	1.5	1.9	2.4	3.3	2.1	1.4	3.2	2.5	2.0	0.6	3.7	1.4	2.8	1.6	1.3
	0.1	0.8	0.9	0.4	0.7	1.1	1.7	0.8	0.5	1.8	1.2	0.8	<u>0.1</u>	2.4	0.5	1.8	0.5	0.4
1000	20.0	21.8	21.8	21.6	21.7	23.5	23.5	21.9	22.6	23.2	23.5	21.8	22.0	22.3	22.6	20.8	22.5	22.7
	10.0	11.9	11.8	11.6	11.8	13.2	13.9	11.9	12.1	13.6	13.3	11.9	11.2	12.5	12.1	11.2	12.3	12.3
	5.0	6.7	6.8	6.5	6.7	8.0	8.5	6.8	7.1	8.3	8.1	6.8	5.9	8.0	7.1	6.6	6.9	6.8
	2.5	4.4	4.3	4.1	4.3	4.8	5.5	4.4	3.9	5.4	5.0	4.4	3.0	5.2	3.9	4.3	4.1	3.9
	1.0	2.4	2.4	2.3	2.4	2.5	3.5	2.4	1.8	3.4	2.6	2.4	1.2	3.4	1.8	2.5	2.2	1.9
	0.5	1.7	1.7	1.6	1.7	1.7	2.4	1.7	1.0	2.5	1.7	1.7	0.6	2.4	1.0	1.9	1.3	1.0
	0.1	0.7	0.7	0.6	0.6	0.7	1.2	0.6	0.3	1.3	0.7	0.7	<u>0.1</u>	1.5	0.3	1.0	0.5	0.3

Note: Empirical size closest to nominal size is underlined.

Table 2: Tests of Over-Identifying Moment Conditions: Asset-Pricing Model
Empirical Size: Pearson-Type Tests

n	Nominal Size	P_n^a *et*	P_n^a *el*	P_n^b *et*	P_n^b *el*	P_n^{alt} (s = 8) *et(n)*	*et(s)*	*et(r)*	*el(n)*	*el(s)*	*el(r)*	P_n^{alt} (s = 16) *et(n)*	*et(s)*	*et(r)*	*el(n)*	*el(s)*	*el(r)*
100	20.0	26.7	28.6	30.4	28.3	25.9	22.5	20.5	26.1	23.0	22.1	25.7	22.5	20.4	25.7	23.0	21.6
	10.0	17.0	19.3	20.4	17.6	15.0	15.5	9.6	14.5	14.8	8.8	14.9	15.7	9.8	14.1	15.1	8.6
	5.0	11.8	14.0	14.6	11.2	9.2	11.0	4.2	7.8	10.4	2.6	9.2	11.2	4.9	7.9	10.6	2.7
	2.5	8.9	10.8	10.9	7.4	5.6	8.0	1.4	4.2	7.1	0.4	5.8	8.4	2.0	4.3	7.6	0.6
	1.0	6.4	8.1	7.9	4.3	2.9	4.9	0.2	1.7	4.1	0.0	3.1	5.3	0.4	1.9	4.7	0.1
	0.5	5.0	7.0	6.2	2.9	1.7	2.9	0.0	0.8	2.5	0.0	2.1	3.6	0.1	1.1	3.0	0.0
	0.1	2.8	4.6	4.3	1.1	0.5	0.8	0.0	0.2	0.7	0.0	0.8	1.2	0.0	0.2	1.1	0.0
200	20.0	25.5	27.2	28.5	26.8	25.6	21.6	21.5	25.8	23.4	22.7	25.4	21.6	21.4	25.5	22.9	22.4
	10.0	15.4	17.3	18.1	15.7	14.7	14.2	10.3	14.3	13.4	10.7	14.5	14.3	10.2	14.0	13.4	10.1
	5.0	10.0	11.9	12.0	9.4	8.7	10.1	4.9	8.0	8.8	4.4	8.6	10.1	5.0	7.6	8.9	4.0
	2.5	7.0	8.5	8.7	5.8	5.1	7.2	2.5	4.5	6.0	1.4	5.0	7.3	2.6	4.3	6.0	1.3
	1.0	4.7	6.0	5.7	2.9	2.8	4.8	0.8	2.1	3.7	0.4	2.7	4.8	1.0	1.8	3.8	0.3
	0.5	3.5	4.8	4.3	1.9	1.7	3.4	0.3	1.1	2.6	0.1	1.7	3.5	0.4	1.0	2.5	0.1
	0.1	1.9	3.0	2.7	0.7	0.6	1.4	0.0	0.4	1.1	0.0	0.7	1.6	0.0	0.3	1.0	0.0
500	20.0	23.6	25.1	26.0	24.7	24.4	21.1	21.6	24.4	22.8	22.2	24.2	21.0	21.3	24.2	22.6	22.0
	10.0	13.5	15.0	15.6	13.5	13.7	12.4	10.5	13.4	12.7	11.0	13.5	12.3	10.4	12.8	12.3	10.7
	5.0	8.3	9.7	9.9	7.9	8.0	8.1	5.1	7.9	7.2	5.6	7.8	8.1	5.0	7.6	7.0	5.3
	2.5	5.3	6.6	6.8	4.6	4.7	5.6	2.6	4.7	4.5	2.6	4.6	5.6	2.5	4.3	4.4	2.4
	1.0	3.3	4.2	4.2	2.4	2.5	3.6	1.0	2.4	2.7	1.0	2.4	3.6	1.0	2.1	2.6	0.7
	0.5	2.3	3.2	3.1	1.4	1.5	2.6	0.5	1.5	1.8	0.5	1.5	2.6	0.5	1.2	1.7	0.3
	0.1	1.0	1.8	1.7	0.5	0.5	1.2	0.1	0.6	0.7	0.1	0.4	1.1	0.1	0.4	0.6	0.0
1000	20.0	22.1	23.2	23.8	22.6	23.0	20.8	21.1	22.7	22.0	21.2	23.0	20.6	20.9	22.6	21.7	21.1
	10.0	12.3	13.6	13.6	12.1	12.7	11.5	10.2	12.3	11.9	10.6	12.6	11.5	10.1	12.2	11.8	10.3
	5.0	7.2	8.3	8.4	7.1	7.2	7.1	5.2	7.1	6.5	5.6	7.1	7.0	5.2	7.0	6.3	5.3
	2.5	4.6	5.4	5.4	3.9	4.2	4.8	2.7	4.1	4.0	2.6	4.1	4.7	2.7	3.9	4.0	2.5
	1.0	2.6	3.4	3.0	1.8	2.3	3.0	1.1	2.0	2.3	1.0	2.2	3.0	1.0	1.8	2.2	0.9
	0.5	1.8	2.5	2.1	1.0	1.2	2.1	0.6	1.2	1.5	0.5	1.2	2.1	0.5	1.1	1.4	0.4
	0.1	0.7	1.3	1.0	0.3	0.5	1.0	0.1	0.4	0.5	0.1	0.4	1.0	0.1	0.4	0.5	0.1

Note: Empirical size closest to nominal size is underlined.

Table 3: Tests of Over-Identifying Moment Conditions: Chi-Square Moments Model
Empirical Size: S_n, LM_n and $GELR_n$

n	Nominal	S_n									LM_n						$GELR_n$	
	Size	*2s*	*i*	*cue*	*et(n)*	*et(s)*	*et(r)*	*el(n)*	*el(s)*	*el(r)*	*et(n)*	*et(s)*	*et(r)*	*el(n)*	*el(s)*	*el(r)*	*et*	*el*
100	20.0	34.6	34.6	34.6	34.7	38.0	38.0	35.0	36.5	37.2	37.7	34.1	35.1	33.8	36.5	31.2	35.7	36.3
	10.0	27.0	26.9	26.9	27.2	28.0	30.1	27.5	25.9	29.5	27.6	27.0	23.4	25.4	25.9	22.8	26.5	26.0
	5.0	22.3	22.3	22.3	22.6	21.5	25.2	23.1	19.3	24.9	21.3	22.3	16.9	20.3	19.3	17.8	20.7	19.3
	2.5	18.8	18.8	18.8	19.1	17.6	21.7	19.8	14.8	21.8	17.6	19.3	12.7	16.7	14.8	14.4	17.1	15.5
	1.0	15.5	15.5	15.5	15.8	13.4	18.2	16.7	11.1	18.9	13.5	16.2	9.6	13.4	11.1	11.3	13.9	11.5
	0.5	13.4	13.4	13.4	13.8	11.5	16.4	14.7	9.0	17.3	11.7	14.2	8.0	11.8	9.0	9.5	12.0	9.9
	0.1	9.8	9.8	9.8	10.2	8.1	13.0	11.3	5.9	14.3	8.6	10.8	5.4	8.9	5.9	7.3	8.9	6.8
200	20.0	29.0	29.0	29.0	29.0	30.9	31.9	29.1	29.6	31.2	30.5	28.5	28.2	26.0	29.6	23.1	29.6	30.0
	10.0	20.9	20.9	20.9	21.0	21.1	23.8	21.1	19.5	22.9	20.8	20.8	16.8	17.4	19.5	14.7	20.2	19.8
	5.0	16.4	16.4	16.4	16.5	14.6	19.0	16.8	12.9	18.4	14.4	16.5	10.5	12.8	12.9	10.8	15.1	13.6
	2.5	13.8	13.8	13.8	13.8	11.0	15.8	14.1	8.9	15.7	10.9	13.9	6.9	9.8	8.9	8.5	11.7	9.7
	1.0	10.4	10.4	10.4	10.6	7.6	12.7	11.0	5.8	13.1	7.6	10.6	4.4	7.8	5.8	6.3	8.7	6.5
	0.5	9.1	9.0	9.0	9.2	6.1	10.8	9.5	4.1	11.3	6.1	9.2	3.1	6.5	4.1	5.3	7.2	5.1
	0.1	6.3	6.3	6.3	6.5	3.5	8.4	6.9	2.1	9.0	3.5	6.6	1.7	4.7	2.1	3.8	4.7	2.9
500	20.0	25.4	25.4	25.4	25.4	26.3	27.9	25.5	25.3	27.3	26.1	25.2	24.0	21.7	25.3	19.1	26.4	26.1
	10.0	16.4	16.4	16.4	16.5	15.6	18.8	16.5	14.4	18.3	15.4	16.3	12.5	13.1	14.4	11.3	15.7	14.8
	5.0	11.5	11.5	11.5	11.5	9.8	13.3	11.6	8.7	13.0	9.7	11.4	6.9	9.0	8.7	7.8	10.2	9.1
	2.5	8.6	8.6	8.6	8.6	6.6	10.3	8.6	5.3	10.2	6.5	8.5	3.7	6.7	5.3	5.8	7.1	5.9
	1.0	6.3	6.3	6.3	6.4	3.7	7.5	6.4	2.7	7.7	3.6	6.3	1.7	4.9	2.7	4.0	5.0	3.3
	0.5	5.2	5.2	5.2	5.2	2.6	6.2	5.3	1.8	6.5	2.6	5.2	1.0	4.0	1.8	3.2	3.6	2.3
	0.1	3.0	3.0	3.0	3.1	1.2	4.2	3.2	0.7	4.7	1.2	3.1	0.3	2.7	0.7	2.0	2.0	1.0
1000	20.0	23.2	23.2	23.2	23.2	24.4	25.1	23.3	23.6	24.6	24.2	23.1	22.5	20.8	23.6	18.8	23.8	23.5
	10.0	14.0	14.0	14.0	14.1	13.9	16.0	14.1	12.9	15.7	13.7	14.0	11.6	12.0	12.9	10.7	13.9	13.2
	5.0	9.1	9.1	9.1	9.1	8.4	10.8	9.1	7.4	10.3	8.3	9.1	6.1	8.0	7.4	7.0	8.5	7.9
	2.5	6.5	6.5	6.5	6.5	5.0	7.9	6.5	4.1	7.7	4.8	6.4	2.9	5.8	4.1	5.2	5.6	4.5
	1.0	4.3	4.3	4.3	4.3	2.7	5.3	4.3	2.0	5.5	2.7	4.2	1.1	4.2	2.0	3.6	3.2	2.2
	0.5	3.2	3.2	3.2	3.2	1.7	4.1	3.2	1.2	4.3	1.7	3.2	0.6	3.4	1.2	2.8	2.1	1.3
	0.1	1.7	1.7	1.7	1.7	0.7	2.2	1.7	0.4	2.5	0.7	1.7	0.1	2.1	0.4	1.6	0.9	0.4

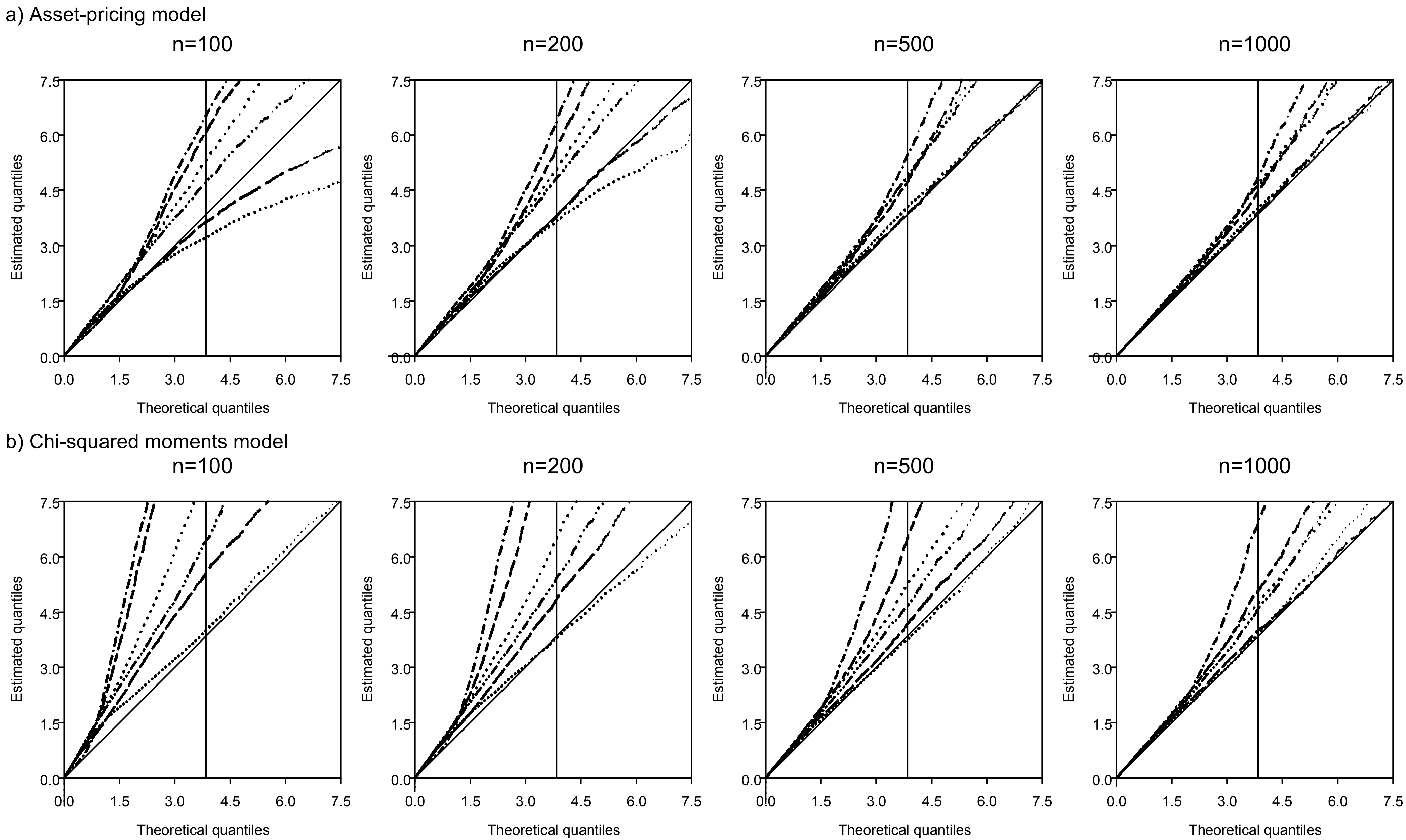
Note: Empirical size closest to nominal size is underlined.

Table 4: Tests of Over-Identifying Moment Conditions: Chi-Square Moments Model
Empirical Size: Pearson-Type Tests

n	Nominal	P_n^a		P_n^b		P_n^{alt} (s = 8)						P_n^{alt} (s = 16)					
	Size	*et*	*el*	*et*	*el*	*et(n)*	*et(s)*	*et(r)*	*el(n)*	*el(s)*	*el(r)*	*et(n)*	*et(s)*	*et(r)*	*el(n)*	*el(s)*	*el(r)*
100	20.0	35.6	37.2	38.4	36.5	32.7	33.7	28.0	32.4	33.7	<u>27.7</u>	34.6	34.9	30.0	35.5	34.2	32.7
	10.0	27.9	29.5	28.5	25.9	21.6	28.0	17.2	19.9	26.1	<u>12.1</u>	24.8	29.9	21.2	24.5	28.8	19.4
	5.0	23.5	24.9	22.1	19.3	15.4	23.4	10.5	12.4	21.2	<u>5.5</u>	18.6	25.8	15.5	17.3	24.9	10.7
	2.5	20.0	21.8	18.2	14.8	10.9	19.5	6.3	7.8	17.2	<u>2.7</u>	14.8	22.5	11.1	12.9	21.6	5.2
	1.0	17.1	18.9	14.2	11.1	7.4	15.3	2.9	4.6	13.5	<u>1.2</u>	11.1	18.9	6.4	8.8	18.2	2.5
	0.5	15.2	17.3	12.2	9.0	5.4	12.5	1.6	3.3	11.0	<u>0.5</u>	9.1	16.8	3.9	6.6	16.5	1.5
	0.1	12.0	14.3	9.0	5.9	3.0	5.8	<u>0.0</u>	1.7	5.9	<u>0.0</u>	5.4	12.5	1.0	3.6	12.8	0.3
200	20.0	29.8	31.3	31.3	29.6	28.1	27.8	23.8	27.4	28.3	23.9	29.1	28.6	25.0	29.2	28.5	26.7
	10.0	21.5	22.9	21.8	19.5	17.8	21.9	13.5	16.0	19.7	<u>10.9</u>	19.3	23.2	15.9	18.8	21.3	14.7
	5.0	17.2	18.4	15.5	12.9	11.4	17.9	7.8	9.5	15.2	<u>4.9</u>	13.6	19.4	10.7	12.2	17.5	7.6
	2.5	14.4	15.7	11.7	8.9	7.7	15.1	4.7	5.8	12.0	<u>2.3</u>	9.8	16.6	7.3	8.3	14.9	3.5
	1.0	11.5	13.1	8.4	5.8	4.8	11.8	2.2	3.3	9.0	<u>0.8</u>	6.9	13.9	4.5	5.1	12.0	1.3
	0.5	9.8	11.3	6.8	4.1	3.3	9.8	1.4	2.1	7.2	0.4	5.4	11.9	2.9	3.4	10.2	<u>0.5</u>
	0.1	7.4	9.0	4.1	2.1	1.6	6.8	0.4	0.7	4.6	<u>0.1</u>	3.0	9.0	1.2	1.6	7.5	<u>0.1</u>
500	20.0	26.0	27.3	26.8	25.3	25.4	23.8	22.6	23.7	25.3	21.1	26.1	24.1	23.2	25.1	25.2	22.7
	10.0	17.2	18.3	16.1	14.4	14.2	17.0	11.1	13.0	15.0	<u>10.0</u>	15.3	17.7	11.9	14.1	16.1	11.3
	5.0	12.0	13.0	10.5	8.7	8.5	12.7	5.8	7.5	9.7	<u>4.8</u>	9.5	13.8	7.0	8.4	11.2	5.8
	2.5	9.1	10.2	7.2	5.3	5.4	9.9	3.3	4.3	7.0	<u>2.3</u>	6.3	11.0	4.5	5.5	8.5	<u>2.7</u>
	1.0	6.6	7.7	4.5	2.7	2.8	7.3	1.4	2.1	4.9	<u>1.1</u>	3.9	8.4	2.4	2.6	6.2	<u>0.9</u>
	0.5	5.6	6.5	3.1	1.8	1.9	6.1	0.8	1.4	3.7	0.6	2.6	6.9	1.5	1.7	5.2	<u>0.5</u>
	0.1	3.6	4.7	1.6	0.7	0.6	4.0	0.2	0.6	2.0	<u>0.1</u>	1.2	5.0	0.5	0.7	3.2	<u>0.1</u>
1000	20.0	23.6	24.6	24.7	23.6	23.5	21.9	21.6	22.5	23.2	<u>20.5</u>	23.8	22.0	21.9	23.2	23.2	21.3
	10.0	14.6	15.7	14.4	12.9	13.2	13.9	10.8	12.2	13.1	<u>10.2</u>	13.6	14.5	11.3	12.8	13.7	10.5
	5.0	9.5	10.3	9.0	7.4	7.8	9.8	5.4	6.8	7.8	<u>5.2</u>	8.3	10.5	6.1	7.5	8.6	5.3
	2.5	6.8	7.7	5.5	4.1	4.4	7.5	<u>2.5</u>	4.0	5.2	2.9	4.9	8.1	3.2	4.3	6.1	2.7
	1.0	4.6	5.5	3.1	2.0	2.1	5.4	<u>1.0</u>	2.3	2.9	1.3	2.5	6.0	1.5	2.1	3.9	1.2
	0.5	3.5	4.3	2.1	1.2	1.2	4.1	0.6	1.4	1.9	0.8	1.6	4.7	0.9	1.4	2.8	<u>0.5</u>
	0.1	1.9	2.5	1.0	0.4	0.4	2.2	0.2	0.6	0.8	0.2	0.6	2.7	0.4	0.5	1.4	<u>0.1</u>

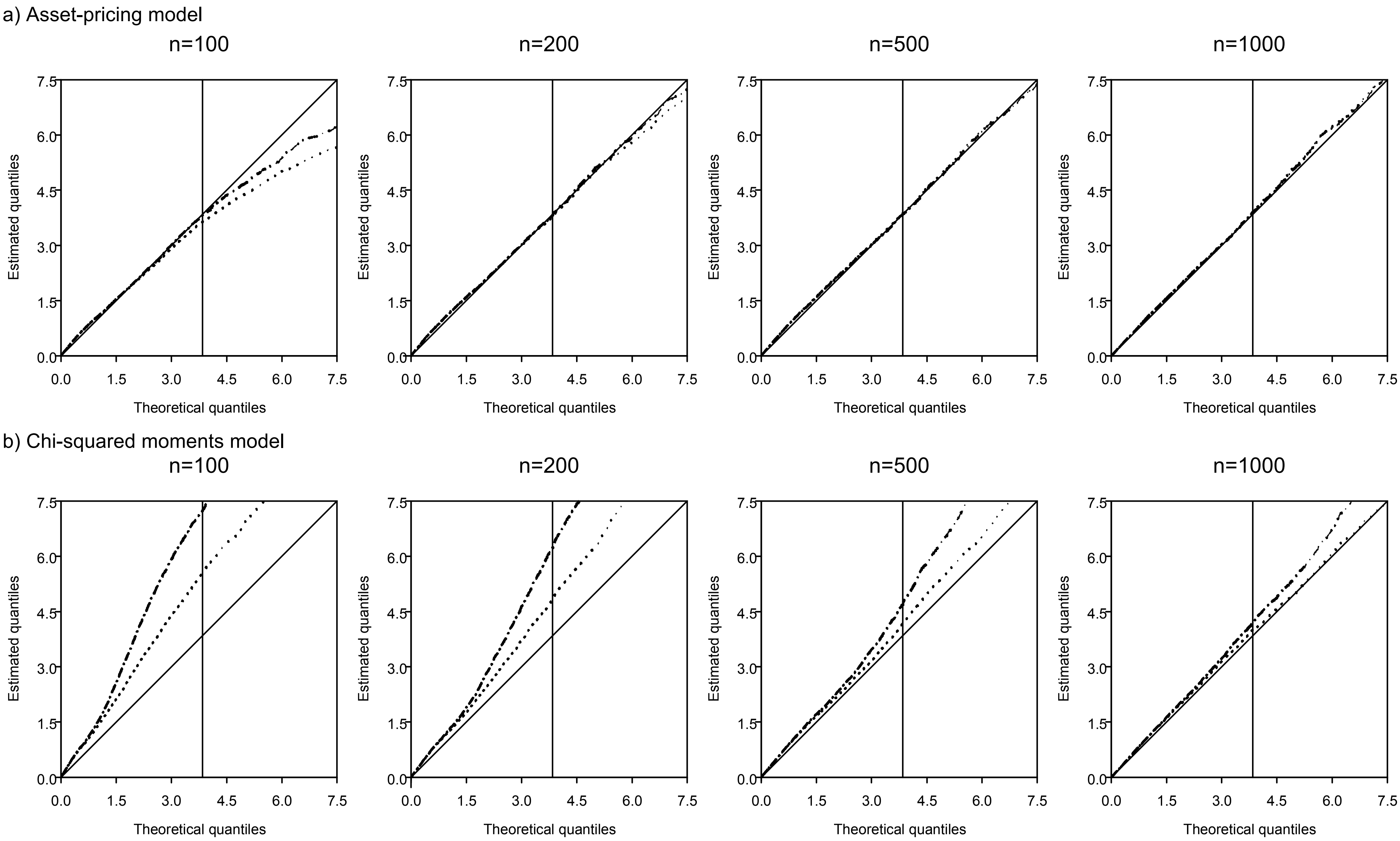
Note: Empirical size closest to nominal size is underlined.

Figure 1: QQ-plots of Palt tests of overidentifying moment conditions (s=8; 10 000 replications)



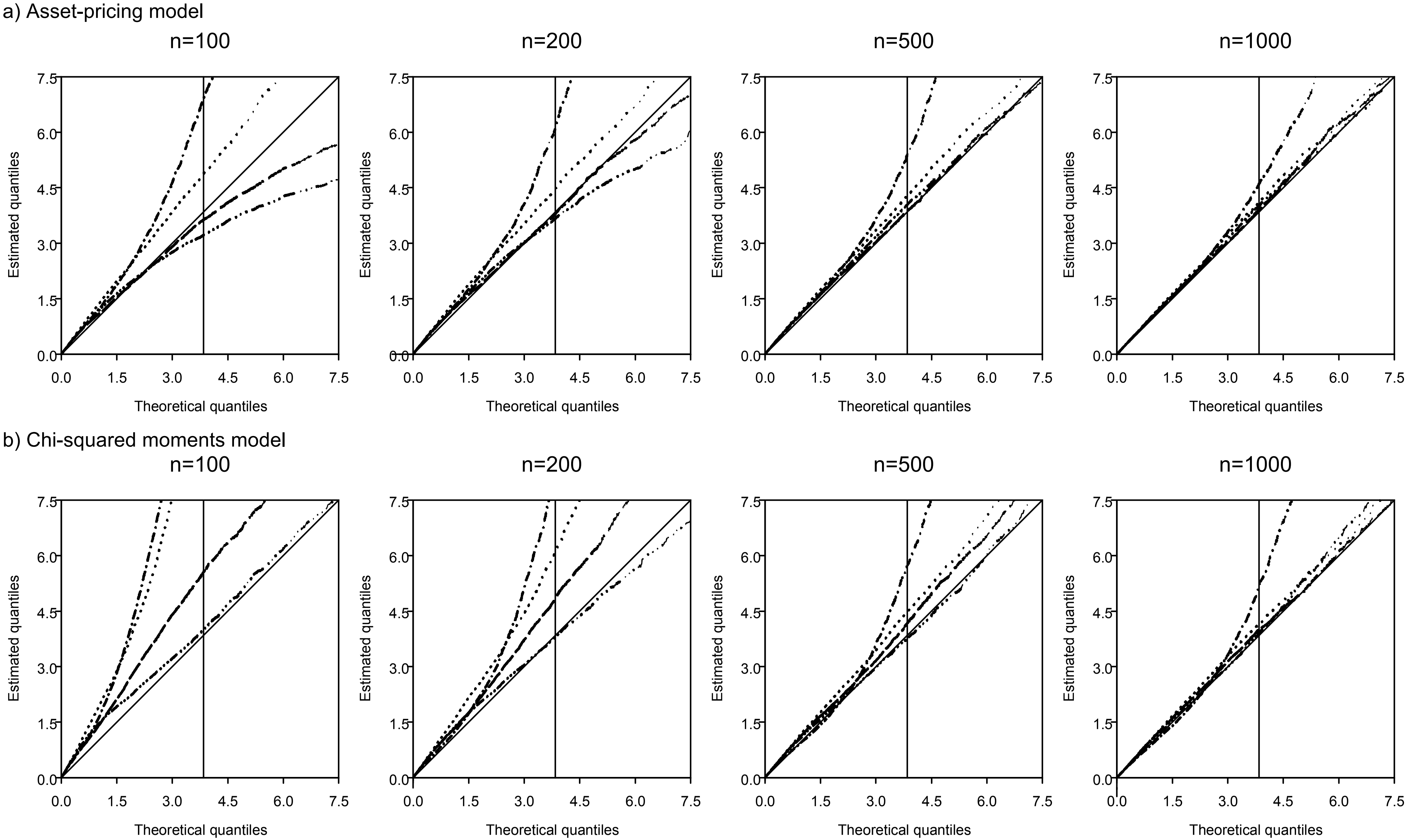
Notes: et(n) (dotted line), et(s) (dot-dashed line), et(r) (dashed line), el(n) (three-dot-dashed line), el(s) (two-dashed line), el(r) (frequent-dotted line).

Figure 2: QQ-plots of Palt-et(r) tests of overidentifying moment conditions (10 000 replications)



Notes: s=8 (dotted line), s=16 (dot-dashed line).

Figure 3: QQ-plots of robust forms of LM and Palt (s=8) tests of overidentifying moment conditions (10 000 replications)



Notes: LMet(r) (dotted line), LMel(r) (dot-dashed line), Palt-et(r) (dashed line), Palt-el(r) (three-dot-dashed line).