UvA ECONOMETRICS

Discussion Paper: 2008/02

# GEL methods for non-smooth moment indicators

Paulo M.D.C. Parente and Richard J. Smith

**www.feb.uva.nl/ke/UvA-Econometrics**

**Amsterdam School of Economics**
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

# GEL Methods for Non-Smooth Moment Indicators*

Paulo M.D.C. Parente
Department of Economics
University of Warwick

Richard J. Smith
cemmap
U.C.L and I.F.S.
and
Faculty of Economics
University of Cambridge

First Draft: November 2004
This Draft: June 2008

**Abstract**

This paper considers the first order large sample properties of the GEL class of estimators for models specified by non-smooth indicators. The GEL class includes a number of estimators recently introduced as alternatives to the efficient GMM estimator which may suffer from substantial biases in finite samples. These include EL, ET and the CUE. This paper also establishes the validity of tests suggested in the smooth moment indicators case for over-identifying restrictions and specification. In particular, a number of these tests avoid the necessity of providing an estimator for the Jacobian matrix which may be problematic for the sample sizes typically encountered in practice.

**JEL Classification:** C13, C30

**Keywords:** Non-Smooth Moment Indicators, Overidentifying Moments, Parametric Restrictions, Additional Moment Restrictions.

*We are grateful for the constructive comments and criticisms of a Joint Editor and two referees on an earlier version of the paper. This research has been presented at the University of Warwick, the Netherlands Econometric Study Group Conference 2005, Amsterdam, the 2005 ESRC Econometric Study Group Annual Conference, Bristol, and the 2005 World Congress of the Econometric Society, U.C.L., London. We are thankful to R. Dupleich Ulloa, J.A.F. Machado, M.B. Stewart and to participants at these meetings for useful comments. The Fundação para a Ciência e Tecnologia and a 2002 Leverhulme Major Research Fellowship provided financial support for Parente and Smith respectively. Any remaining errors are solely the responsibility of the authors.

# 1 Introduction

In his monograph on empirical likelihood (EL), [Owen, 2001], Owen proffered a list of challenges that EL had yet to confront, one of which concerned the lack of theoretical underpinnings for EL when applied to non-smooth estimating equations; see Owen (2001, section 10.6). This particular issue is addressed through consideration of the generalised empirical likelihood (GEL) class of estimators [Smith, 1997, 2001]. GEL methods encompass a large number of estimators encountered in the literature. Indeed, EL [Qin and Lawless, 1994, Imbens, 1997], exponential tilting (ET) [Kitamura and Stutzer, 1997, Imbens *et al.*, 1998] and the continuous updating estimator (CUE) [Hansen et al., 1996], see Newey and Smith (2004), (NS henceforth), belong to the GEL class. Furthermore, NS demonstrated that if the moment indicators are continuously differentiable the minimum discrepancy (MD) estimators of Corcoran (1998) have a dual GEL version when the discrepancy function belongs to the Cressie and Read (1984) family.

NS proved consistency for GEL without requiring differentiable moment indicators although their proof of asymptotic normality uses their differentiability. This assumption does not hold in a number of important models such as the quantile regression (QR) model [Koenker and Bassett, 1978], the censored QR model [Powell, 1984, 1986] and asymmetric least squares [Newey and Powell, 1987]. Differentiability is also required for the NS GMM and GEL asymptotic bias expressions, which account for the large GMM and smaller GEL biases obtained in some simulation experiments; see, e.g., Newey, Ramalho and Smith (2005) and Ramalho (2001).

The main objective of this article is to provide a unified first order asymptotic theory for the GEL class of estimators when the moment indicators are not differentiable at the true value of the parameter.[1] The assumptions required are precisely those given by Newey and McFadden (1994) for the two-step GMM (2S-GMM) estimator, i.e., the regularity conditions for GEL asymptotic normality are no more stringent than those for

[1]An alternative method would smooth the moment indicators similarly to Horowitz (1992, 1998). Whang (2003) adapts this procedure for EL but is, however, rather restrictive, only examining the coverage accuracy of EL confidence regions rather than bias.

2S-GMM. A number of disparate results are already available in the literature. Asymptotic normality for CUE follows immediately from the results of Pakes and Pollard (1989). ET asymptotic normality was proved by Christoffersen et al. (1999). Zhang and Gijbels (2004) demonstrated the same result for a version of EL for conditional moment indicators with bounded regressor support; Otsu (2007) showed that for QR this assumption can be dropped and proposed statistics for testing parametric restrictions and specification. Chernozhukov and Hong (2003) proved an asymptotic equivalence between the 2S-GMM estimator and GEL objective functions evaluated at the GEL optima.[2] Apart from Pakes and Pollard (1989), the requisite assumptions used in these papers differ from those for 2S-GMM asymptotic normality.[3]

The validity of a battery of GEL-based inference procedures proposed for the smooth case by Smith (1997, 2000, 2001) and Ramalho and Smith (2004) is also proven, including tests of overidentifying moment conditions, parametric restrictions and additional moment conditions. An important advantage of some of these statistics is that estimation of the asymptotic variance matrix of the GEL estimator is not required; although several estimation methods have been proposed, none appears to be particularly reliable in practice for the non-smooth case.[4] GEL methods, of course, share this particular feature with those based on efficient GMM. Indeed, Newey and West (1987) proposed GMM likelihood ratio-like tests when the moment indicators are smooth. These results remain valid for the non-smooth case, yielding an easily implementable test under heterogeneity for parametric restrictions in the standard and, thus, just-identified QR framework. To the best of our knowledge this statistic is new in the QR literature; see, e.g., the recent monograph Koenker (2005). We also note a close relationship with the QR statistics of Koenker and Basset (1982) and Weiss (1991).

[2]Although GEL asymptotic normality was not shown, this result could provide a basis for a proof; see section 2.2.

[3]In independent work, Kemp (2005) extends the GEL asymptotic normality result to weakly dependent data.

[4]Buchinsky (1995) compares estimators based on bootstrap and kernel methods for QR. However, even though the bootstrap performs well, as yet there is no formal proof that the bootstrap estimator is consistent in this framework.

Kitamura (2001) is the exception to a general absence of theoretical results available for discrimination between asymptotically equivalent tests in a moment condition setting. Kitamura (2001) proves the large deviation optimality of EL-based tests of over-identifying moment conditions under assumptions also appropriate for the non-smooth moment indicator context studied here. Parametric restrictions test performance together with estimator bias are therefore examined in a set of Monte Carlo experiments. The results are rather limited since optimisation of the GMM and GEL objective functions requires procedures which are extremely time consuming. Consequently, we are unable to rank 2S-GMM and GEL estimators unequivocally in terms of bias. These estimators are, however, generally less biased than GMM with an identity matrix as metric. Likelihood ratio statistics seem most efficacious among all test statistics examined.[5]

The article is organised as follows. Section 2 describes GMM and GEL and presents the main results on estimation. Tests for overidentifying conditions, parametric restrictions and additional moment conditions are considered in sections 3 and 4. Section 5 presents simulation evidence on the efficacy of GEL. Finally, section 6 concludes. Proofs of results in the text are given in the Appendix.

# 2 GMM and GEL Estimation

Suppose the following moment conditions hold

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

where $E[\cdot]$ denotes expectation taken with respect to the distribution of $z$, $\beta_0$ is an unknown $p$-vector and $g(z,\beta)$ a known $m$-vector of functions with $m\geq p$.

Let $z_i$, $(i=1,...,n)$ denote a random sample of data observations drawn from the distribution of $z$. Also let $\Omega=E\left[g\left(z,\beta_0\right)g\left(z,\beta_0\right)'\right]$ and $\hat{\Omega}\left(\beta\right)=\sum_{i=1}^{n}g_i\left(\beta\right)g_i\left(\beta\right)'/n$.

[5] Note that while in the just-identified case bootstrap inference can be a reliable alternative to the procedures discussed here, for over-identified models the implementation of such resampling methods may not be practical for non-smooth moment indicators since it becomes necessary to solve for GEL in each bootstrap sample.

## 2.1 GMM and GEL

Given $\beta$, the sample analog of the population expectation $E\left[g\left(z,\beta\right)\right]$ is $\hat{g}\left(\beta\right)=\sum_{i=1}^{n}g_i(\beta)/n$ where $g_i\left(\beta\right)=g\left(z_i,\beta\right)$, $(i=1,...,n)$. The moment condition (2.1) and uniform convergence of $\hat{g}\left(\beta\right)$ to $E\left[g\left(z,\beta\right)\right]$ under suitable regularity conditions suggests estimation of $\beta_0$ by the GMM estimator $\hat{\beta}$ obtained by minimisation of the GMM criterion; *viz.*

$$\hat{\beta}_{GMM}=\arg\min_{\beta\in\mathcal{B}}\hat{g}(\beta)'\hat{W}\hat{g}(\beta) \tag{2.2}$$

where $\hat{W}$ is p.s.d. and $\mathcal{B}$ denotes the parameter space. With smooth indicators, Hansen (1982) showed that if $\hat{W}\xrightarrow{p}W$, $W$ is p.d., then, under some additional regularity conditions, the GMM estimator is consistent and asymptotically normally distributed. Moreover, among the class of GMM estimators defined by (2.2) the efficient GMM estimator sets $W=\Omega^{-1}$. The efficient 2S-GMM estimator utilises an initial consistent GMM estimate $\tilde{\beta}_{GMM}$ for $\beta_0$, obtained, e.g., by setting $\hat{W}=I_m$, and replaces $\hat{W}$ in (2.2) by $\hat{\Omega}(\tilde{\beta}_{GMM})^{-1}$.

Let

$$\hat{P}_n(\beta,\lambda)=\sum\nolimits_{i=1}^{n}(\rho(\lambda'g_i(\beta))-\rho_0)/n, \tag{2.3}$$

where $\rho\left(\cdot\right)$ is a concave function on its domain, an open interval $\mathcal{V}$ containing zero, $\rho_j\left(v\right)=\partial^j\rho\left(v\right)/\partial\nu^j$ with $\rho_j\left(0\right)=\rho_j$, $(j=0,1,...)$, and normalised without loss of generality by setting $\rho_1=\rho_2=-1$. The GEL estimator is then defined as

$$\hat{\beta}=\arg\min_{\beta\in\mathcal{B}}\sup_{\lambda\in\hat{\Lambda}_n(\beta)}\hat{P}_n\left(\beta,\lambda\right), \tag{2.4}$$

where $\hat{\Lambda}_n\left(\beta\right)=\{\lambda:\lambda'g_i\left(\beta\right)\in\mathcal{V},\ i=1,...,n\}$; see NS and Smith (1997, 2001). EL and ET estimators are obtained with $\rho(v)=\log(1-v)$ and $\mathcal{V}=(-\infty,1)$ [Qin and Lawless (1994), Smith (1997)] and $\rho(\cdot)=-\exp(v)$ [Kitamura and Stutzer, 1997, Imbens et al., 1998, Smith, 1997] whereas CUE $\hat{\beta}_{CUE}=\arg\min_{\beta\in\mathcal{B}}\hat{g}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{g}(\beta)$ [Pakes and Pollard, 1989, Hansen et al., 1996] is a GEL estimator when $\rho(\cdot)$ is quadratic [NS, Theorem 2.1, p.223]. Moreover, MD estimators [Corcoran, 1998] are GEL if the discrepancy function belongs to the Cressie and Read (1984) family [NS, Theorem 2.2, p.224]. NS and Smith

(1997, 2001) show that $\hat{\beta}$ is first order asymptotically equivalent to efficient GMM when the moment indicators are smooth.

For non-smooth moment indicators GEL is no longer required to minimize (2.3) but rather to satisfy

$$\hat{P}_n(\hat{\beta}, \hat{\lambda}) \leq \inf_{\beta \in \mathcal{B}} \sup_{\lambda \in \hat{\Lambda}_n(\beta)} \hat{P}_n(\beta, \lambda) + o_p\left(n^{-\tau}\right)$$

where $\tau$ is non-negative and $\hat{\lambda} = \arg\max_{\lambda \in \hat{\Lambda}_n(\hat{\beta})} \hat{P}_n(\hat{\beta}, \lambda)$.[6] This definition of $\hat{\beta}$ is analogous to that of Pakes and Pollard (1989); see also Newey and McFadden (1994, section 7).

## 2.2 Asymptotic Properties

This sub-section shows that conditions sufficient to ensure consistency and asymptotic normality for 2S-GMM [Newey and MacFadden, 1994, section 7] are precisely those for GEL with non-smooth moment indicators.

NS, Theorem 3.1, p.226, reproduced here for ease of reference, gives GEL consistency without requiring differentiability.[7]

**Assumption 2.1** **(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)** *$E[\sup_{\beta \in \mathcal{B}} \|g(z, \beta)\|^2] < \infty$ ;* **(e)** *$\Omega$ is nonsingular;* **(f)** *$\rho(v)$ is twice continuously differentiable in a neighbourhood of zero.*

Cf. NS, Assumption 1, p.226. Assumption 2.1 (d) relaxes the boundedness condition in NS, i.e., $E\left[\sup_{\beta \in \mathcal{B}} \|g(z, \beta)\|^\alpha\right] < \infty$ for some $\alpha > 2$. Using Lemma 3 of Owen (1990), Guggenberger and Smith (2005, p.673) show $\alpha = 2$ permits the proof of Lemma A1 in NS without further modification.

**Theorem 2.1** *If Assumption 2.1 is satisfied, then $\hat{\beta} \xrightarrow{p} \beta_0$, $\hat{g}(\hat{\beta}) = O_p(n^{-1/2})$ and $\hat{\lambda} = O_p(n^{-1/2})$.*

[6] Theorem 2.1 (consistency) below sets $\tau = 0$ whereas Theorem 2.2 (asymptotic normality) requires $\tau = 1$.

[7] A consistency proof for ET is given in Kitamura and Stutzer (1997) which also does not require moment indicator differentiability.

Previous studies for non-smooth moment conditions considered the M-estimator class. Asymptotic normality for M-estimators defined in terms of non-smooth objective functions has been discussed *inter alia* in Daniels (1961), Huber (1967), Pakes and Pollard (1989), Pollard (1985), Newey and McFadden (1994) and Van der Vaart (1998). All of these authors present different sufficient conditions to ensure asymptotic normality. Asymptotic normality for CUE follows immediately from the results of Pakes and Pollard (1989). For ET asymptotic normality of $\hat{\beta}$ and $\hat{\lambda}$ was proved by Christoffersen et al. (1999) and Zhang and Gijbels (2004) demonstrated the same result for a version of EL with conditional moment indicators.

The proof for the joint asymptotic normality of $\hat{\beta}$ and $\hat{\lambda}$ stated in Theorem 2.2 below follows closely that of Pakes and Pollard (1989, Theorem 3.3, p.1040) for CUE. Let $G = \partial E\left[g\left(z, \beta_0\right)\right] / \partial \beta'$.

**Assumption 2.2** **(a)** *$\beta_0 \in int(\mathcal{B})$;* **(b)** *$g(\beta) = E\left[g(z,\beta)\right]$ is differentiable at $\beta_0$;* **(c)** *$rank(G) = p$;* **(d)** *write $v_n(\beta) = \sqrt{n}\left[\hat{g}(\beta) - g(\beta)\right]$, then for any $\delta_n \to 0$*

$$\sup_{\|\beta-\beta_0\|\leq\delta_n} \frac{\|v_n(\beta) - v_n(\beta_0)\|}{1+\sqrt{n}\,\|\beta-\beta_0\|} \xrightarrow{p} 0.$$

Assumption 2.2 (b) substitutes differentiability of $g(z,\beta)$ at $\beta_0$ by differentiability of its expected value; cf. NS, Assumption 2, p.226. Assumption 2.2 (d) is also used in Pakes and Pollard (1989, Theorem 3.3, p.1040) and by Newey and McFadden (1994, Theorem 7.2, p.2186) for GMM. Note that

$$\sup_{\|\beta-\beta_0\|\leq\delta_n} \frac{\|v_n(\beta) - v_n(\beta_0)\|}{1+\sqrt{n}\,\|\beta-\beta_0\|} \leq \sup_{\|\beta-\beta_0\|\leq\delta_n} \|v_n(\beta) - v_n(\beta_0)\|;$$

the right hand side tends to zero in probability if the sequence $\{v_n(\beta), n \geq 1\}$ is stochastically equicontinuous. Primitive conditions for stochastic equicontinuity in different set-ups arise from empirical process theory; see, e.g., Pollard (1984), Pakes and Pollard (1989) and Andrews (1994).

Let $\Sigma = \left(G'\Omega^{-1}G\right)^{-1}$ and $P = \Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1}$.

**Theorem 2.2** *Let Assumptions 2.1 and 2.2 hold. Then*

$$\sqrt{n}\begin{pmatrix} \hat{\beta}-\beta_0 \\ \hat{\lambda} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \mathtt{diag}\left(\Sigma, P\right)\right).$$

The structure of the proof is as follows. Since a proof of asymptotic normality based on the GEL first order conditions is no longer applicable, the objective function $\hat{P}_n(\beta,\lambda)$ is approximated by the smooth well-behaved (although infeasible) function $\hat{L}_n(\beta,\lambda) = [-G(\beta-\beta_0)]'\lambda - \hat{g}(\beta_0)'\lambda - \frac{1}{2}\lambda'\Omega\lambda$; cf. Pakes and Pollard (1989). Using standard arguments based on the first order conditions of this problem the estimators $\tilde{\beta}$ and $\tilde{\lambda}$ that solve $\min_{\beta\in\mathcal{B}}\sup_{\lambda\in\mathcal{R}^m}\hat{L}_n(\beta,\lambda)$ have the limiting normal distribution of Theorem 2.2; see NS, Theorem 3.2, p.226. Given Assumption 2.2 and using Lemmata A.1-A.3 in the Appendix the GEL estimators $\hat{\beta}$ and $\hat{\lambda}$ are then shown to be asymptotically equivalent to $\tilde{\beta}$ and $\tilde{\lambda}$.

An alternative proof of Theorem 2.2 could be based on Newey and McFadden (1994, Theorem 7.2, p.2186). It is immediately apparent from (2.4) that GEL may be cast as an M-estimation problem by defining $\hat{\lambda}(\beta) = \arg\sup_{\lambda\in\hat{\Lambda}_n(\beta)}\hat{P}_n(\beta,\lambda)$, the existence of which follows from the implicit function theorem as $\rho(\cdot)$ is twice differentiable. See the proof of Theorem 2.2, p.238, in NS; cf. Smith (1997, p.507, 2001, section 2.3). Chernozhukov and Hong (2003) exploit the structure of the first order conditions with respect to $\lambda$, *viz.* $\sum_{i=1}^n \rho_1(\hat{\lambda}(\beta)'g_i(\beta))g_i(\beta) = 0$. Applying the mean value theorem to $\rho_1(\hat{\lambda}(\beta)'g_i(\beta))$, $-\hat{g}(\beta) - \Omega_n\hat{\lambda}(\beta) = 0$ where $\Omega_n(\beta) = -\sum_{i=1}^n \rho_2(\bar{\lambda}(\beta)'g_i(\beta))g_i(\beta)g_i(\beta)'/n$ and $\bar{\lambda}(\beta)$ lies between 0 and $\hat{\lambda}(\beta)$. Since $\Omega_n$ is p.d. w.p.a.1 in a neighbourhood of $\beta_0$,

$$\hat{\lambda}(\beta) = -\Omega_n(\beta)^{-1}\hat{g}(\beta). \tag{2.5}$$

Plugging (2.5) into a second order Taylor expansion of the GEL criterion function $\hat{P}_n(\hat{\beta},\hat{\lambda})$ around $\lambda = 0$, as $\hat{\lambda} \xrightarrow{p} 0$ by Theorem 2.1, Chernozhukov and Hong (2003) showed that under some additional regularity conditions $\hat{P}_n(\hat{\beta},\hat{\lambda}) = \frac{1}{2}\hat{g}(\hat{\beta})'\Omega^{-1}\hat{g}(\hat{\beta}) + o_p(n^{-1})$, i.e., the GEL objective function at $\hat{\beta}$ and $\hat{\lambda}$ is asymptotically equivalent to the efficient GMM criterion function. Application of Newey and McFadden (1994, Theorem 7.2, p.2186) provides the asymptotic distribution of the GEL estimator $\hat{\beta}$. That for $\hat{\lambda}$ would follow directly from (2.5).[8]

The following example shows the usefulness of Theorem 2.2.

[8] We are grateful to a referee for noting that these arguments also hold under our assumptions.

## 2.3 Example: IV Estimation for QR

Suppose that the $\theta$-quantile of $y$ conditional on $w$ is defined by $Q_\theta(y|w) = w'\beta_\theta$ where $Q_\theta(y|w) = \inf\{q : \mathcal{P}\{y \leq q|w\} \geq \theta\}$. The linear QR model is then

$$y = w'\beta_\theta + \xi,$$

where $\mathcal{P}\{\xi \leq 0|w\} = \theta$. Given a random sample $z_i = (y_i, w_i)$, $(i = 1, ..., n)$, Koenker and Basset (1978) showed that the estimator obtained by minimization of the following program leads to a consistent estimator for $\beta_\theta$

$$\sum_{i=1}^n \rho_\theta(y_i - w_i'\beta_\theta), \tag{2.6}$$

where $\rho_\theta(\cdot)$ is the check function $\rho_\theta(\xi) = \xi[\theta - I(\xi < 0)]$ and $I(\cdot)$ denotes an indicator function. This estimator is consistent since $E[sgn_\theta(\xi)w] = 0$, where $sgn_\theta(\xi) = \theta - I(\xi < 0)$ is the $\theta$-weight sign function defined in Fitzenberger (1997).

If $\mathcal{P}\{\xi \leq 0|w\} \neq \theta$, then $E[sgn_\theta(\xi)w] \neq 0$ rendering the above estimator inconsistent. Suppose that the $m$-vector of instruments $x$ is such that $\mathcal{P}\{\xi \leq 0|x\} = \theta$ implying the moment conditions $E[sgn_\theta(\xi)x] = 0$; cf. Chernozhukov and Hong (2003) and Honore and Hu (2004). GMM is then based on the sample analogue

$$\sum_{i=1}^n sgn_\theta(y_i - w_i'\hat{\beta}_\theta)x_i/n = 0. \tag{2.7}$$

Here $\Omega = E[sgn_\theta(\xi)^2 xx'] = \theta(1-\theta)E[xx']$. Newey and McFadden (1994) proved consistency [Theorem 2.6, p.2132] and asymptotic normality [Theorem 7.2, p.2186] for GMM with non-smooth moment indicators.

The following assumption gives sufficient conditions for consistency of GEL based on the moment conditions (2.7).

**Assumption E.1:** *(a)* $\xi$ *is continuously distributed given* $x$ *and* $\mathcal{P}\{\xi \leq 0|x\} = \theta$; **(b)** $\mathcal{B}$ *is compact;* **(c)** $E[\|x\|^2] < \infty$; **(d)** $E[xx']$ *is nonsingular;* **(e)** $\rho(\cdot)$ *is twice continuously differentiable in a neighbourhood of zero.*

**Theorem E.1:** *Under Assumption E.1 the GEL estimator based on the moment condition (2.7) is a consistent estimator for* $\beta_\theta$.

For asymptotic normality the following additional assumption is required.

**Assumption E.2:** *(a)* $\beta_\theta \in int(\mathcal{B})$; *(b) the distribution function of* $\xi$ *conditional on* $x$ *and* $w$ *is differentiable at* $0$ *with derivative* $f_\xi(0|x,w)$; **(c)** $E[f_\xi(0|x,w)xw']$ *is full column rank.*

**Theorem E.2:** *Under Assumptions E.1 and E.2 the limiting distribution of the GEL estimator for* $\beta_\theta$ *based on the moment condition (2.7) is given by*

$$\sqrt{n}(\hat{\beta}-\beta_\theta) \xrightarrow{d} N(0,\theta(1-\theta)\left(E[f_\xi(0|x,w)wx']E[xx']^{-1}E[f_\xi(0|w,x)xw']\right)^{-1}).$$

## 2.4 Asymptotic Variance Matrix Estimation

To apply Theorem 2.2 in practice requires the consistent estimation of $\Omega$ and $G$. Using a uniform WLLN, by Assumption 2.1, the OPG estimator $\hat{\Omega} = \sum_{i=1}^n g_i(\hat{\beta})g_i(\hat{\beta})'/n$ is consistent for $\Omega$.

Estimation of $G$ is more problematic. Pakes and Pollard (1989) proposed an estimator $\hat{G}^{PP}$ based on numerical derivatives of the empirical moment indicators. The $j$th column of $\hat{G}^{PP}$ is given by $\hat{G}_j^{PP} = \varepsilon_n^{-1}\left[\hat{g}(\hat{\beta}+e_j\varepsilon_n)-\hat{g}(\hat{\beta})\right]$, where $e_j$ is the $j$th unit vector and $\varepsilon_n \xrightarrow{p} 0$. Other estimators may be derived for specific cases, e.g., for the IV QR example, Powell's (1984) consistent estimator for $G$ is $\hat{G}^P = \sum_{i=1}^n 1(\left|y_i - w_i'\hat{\beta}_\theta\right| < \hat{c}_n)x_i w_i'/2\hat{c}_n n$ where $\hat{c}_n \xrightarrow{p} 0$ and satisfies the other regularity conditions stated in Powell (1984); see section 5.1. The Monte-Carlo study of Buchinsky (1995) for QR showed that the performance of kernel-based estimators, e.g., $\hat{G}^P$, depends critically on the choice of kernel and bandwidth, the latter problem also shared by the numerical derivative estimator.

An important feature of the GEL framework is that a number of test statistics do not require an estimator of $G$. See sections 3 and 4.

# 3 Overidentifying Moment Conditions

This section focuses on tests to gauge the validity of the moment conditions (2.1). The traditional test statistic proposed by Hansen (1982) has been criticised in the literature due to its poor finite sample properties. Consequently a number of authors, including

Imbens et al. (1998), Kitamura and Stutzer (1997) and Smith (1997, 2000, 2001), have proposed alternative test statistics based on classical principles. Theorem 3.1 shows that these statistics are also valid when the moment conditions are non-smooth.

**Theorem 3.1** *Let $\hat{\Omega} = \Omega + o_p(1)$ and Assumptions 2.1, 2.2 hold. Then the likelihood ratio (LR) statistic*

$$\mathcal{LR} = 2n\hat{P}_n(\hat{\beta}, \hat{\lambda}),$$

*the Lagrange multiplier (LM) statistic*

$$\mathcal{LM} = n\hat{\lambda}'\hat{\Omega}\ \hat{\lambda},$$

*and the score (S) statistic*

$$\mathcal{S} = n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta})$$

*are asymptotically equivalent. In particular,* $\mathcal{LR}$, $\mathcal{LM}$, $\mathcal{S} \xrightarrow{d} \chi^2_{m-p}$.

Under Assumption 2.1, the hypotheses of Kitamura (2001, Theorem 2, pp.1664-5) apply. Therefore, the EL version of $\mathcal{LR}$ defines a $\delta$-optimal test of (2.1). Imposition of a further regularity condition results in this test being asymptotically efficient in the Hoeffding (1965) sense; see Kitamura (2001, Corollary 1, p.1665).

Implied probabilities may be defined in the GEL framework, *viz.*

$$\hat{p}_i = \frac{\rho_1(\hat{\lambda}'g_i(\hat{\beta}))}{\sum_{j=1}^n \rho_1(\hat{\lambda}'g_j(\hat{\beta}))}, (i = 1, ..., n).$$

See Brown and Newey (2002) and NS, p.223. Ramalho and Smith (2004) proposed Pearson-type test statistics for overidentifying moment conditions based on these implied probabilities, *viz.*

$$\mathcal{P}_n^a = \sum\nolimits_{i=1}^n (n\hat{p}_i - 1)^2$$

and

$$\mathcal{P}_n^b = \sum\nolimits_{i=1}^n \frac{(n\hat{p}_i - 1)^2}{n\hat{p}_i}.$$

Theorem 3.2 shows that $\mathcal{P}_n^a$ and $\mathcal{P}_n^b$ are asymptotically equivalent to $\mathcal{LR}$, $\mathcal{LM}$ and $\mathcal{S}$.

**Theorem 3.2** *If $\hat{\Omega} = \Omega + o_p(1)$ and Assumptions 2.1 and 2.2 are satisfied, then the Pearson-type statistics $\mathcal{P}_n^a$ and $\mathcal{P}_n^b$ are asymptotically equivalent to $\mathcal{LR}$, $\mathcal{LM}$ and $\mathcal{S}$. Therefore, $\mathcal{P}_n^a$, $\mathcal{P}_n^b \xrightarrow{d} \chi^2_{m-p}$.*

Note that none of these statistics requires the estimation of $G$. Cf. section 2.4.

# 4 Specification Tests

This section is concerned with the validity, or otherwise, of additional moment restrictions together with parametric restrictions on $\beta_0$; *viz.*

$$E[q(z, \beta_0)] = 0, r(\beta_0) = 0 \tag{4.1}$$

where $q(z, \beta)$ is a known $s$-vector of moment indicators and $r(\beta)$ a $r$-vector of constraints, $r \leq p$. The results given below are easily specialised to the pure additional moments or parametric restrictions cases.

Model specification tests are typically based on moment conditions of the type $E[q(z, \beta_0)] = 0$, e.g., tests of functional form, heteroskedasticity and endogeneity. Such tests were introduced by Newey (1985) and Tauchen (1985) for M-estimators, the latter paper being particularly notable since non-differentiability of $q(z, \beta)$ at $\beta_0$ is permitted. Newey and West (1987) proposed tests for parametric restrictions of the form $r(\beta_0) = 0$ based on efficient GMM estimation in the smooth moment indicator context. A number of authors have considered GEL-based tests of additional moment and parametric restrictions based on GEL and its variants. See, e.g., Kitamura and Stutzer (1997) and Smith (1997, 2000, 2001).

Let $h(z, \beta) = (g(z, \beta)', q(z, \beta)')'$, $q_i(\beta) = q(z_i, \beta)$ and $h_i(\beta) = h(z_i, \beta)$, $(i = 1, ..., n)$, $\hat{q}(\beta) = \sum_{i=1}^n q_i(\beta)/n$ and $\hat{h}(\beta) = \sum_{i=1}^n h_i(\beta)/n$. Also let $Q = \partial E[q(z, \beta_0)]/\partial\beta'$, $H = (G', Q')'$, $\Xi(\beta) = E[h(z, \beta)h(z, \beta)']$, $\Xi = \Xi(\beta_0)$, $R(\beta) = \partial r(\beta)/\partial\beta'$ and $R = R(\beta_0)$. Define the restricted parameter space $\mathcal{B}^r = \{\beta \in \mathcal{B} : r(\beta) = 0\}$.

The following additional assumption is required to establish the distribution of test statistics under (4.1) and (2.1).

**Assumption 4.1** **(a)** *$\beta_0$ is the unique solution of $E[h(z,\beta)] = 0$, $r(\beta) = 0$;* **(b)** *$q(z,\beta)$ is continuous at each $\beta \in \mathcal{B}$ with probability one, $E[\sup_{\beta\in\mathcal{B}} \|q(z,\beta)\|^2] < \infty$ ;* **(c)** *$q(\beta) = E[q(z,\beta)]$ is differentiable at $\beta_0$, $r(\beta)$ is twice continuously differentiable on $\mathcal{B}$;* **(d)** *$rank(R) = r$;* **(e)** *$\Xi$ is nonsingular;* **(f)** *write $w_n(\beta) = \sqrt{n}(\hat{q}(\beta) - q(\beta))$, then for any $\delta_n \to 0$,*

$$\sup_{\|\beta-\beta_0\|\leq\delta_n} \frac{\|w_n(\beta) - w_n(\beta_0)\|}{1+\sqrt{n}\|\beta-\beta_0\|} \xrightarrow{p} 0.$$

## 4.1 GMM Tests

Of particular importance is a LR-type test that does not require the estimation of the Jacobian matrix $H$. Although the main focus of this paper concerns GEL methods, Theorem 4.1 below states that the standard limiting chi-square distributional result for the LR-type statistic based on an efficient GMM estimator holds even when the moment conditions are not smooth. This result is presented as it is thought it may be of independent interest.

Let

$$\hat{Q}_n(\beta) = \hat{g}(\beta)' \hat{\Omega}^{-1} \hat{g}(\beta), \hat{Q}_n^r(\beta) = \hat{h}(\beta)'\hat{\Xi}^{-1}\hat{h}(\beta).$$

Also let $\hat{\beta}_{GMM}$ and $\hat{\beta}_{GMM}^r$ denote the efficient unrestricted and restricted GMM estimators respectively obtained from minimization of $\hat{Q}_n(\beta)$ over $\mathcal{B}$ and $\hat{Q}_n^r(\beta)$ over the restricted parameter space $\mathcal{B}^r$.

**Theorem 4.1** *Suppose Assumptions 2.1, 2.2 and 4.1 hold and $\hat{\Xi} = \Xi + o_p(1)$. Then the GMM LR statistic*

$$\mathcal{LR}^{GMM} = n(\hat{Q}_n^r(\hat{\beta}_{GMM}^r) - \hat{Q}_n(\hat{\beta}_{GMM})) \xrightarrow{d} \chi^2_{r+s}.$$

Notice that in the exactly identified case $\hat{Q}_n(\hat{\beta}_{GMM}) = o_p(1)$ and therefore the LR statistic takes the more familiar OPG form of the LM test discussed, e.g., in Davidson and MacKinnon (1983). The implementation of this test is then straightforward, the test statistic being $n - SSR$, where $SSR$ is the sum of the squared residuals obtained from a regression of a vector of ones on $h_i(\hat{\beta}_{GMM}^r)$, $(i = 1, ..., n)$.

## 4.2 Example (cont.)

Confining attention to parametric restrictions $r(\beta_0) = 0$ only, the above result has a rather interesting consequence for QR in the absence of endogeneity. Since the model is now exactly identified, the QR estimator is asymptotically equivalent to efficient GMM with the LR statistic given by

$$\mathcal{LR}^{GMM} = n\hat{Q}_n(\hat{\beta}^r_{GMM})$$

where $\hat{\beta}^r_{GMM}$ is the restricted efficient GMM estimator. Defining $\hat{\Omega} = \theta\,(1-\theta)\sum_{i=1}^n w_i w_i'/n$,

$$\mathcal{LR}^{GMM} = \sum\nolimits_{i=1}^n w_i' sgn_\theta(y_i - w_i'\hat{\beta}^r_{GMM})\hat{\Omega}^{-1}\sum\nolimits_{i=1}^n w_i sgn_\theta(y_i - w_i'\hat{\beta}^r_{GMM})$$

has a limiting chi-square distribution with $r$ degrees of freedom. Although this statistic does not depend directly on $G$, it does require efficient estimation under the null hypothesis. Koenker and Bassett (1982) proposed this statistic (as their $LM$ statistic) for the case of linear restrictions and assuming stochastic independence between the regressors and the error term. However, $\mathcal{LR}^{GMM}$ is still valid if this latter assumption does not hold; see Assumptions E.1, E.2 and 4.1.

It is instructive to compare asymptotically equivalent quadratic forms for the GMM LR-type statistic and the standard QR objective function LR-type statistic. For the GMM LR-type statistic,

$$\mathcal{LR}^{GMM} = n\hat{g}\,(\beta_0)'\,G^{-1}R'\left(RG^{-1}\Omega G^{-1}R'\right)^{-1}RG^{-1}\hat{g}\,(\beta_0) + o_p\,(1)\,;$$

see the Appendix. The QR objective function LR-type statistic is given by

$$\mathcal{LR}^{QR} = 2(\sum\nolimits_{i=1}^n \rho_\theta(y_i - w_i'\hat{\beta}_{qr}) - \sum\nolimits_{i=1}^n \rho_\theta(y_i - w_i'\hat{\beta}^r_{qr})),$$

where $\hat{\beta}_{qr}$ and $\hat{\beta}^r_{qr}$ denote the QR and the restricted QR estimators respectively. Using similar arguments to those of Koenker and Basset (1982),

$$\mathcal{LR}^{QR} = n\hat{g}\,(\theta_0)'\,G^{-1}R'\left(RG^{-1}R'\right)^{-1}RG^{-1}\hat{g}\,(\theta_0) + o_p\,(1)\,.$$

Consequently, while $\mathcal{LR}^{GMM}$ converges in distribution to a chi-square with $r$ degrees of freedom, $\mathcal{LR}^{QR}$ converges in distribution to $\sum_{i=1}^{r} \lambda_i Z_i^2$ where $Z_i$, $(i = 1, ..., r)$, are independent $N(0, 1)$ and $\lambda_i$, $(i = 1, ..., r)$, are the eigenvalues of $(RG^{-1}R')^{-1} RG^{-1}\Omega G^{-1}R'$; see Johnson and Kotz (1972, ch. 29). Thus, in general, $\mathcal{LR}^{GMM}$ and $\mathcal{LR}^{QR}$ are not asymptotically equivalent and the asymptotic distribution of $\mathcal{LR}^{QR}$ is non-standard. In the absence of heterogeneity in the density function at the origin, i.e., $f_\xi(0|w) = f_\varepsilon(0)$, then $G = f_\xi(0)\,\Omega/\theta(1-\theta)$. In this case, therefore, $\mathcal{LR}^{QR}$ can be adjusted to provide a statistic that has a limiting chi-square distribution with $r$ degrees of freedom; *viz.*

$$\mathcal{LR}_a^{QR} = \frac{f_\xi(0)}{\theta(1-\theta)}\mathcal{LR}^{QR}$$

which corresponds to the LR statistic proposed by Koenker and Bassett (1982). Weiss (1991) proposes several tests for testing parametric hypotheses in the case when the data is heterogeneous, but does not mention $\mathcal{LR}^{GMM}$. In fact, the statistic $\mathcal{LR}^{GMM}$ and the associated limiting distributional result seem to be new in the QR literature. Note, however, that in the QR framework the well-known rank test statistic, see, e.g., Koenker (1997), does not require the estimation of the asymptotic covariance matrix either. Application of this test though requires solving the dual problem of (2.6) which is not implemented in most of the econometric software, `S-Plus` being a notable exception.

## 4.3 GEL Tests

The above property of the (efficient) GMM criterion function is also shared by the GEL criterion function. Consider the restricted estimator

$$\hat{P}_n^r(\hat{\beta}^r, \hat{\lambda}^r) \le \inf_{\beta\in\mathcal{B}^r} \sup_{\eta\in\hat{\Lambda}_n(\beta)} \hat{P}_n^r(\beta,\eta) + o_p(n^{-1}) \tag{4.2}$$

where the extended GEL criterion $\hat{P}_n^r(\beta,\eta) = \sum_{i=1}^{n}(\rho(\eta' h_i(\beta) - \rho_0))/n$, $\eta = (\lambda', \psi')'$ is a $(m+s)$-vector of auxiliary parameters and $\hat{\Delta}_n(\beta) = \{\eta : \eta' h_i(\beta) \in \mathcal{V},\ i = 1, ..., n\}$. The auxiliary parameter estimator $\hat{\eta}^r = \hat{\eta}^r(\hat{\beta}^r)$ where $\hat{\eta}^r(\beta) = \arg\max_{\eta\in\hat{\Delta}_n(\beta)} \hat{P}_n^r(\beta,\eta)$. Define $\hat{\eta} = (\hat{\lambda}', 0')'$. Let $\mu$ denote the vector of Lagrange multipliers associated with the constraint $r(\beta) = 0$.

Define the $(p+m+s+r)\times(s+r)$ selection matrix $S_{\psi,\mu}$ such that $S'_{\psi,\mu}(\beta',\eta',\mu')' = (\psi',\mu')'$. Additionally, define the matrix

$$\Psi = \begin{pmatrix} 0 & H' & -R' \\ H & \Xi & 0 \\ -R & 0 & 0 \end{pmatrix}.$$

Given $\hat{\beta}$ or $\hat{\beta}^r$, a consistent estimator $\hat{\Psi}^r$ for $\Psi$ is easily constructed from estimators for $\Xi$, $H$ using similar approaches to those described above to estimate $\Omega$, $G$ with $R$ estimated by $R(\hat{\beta})$ or $R(\hat{\beta}^r)$.

Theorem 4.2 shows that the tests proposed in Smith (1997, 2000, 2001) remain valid for testing (4.1) when the moment indicators are no longer smooth.

**Theorem 4.2** *Let $\hat{\Psi}^r = \Psi + o_p(1)$. If Assumptions 2.1, 2.2 and 4.1 hold, then the likelihood ratio statistic*

$$\mathcal{LR}^r = 2n(\hat{P}^r_n(\hat{\beta}^r,\hat{\eta}^r) - \hat{P}_n(\hat{\beta},\hat{\lambda})),$$

*the Lagrange multiplier statistic*

$$\mathcal{LM}^r = n(\hat{\eta}^r - \hat{\eta})'\hat{\Xi}^r(\hat{\eta}^r - \hat{\eta})$$

*the Wald statistic*

$$\mathcal{W}^r = n(\hat{\psi}^{r\prime},\hat{\mu}^{r\prime})(S'_{\psi,\mu}(\hat{\Psi}^r)^{-1}S_{\psi,\mu})^{-1}(\hat{\psi}^{r\prime},\hat{\mu}^{r\prime})',$$

*and the score statistic*

$$\mathcal{S}^r = n\sum_{i=1}^n \hat{p}_i \begin{pmatrix} q_i(\hat{\beta}) \\ r(\hat{\beta}) \end{pmatrix}' S'_{\psi,\mu}(\hat{\Psi}^r)^{-1}S_{\psi,\mu}\sum_{i=1}^n \hat{p}_i \begin{pmatrix} q_i(\hat{\beta}) \\ r(\hat{\beta}) \end{pmatrix}$$

*are asymptotically equivalent to $\mathcal{LR}^{GMM}$. Therefore, $\mathcal{LR}^r$, $\mathcal{LM}^r$, $\mathcal{W}^r$, $\mathcal{S}^r \xrightarrow{d} \chi^2_{r+s}$.*

The implied probabilities associated with the constrained model are

$$\hat{p}^r_i = \frac{\rho_1(\hat{\eta}^{r\prime}h_i(\hat{\beta}^r))}{\sum_{j=1}^n \rho_1(\hat{\eta}^{r\prime}h_j(\hat{\beta}^r))}, (i=1,...,n).$$

Pearson-type tests for parametric restrictions based on a contrast of constrained and unconstrained implied probabilities were introduced by Ramalho and Smith (2004). Theorem 4.3 demonstrates their validity in the non-smooth moment set-up.

**Theorem 4.3** *If Assumptions 2.1, 2.2 and 4.1 hold, the Pearson-type statistics*

$$\mathcal{P}_r^a = \sum\nolimits_{i=1}^{n} \frac{(n\hat{p}_i^r - n\hat{p}_i)^2}{n\hat{p}_i},$$

$$\mathcal{P}_r^b = \sum\nolimits_{i=1}^{n} \frac{(n\hat{p}_i^r - n\hat{p}_i)^2}{n\hat{p}_i^r},$$

$$\mathcal{P}_r^c = \sum\nolimits_{i=1}^{n} (n\hat{p}_i^r - n\hat{p}_i)^2$$

*are asymptotically equivalent to* $\mathcal{LR}^{GMM}$*. Therefore,* $\mathcal{P}_r^a$*,* $\mathcal{P}_r^b$*,* $\mathcal{P}_r^c \xrightarrow{d} \chi^2_{r+s}$*.*

None of the LR-, LM- and the Pearson-type GEL statistics require an estimator of $H$. Of course, the difficulty of finding a reliable estimator of $H$, required, for example, in the definition of the Wald- and score-type statistics, is replaced by that of solving an additional complicated optimization problem, possibly with multiple local optima, see Andrews (1997), Chernozhukov and Hong (2003) and Whang (2003). Nevertheless, well-known algorithms, e.g., the genetic algorithm, see Dorsey and Mayer (1995), can deal with multiple local optima. Moreover, if efficient unconstrained and constrained estimators for $\beta_0$, $\hat{\beta}$ and $\hat{\beta}^r$, are available, computation of these statistics only requires the solution to the optimization problems $\hat{\lambda} = \arg\max_{\lambda \in \hat{\Lambda}_n(\hat{\beta})} \hat{P}_n(\hat{\beta}, \lambda)$ and $\hat{\eta}^r = \arg\max_{\eta \in \hat{\Delta}_n(\hat{\beta}^r)} \hat{P}_n^r(\hat{\beta}^r, \eta)$. E.g., with purely parametric restrictions, the QR estimator of Koenker and Bassett (1978) is asymptotically equivalent to the GEL estimator and is straightforward to compute in both unconstrained and constrained scenarios. Moreover, if the former case is just-identified, $\hat{\lambda} = 0$. Consequently, $\hat{P}_n(\hat{\beta}, \hat{\lambda}) = 0$ and the likelihood ratio statistic reduces to the simple expression $\mathcal{LR}^r = 2n\hat{P}_n(\hat{\beta}^r, \hat{\lambda}^r)$.

## 4.4 Example (cont.)

Note that $\mathcal{S}^r$ adapted for purely parametric restrictions has exactly the same form as the LM statistic proposed by Weiss (1991) although they are evaluated at different estimators, $\mathcal{S}^r$ at the restricted GEL estimator and Weiss' (1991) LM statistic at the restricted QR estimator. These estimators, however, are in general not asymptotically equivalent although $\mathcal{S}^r$ and Weiss' (1991) LM statistic are.[9] To gain some insight into this apparently puzzling result consider asymptotic representations of the standardized first order conditions for the QR estimator evaluated at both estimators.

For the restricted QR estimator, stochastic equicontinuity condition and differentiability of $g(\beta)$ imply that

$$n^{1/2}\hat{g}(\hat{\beta}^r_{qr}) = R'(RG^{-1}R')^{-1}RG^{-1}n^{1/2}\hat{g}(\beta_0) + o_p(1).$$

Hence

$$RG^{-1}n^{1/2}\hat{g}(\hat{\beta}^r_{qr}) = RG^{-1}n^{1/2}\hat{g}(\beta_0) + o_p(1).$$

Likewise, for the restricted GEL estimator,

$$n^{1/2}\hat{g}(\hat{\beta}^r) = \Omega G^{-1}R'\left(RG^{-1}\Omega G^{-1}R'\right)^{-1}RG^{-1}n^{1/2}\hat{g}(\beta_0) + o_p(1)$$

and

$$RG^{-1}n^{1/2}\hat{g}(\hat{\beta}^r) = RG^{-1}n^{1/2}\hat{g}(\beta_0) + o_p(1).$$

Therefore

$$RG^{-1}n^{1/2}\hat{g}(\hat{\beta}^r) = RG^{-1}n^{1/2}\hat{g}(\hat{\beta}^r_{qr}) + o_p(1).$$

Even though $\hat{\beta}^r_{qr}$ and $\hat{\beta}^r$ are not asymptotically equivalent, the standardized first order conditions of the QR estimator evaluated at $\hat{\beta}^r_{qr}$ and $\hat{\beta}^r$ are asymptotically equivalent when premultiplied by $RG^{-1}$.

## 4.5 Confidence Regions

To compute a confidence region for a sub-vector $\beta^1_0$, say, of $\beta_0$, consisting of $p_1$ elements, the results of Theorem 4.2 indicate that a $1-\alpha$ level confidence region using the LR-type

[9] The restricted QR estimator has the linear representation

$$n^{1/2}(\hat{\beta}^r_{qr} - \beta_0) = -(G^{-1} - G^{-1}R'(RG^{-1}R')^{-1}RG^{-1})n^{1/2}\hat{g}(\beta_0) + o_p(1)$$

where $g_i(\beta_0) = w_i sgn_\theta(y_i - w_i'\beta_0)$, $(i = 1, ..., n)$. The restricted GMM estimator has the representation

$$n^{1/2}(\hat{\beta}^r_{GMM} - \beta_0) = -(G^{-1} - G^{-1}\Omega G^{-1}R'(RG^{-1}\Omega G^{-1}R')^{-1}RG^{-1})n^{1/2}\hat{g}(\beta_0) + o_p(1)$$

Note that the restricted GEL estimator is asymptotically equivalent to the restricted efficient GMM estimator.

statistic is given by

$$\{\beta^1 : 2n[\hat{P}_n(\hat{\beta},\hat{\lambda}) - \hat{P}_n((\beta^1,\hat{\beta}^2\left(\beta^1\right)),\lambda(\beta^1,\hat{\beta}^2(\beta^1)))] \leq \chi^2_{p_1}(\alpha)\} \tag{4.3}$$

where $\chi^2_{p_1}(\alpha)$ is the $1-\alpha$ critical value from the chi-square distribution with $p_1$ degrees of freedom, $\beta = (\beta^1, \beta^2)$ and $\hat{\beta}^2(\beta^1)$ solves the program $\min_{\beta^2} \sup_{\lambda\in\hat{\Lambda}_n(\beta^1,\beta^2)} \hat{P}_n\left(\left(\beta^1,\beta^2\right),\lambda\right)$. Additionally, the confidence region (4.3) requires the computation of $\hat{\beta}$ and $\hat{\lambda}$.

In practice this procedure is likely to be extremely cumbersome, requiring a grid-search over points $\beta^1$, each of which involves the solution of an optimization problem. However, a confidence region for $\beta_0$ itself is easily computed based on the result $\sup_{\lambda\in\hat{\Lambda}_n(\beta_0)} 2n\hat{P}_n(\beta_0,\lambda) \xrightarrow{d} \chi^2_p$, which avoids estimation of both parameters and variance matrices. A $1-\alpha$ level confidence region is therefore given by[10]

$$\{\beta : 2n\hat{P}_n((\beta,\lambda(\beta))) \leq \chi^2_p(\alpha)\}.$$

# 5 Simulation Experiments

This section studies the bias of GMM and GEL estimators and the performance of test statistics for parametric restrictions in an IV QR model.

## 5.1 Design

We consider the following design

$$y = \beta_0 + w\beta_1 + \sigma(x_1,x_2)(\varepsilon - q_\varepsilon(\theta)),$$
$$w = (x_1+x_2)/3 + \varepsilon + \eta,$$

where $q_\varepsilon(\theta)$ is the conditional $\theta$-quantile of $\varepsilon$. The scale factor $\sigma(x_1,x_2)$ allows for the presence of conditional heteroskedasticity.

The parameter values are $(\beta_0,\beta_1) = (0,0)$ and $x_1 \sim \chi^2(1)$, $x_2 \sim \chi^2(2)$, $\eta \sim N(0,1)$ are distributed independently of $\varepsilon$. An asymmetric covariate design is chosen since symmetric designs tend to be somewhat benign in their discriminatory power. We set $\sigma(x_1,x_2) = 1$

[10] We are grateful to a referee for pointing out this possibility.

and $\sigma(x_1, x_2) = \sqrt{3/14}(1 + (x_1 + x_2)/3)$ to explore the effect of conditional heteroskedasticity; note that in both cases $E[\sigma^2(x_1, x_2)] = 1$. The distributions for $\varepsilon$ considered are $N(0,1)$, $t(3)$ and $\chi^2(1)$. The illustrative conditional quantile $\theta = 0.3$ is chosen to avoid conditional median-unbiased GMM and GEL estimators when $\varepsilon$ is symmetrically distributed.[11]

The parameters $(\beta_0, \beta_1)$ are estimated using GMM with the identity matrix as metric, 2S-GMM with GMM as initial consistent estimator, CUE, EL and ET. We also consider LS, QR, 2SLS and LIML, the two latter estimators being consistent for $\beta_1$ under conditional homoskedasticity.

To deal with non-smoothness and several local optima of the GMM and GEL objective functions, we used the `MATLAB` implementation of a genetic algorithm due to Houck et al. (1995); see also Dorsey and Mayer (1995). Genetic algorithms are stochastic methods that direct a search of large regions of the parameter space to areas where the global optimum is more likely to be.[12] GEL requires the evaluation of $\hat{\lambda}(\beta)$ (2.5) for which, since $\rho(\cdot)$ is twice continuously differentiable, the Newton method was used.[13] Each

[11] Consider the IV QR model $y = w'\beta_\theta + \xi$ for the $\theta$-quantile of $\xi$ and assume $x$ is a vector of intruments, i.e., $\mathcal{P}\{\xi \leq 0|x\} = \theta$. The corresponding moment indicators are $g(z,\beta) = x[\theta - I(y - w'\beta < 0)]$. Denote the GEL estimator by $\hat{\beta}_\theta$. Let $\xi^- = -\xi$ and write $y^- = x'\beta_\theta + \xi^-$; thus $y^- = -y + 2x'\beta_\theta$. Denote the GEL estimator by $\hat{\beta}^-_{1-\theta}$ from the IV QR model $y^- = w'\beta_{1-\theta} + \xi^-$ for the $(1-\theta)$-quantile of $y$ conditional on $x$. The associated moment indicators are $g^-(z,\beta) = x[1 - \theta - I(y^- - w'\beta < 0)] = -x[\theta - I(y - 2w'\beta_\theta + w'\beta \leq 0)]$. Hence, $\hat{\beta}_\theta = -\hat{\beta}^-_{1-\theta} + 2\beta_\theta$, i.e., $\hat{\beta}_\theta - \beta_\theta = -(\hat{\beta}^-_{1-\theta} - \beta_\theta)$. Therefore, only when $\theta = 0.5$ is the GEL estimator median-unbiased. We are grateful to the Joint Editor for this point.

[12] An initial population of points is randomly chosen in a predefined parameter space. For GMM and GEL, to speed up computation, 2000 points were selected in the space $[-2,2] \times [-2,2]$ known to contain the true parameter values. In applications, a larger parameter space and a bigger population size would need to be considered. This procedure did not yield global minima as a second run of the genetic algorithm restricting the parameter space to a neighbourhood of the initial estimates usually obtained an improvement. Therefore these latter minima were used as the final estimates. The second run was initialized by choosing 1000 points in the space $\prod_{i=0}^{1}[-0.5 + b_i, 0.5 + b_i]$ where $(b_0, b_1)$ are the initial estimates. These new estimates initialised the implementation of the `MATLAB` simplex search algorithm to ensure local optima were obtained.

[13] For EL since logarithms do not admit negative arguments the `MATLAB` code due to Owen, available at http://www-stat.stanford.edu/˜owen/empirical/, was used in which logarithms are replaced by

$$f(x) = \begin{cases} \log(x) & \text{if } x \geq \xi \\ \log(\xi) - 1.5 + 2(x/\xi) - 0.5(x/\xi)^2 & \text{if } x < \xi \end{cases}.$$

which has support $\mathcal{R}$. See eq. (12.3), p.235, in Owen (2001).

experiment is based on 1000 replications with sample size $n = 100$.[14]

We also consider LR-, Wald- and score-type statistics for the parametric restriction $\beta_1 = 0$; see section 4. Implementation of the latter statistics requires consistent estimation of $\Omega$ and $G$. The estimators considered for $\Omega$ are

$$\hat{\Omega}_1 = \sum\nolimits_{i=1}^{n} (\theta - I(y_i < \hat{\beta}_0 + \hat{\beta}_1 w_i))^2 x_i x_i'/n, \hat{\Omega}_2 = \sum\nolimits_{i=1}^{n} \hat{p}_i(\theta - I(y_i < \hat{\beta}_0 + \hat{\beta}_1 w_i))^2 x_i x_i',$$

where $\hat{p}_i$, $(i = 1, ..., n)$, are the GEL implied probabilities. Estimation of $G$ is more problematic as it depends of the conditional density function of the error term. First, we consider Powell's (1984) estimator

$$\hat{G}_1 = \sum\nolimits_{i=1}^{n} I(\left| y_i - \hat{\beta}_0 - \hat{\beta}_1 w_i \right| < c_n)) x_i w_i' / \left(2 c_n n\right),$$

where the bandwidth $\hat{c}_n \to 0$ and $\sqrt{n}\hat{c}_n \to 0$ with $\hat{c}_n = \kappa(\Phi^{-1}(\theta + h_n) - \Phi^{-1}(\theta - h_n))$, see Koenker (2005, p.81); here $h_n = 0.1 + n^{-1/3}$, $\kappa$ is Hogg's (1979) robust scale estimate and $\Phi^{-1}(\cdot)$ is the inverse $N(0,1)$ distribution function. Secondly, we also used Buchinsky's (1995) estimator

$$\hat{G}_2 = \sum\nolimits_{i=1}^{n} \phi((y_i - \hat{\beta}_0 - \hat{\beta}_1 w_i)/\hat{b}_n)) x_i w_i'/(\hat{b}_n n),$$

where $\phi(\cdot)$ is the $N(0,1)$ density function and $\hat{b}_n = (3n/4)^{-1/5}\kappa$, the optimal bandwidth for density function estimation when the true density is normal, see Bowman and Azzalini (1997, p.31). GEL efficient versions of these estimators are

$$\hat{G}_3 = \sum\nolimits_{i=1}^{n} \hat{p}_i I(\left| y_i - \hat{\beta}_0 - \hat{\beta}_1 w_i \right| < \hat{c}_n)) x_i w_i' / \left(2\hat{c}_n\right), \hat{G}_4 = \sum\nolimits_{i=1}^{n} \hat{p}_i \phi((y_i - \hat{\beta}_0 - \hat{\beta}_1 w_i)/\hat{b}_n)) x_i w_i'/\hat{b}_n;$$

see Brown and Newey (2002). Test statistics were computed based on 2S-GMM, CUE, EL and ET estimators under the alternative hypothesis.[15] Wald, and score statistics are denoted as $\mathcal{W}_{ij}$ and $\mathcal{S}_{ij}$, where $i$ and $j$ indicate the estimators employed for $\Omega$ and $G$ respectively. Nominal test size is set as 0.05 and power calculations are based on size-corrected tests.[16]

[14] Computations were extremely time-consuming with an average of 4 and 7 days respectively for each design on an Intel Core Duo T2500 2.00 GHz and Pentium 4 with 2 Ghz.

[15] It might be expected that the performance of the tests would be improved if estimation of $\Omega$ and $G$ was performed under the null hypothesis. However, results not reported here indicated that in over-identified models Wald and score tests displayed little or no power.

[16] The performance of Wald and score statistics with the variance estimator $\hat{\Omega} = \theta(1-\theta)\sum\nolimits_{i=1}^{n} x_i x_i'/n$

## 5.2 Results

### 5.2.1 Estimator Bias

**Tables 1 and 2 about here**

Tables 1 and 2 present results on estimator mean bias (MeanB), root mean square error (RMSE), median bias (MedB) and median absolute error (MAE) under conditional homoskedasticity and heteroskedasticity respectively. Given the possibility of non-existence of moments we concentrate on robust measures of central tendency and dispersion.

In general, MedB for $\beta_1$ estimation deteriorates under heteroskedasticity with 2S-GMM and GEL dominating in terms of bias and MAE in Table 2; note though that only GMM- and GEL-based estimators are consistent under heteroskedasticity. In the homoskedastic case, Table 1 shows that LIML, although inconsistent for $\beta_0$, is very competitive not only in terms of bias but also in terms of MAE.[17] With symmetric $t_3$ and asymmetric $\chi^2_1$ errors, 2SLS and LIML are dominated by 2S-GMM and EL respectively in terms of median bias and by 2S-GMM and GEL in respect of MAE with $\chi^2_1$ errors.

In Table 2, GMM is more median biased than 2S-GMM and GEL in most cases apart from CUE and $t_3$ errors. There is no best estimator among the GEL class though in general EL and ET are less biased than CUE. Note that, in some cases, 2S-GMM performs better than GEL in terms of median bias; see, e.g., the $\chi^2_1$ error results in Tables 1 and 2.

### 5.2.2 Test Statistic Performance

**Tables 3-8 about here**

were very similar to that with $\hat{\Omega}_1$. For brevity, results for LM and Pearson tests are not presented since all forms have empirical size considerably in excess of nominal size and have poor size-adjusted power characteristics. Results are available from the authors upon request.

[17]Note that, with homoskedastic errors, after reparameterisation of the intercept coefficient, 2SLS and LIML are first order asymptotically efficient for $\beta_1$ given the mean-based moment restrictions $E[\varepsilon - E[\varepsilon]] = 0$ and $E[x_j(\varepsilon - E[\varepsilon])] = 0$, $(j = 1, 2)$, i.e., $\sqrt{n}(\hat{\beta}^{2SLS} - \beta_0) \xrightarrow{d} N(0, \sigma^2_\varepsilon \left(E[wx']\, E[xx']^{-1} E[xw']\right)^{-1})$. We are grateful to the Joint Editor for this point.

Test statistic performance under homoskedasticity is reported in Tables 3 to 5. Wald tests are generally substantially oversized under $N(0,1)$ and $t_3$ error schemes, which accords with Buchinsky (1995) and Koenker (2005, section 3.10) on Wald tests computed using a kernel-based estimator of $G$. There is, however, some improvement under $\chi_1^2$ errors although these tests are now typically undersized. Overall their power properties are quite similar and tend to exceed that of other test statistics, often substantially. Score tests tend to underreject considerably; there is some evidence that efficient estimation of $\Omega$ and $G$ improves size. Among score tests $\mathcal{S}_{11}$ statistics perform best and display similar power. The empirical sizes of the GEL versions of $\mathcal{LR}$ tests are reasonable with those of $\mathcal{LR}^{\text{EL}}$ and $\mathcal{LR}^{\text{ET}}$ close to nominal size. Generally $\mathcal{LR}^{\text{CUE}}$ displays best power followed by $\mathcal{LR}^{\text{ET}}$ then $\mathcal{LR}^{\text{EL}}$, their power properties being not dissimilar to those of the $\mathcal{S}_{11}$ statistics.

Tables 6 to 8 record test statistic performance under heteroskedasticity, with power generally decreasing relative to that under homoskedasticity. Once again, the Wald statistics are over-sized with $\mathcal{W}_{1\cdot}$ statistics having quite similar and best power properties over all statistics. The $\mathcal{S}_{2\cdot}^{\text{ET}}$ statistics are reasonably sized with others undersized; the $\mathcal{S}_{11}$ statistics again have best and similar power characteristics. The empirical size of $\mathcal{LR}^{\text{2S-GMM}}$ is closest to nominal size; $\mathcal{LR}^{\text{CUE}}$ again displays best power quite similar to that of the $\mathcal{S}_{11}$ statistics.

Overall the LR-type class of tests seems most reliable in terms of size with reasonable power properties, $\mathcal{LR}^{\text{ET}}$ and $\mathcal{LR}^{\text{CUE}}$, although somewhat oversized, being preferable in homoskedastic and heteroskedastic environments respectively.

# 6 Conclusions

This paper obtains the first order asymptotic theory for GEL estimators when the moment indicators are non-smooth. The validity of test statistics for overidentifying moment conditions, parametric restrictions and additional moment conditions previously suggested for the smooth moment indicator set-up is demonstrated. We also show that the Pearson-type tests proposed in Ramalho and Smith (2004) are also valid here. An advantage of tests based on likelihood ratio, Lagrange multiplier and Pearson-type statistics is that estimation of the Jacobian matrix $G$ possibly difficult in practice is not required although an additional optimization of a non-smooth objective function is needed. When the parameters are just-identified these tests are extremely easy to implement if an efficient consistent parameter estimator is available, e.g., Koenker and Bassett's (1978) QR estimator.

The bias of GMM and GEL and test statistic performance are examined in a simulation study. Results indicate that there is no unequivocal ranking of the different estimators in the GEL class. Indeed, 2S-GMM outperforms GEL in some cases. However, 2S-GMM and GEL dominate GMM with identity matrix as metric. Although Wald statistics dominate in terms of power they are typically considerably oversized. The most reliable statistics in terms of empirical size and power appear to be members of the LR class with those based on ET and CUE most suitable for homoskedastic and heteroskedastic enviroments respectively.

# Appendix: Proofs of Results

Throughout the Appendix, $C$ denotes a generic positive constant that may be different in different uses, and CS, M, and T the Cauchy-Schwarz, Minkowski, and triangle inequalities respectively. Also, with probability approaching one is abbreviated as w.p.a.1, positive semi-definite as p.s.d., UWL denotes a uniform weak law of large numbers such as Lemma 2.4 of Newey and McFadden (1994), and CLT is the Lindeberg-Lévy central limit theorem. Define the norm $\|x\|_A = (x'Ax)^{1/2}$ with p.s.d. matrix $A$ as metric.

The proof of asymptotic normality is similar to the approach of Pakes and Pollard (1989) for their Theorem 3.3. We require the following Lemma and Lemmata A2-A3 of Newey and Smith (2004) which are reproduced below.

**Lemma A.1** *If Assumption 2.1 holds then for* $\Lambda_n = \{\lambda : \|\lambda\| \leq Cn^{-1/2}\}$

$$\sup_{\beta\in\mathcal{B},\lambda\in\Lambda_n,1\leq i\leq n} |\lambda' g_i(\beta)| \xrightarrow{p} 0.$$

**Proof:** Write $b_i = \sup_{\beta \in \mathcal{B}} \|g_i(\beta)\|^2$. Now $E[b_i] < \infty$ for $1 \leq i \leq n$ by Assumption 2.1. Then by Lemma 3 of Owen (1990) $\max_{1 \leq i \leq n} b_i = o_p(n^{1/2})$. Hence by CS $\max_{\beta \in \mathcal{B}, \lambda \in \Lambda_n, i \leq n} |\lambda' g_i(\beta)| \leq C n^{-1/2} o_p(n^{1/2}) = o_p(1)$. ■

**Lemma A.2** *If Assumption 2.1 is satisfied, $\bar{\beta} \in \mathcal{B}$, $\bar{\beta} \xrightarrow{p} \beta_0$ and $\hat{g}(\bar{\beta}) = O_p(n^{-1/2})$, then $\bar{\lambda} = \arg\max_{\lambda \in \hat{\Lambda}_n(\bar{\beta})} \hat{P}_n(\bar{\beta}, \lambda)$ exists w.p.a.1, $\bar{\lambda} = O_p(n^{-1/2})$ and $\sup_{\lambda \in \hat{\Lambda}_n(\bar{\beta})} \hat{P}_n(\bar{\beta}, \lambda) \leq \rho_0 + O_p(n^{-1})$.*

**Lemma A.3** *If Assumption 2.1 is satisfied then $\left\|\hat{g}(\hat{\beta})\right\| = O_p(n^{-1/2})$.*

The following is Lemma A.1 of Ramalho and Smith (2004) which is proven there. As stated, Lemma A.1 in Ramalho and Smith (2004) assumes the differentiability of $g(z, \beta)$ at $\beta_0$. However, this assumption is not used at any stage of their proof and therefore their Lemma A.1 is applicable to the context studied here.

**Lemma A.4** *Let Assumption 2.1 hold. Then $n\hat{p}_i = 1 + o_p(1)$ and*

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

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

**Proof of Theorem 2.2:** By T it follows that

$$\left\|g(\hat{\beta})\right\| \leq \left\|\hat{g}(\hat{\beta}) - \hat{g}(\beta_0) - g(\hat{\beta})\right\| + \left\|\hat{g}(\hat{\beta})\right\| + \left\|\hat{g}(\beta_0)\right\|.$$

From Theorem 2.1 and Lemma A.3, $\left\|\hat{g}(\hat{\beta})\right\| = O_p(n^{-1/2})$ and, by Assumption 2.1 (d) and CLT, $\sqrt{n}\left\|\hat{g}(\beta_0)\right\| = O_p(1)$. Now Assumption 2.2 (d) implies that

$$\sqrt{n}\left\|\hat{g}(\hat{\beta}) - \hat{g}(\beta_0) - g(\hat{\beta})\right\| \leq (1 + \sqrt{n}\left\|\hat{\beta} - \beta_0\right\|)o_p(1).$$

Hence,

$$\sqrt{n}\left\|g(\hat{\beta})\right\| \leq (1 + \sqrt{n}\left\|\hat{\beta} - \beta_0\right\|)o_p(1) + O_p(1).$$

Since $g(\beta)$ is differentiable at $\beta_0$, $\left\|g(\hat{\beta})\right\| \geq C\left\|\hat{\beta}-\beta_0\right\|$. Thus

$$\sqrt{n}\left\|\hat{\beta}-\beta_0\right\| \leq (1+\sqrt{n}\left\|\hat{\beta}-\beta_0\right\|)o_p(1)+O_p(1).$$

Hence $(1-o_p(1))\sqrt{n}\left\|\hat{\beta}-\beta_0\right\| \leq O_p(1)$ and therefore $\sqrt{n}\left\|\hat{\beta}-\beta_0\right\| = O_p(1)$.

Next define $\hat{\theta} = (\hat{\beta}', \hat{\lambda}')'$ and $\theta_0 = (\beta_0', 0')'$. We now show that, near its optima, $\hat{P}_n(\beta, \lambda)$ is well approximated by the function

$$\hat{L}_n(\beta,\lambda) = [-G(\beta-\beta_0)]'\lambda - \hat{g}(\beta_0)'\lambda - \frac{1}{2}\lambda'\Omega\lambda.$$

Indeed, we prove that $\left|\hat{P}_n(\hat{\beta},\hat{\lambda}) - \hat{L}_n(\hat{\beta},\hat{\lambda})\right| = o_p(n^{-1})$. A Taylor expansion of $\hat{P}_n(\hat{\beta},\hat{\lambda})$ around $\lambda = 0$ (with Lagrange remainder) gives

$$\hat{P}_n(\hat{\beta},\hat{\lambda}) = -\hat{\lambda}'\hat{g}(\hat{\beta}) + \frac{1}{2}\hat{\lambda}'(\textstyle\sum_{i=1}^n \rho_2(\dot{\lambda}'g_i(\hat{\beta}))g_i(\hat{\beta})g_i(\hat{\beta})'/n)\hat{\lambda}$$

for some $\dot{\lambda}$ on the line segment between $\hat{\lambda}$ and 0. Therefore, by T,

$$\begin{aligned}\left|\hat{P}_n(\hat{\beta},\hat{\lambda}) - \hat{L}_n(\hat{\beta},\hat{\lambda})\right| &\leq \left|-(\hat{g}(\hat{\beta}) - \hat{g}(\beta_0) - G(\hat{\beta}-\beta_0))'\hat{\lambda}\right| \\ &+\left|\frac{1}{2}\hat{\lambda}'(\textstyle\sum_{i=1}^n \rho_2(\dot{\lambda}'g_i(\hat{\beta}))g_i(\hat{\beta})g_i(\hat{\beta})'/n + \Omega)\hat{\lambda}\right|\end{aligned}$$

By CS, Lemmata A.1-A.2, Assumption 2.1 (d) and UWL

$$\begin{aligned}&\left|\hat{\lambda}'(\textstyle\sum_{i=1}^n \rho_2(\dot{\lambda}'g_i(\hat{\beta}))g_i(\hat{\beta})g_i(\hat{\beta})'/n + \Omega)\hat{\lambda}\right| \\ \leq\ &\left\|\hat{\lambda}\right\|^2 \left\|\textstyle\sum_{i=1}^n \rho_2(\dot{\lambda}'g_i(\hat{\beta}))g_i(\hat{\beta})g_i(\hat{\beta})'/n + \Omega)\right\| \\ =\ &O_p\left(n^{-1}\right)o_p(1) = o_p\left(n^{-1}\right).\end{aligned}$$

Finally,

$$\left|-(\hat{g}(\hat{\beta}) - \hat{g}(\beta_0) - G(\hat{\beta}-\beta_0))'\hat{\lambda}\right| \leq \left\|-(\hat{g}(\hat{\beta}) - \hat{g}(\beta_0) - G(\hat{\beta}-\beta_0))\right\|\left\|\hat{\lambda}\right\|.$$

From Assumptions 2.2 (b) and (d), CS and T,

$$\begin{aligned}\left\|-(\hat{g}(\hat{\beta}) - \hat{g}(\beta_0) - G(\hat{\beta}-\beta_0))\right\| &\leq \left\|-(\hat{g}(\hat{\beta}) - \hat{g}(\beta_0) - g(\hat{\beta})\right\| + \left\|G(\hat{\beta}-\beta_0) - g(\hat{\beta})\right\| \qquad \text{(A.1)} \\ &\leq (1+\sqrt{n}\left\|\hat{\beta}-\beta_0\right\|)o_p(n^{-1/2}) + o_p(\left\|\hat{\beta}-\beta_0\right\|) = o_p(n^{-1/2})\end{aligned}$$

Hence, since $\hat{\lambda} = O_p(n^{-1/2})$ by Lemma A.2,

$$\left| -(\hat{g}(\hat{\beta}) - \hat{g}(\beta_0) - G(\hat{\beta} - \beta_0))'\hat{\lambda} \right| \leq o_p\left(n^{-1}\right).$$

Therefore,

$$\left| \hat{P}_n(\hat{\beta}, \hat{\lambda}) - \hat{L}_n(\hat{\beta}, \hat{\lambda}) \right| = o_p\left(n^{-1}\right). \tag{A.2}$$

Now consider the problem $\min_{\beta \in \mathcal{B}} \sup_{\lambda \in \mathcal{R}^m} \hat{L}_n(\beta, \lambda)$. Since $\hat{L}_n(\beta, \lambda)$ is concave in $\lambda$ and $\mathcal{B}$ is compact by Assumption 2.1 (b), the first order conditions for an interior global maximum are satisfied at $\tilde{\theta} = \left(\tilde{\beta}', \tilde{\lambda}'\right)'$, i.e.,

$$-G'\tilde{\lambda} = 0, -G\left(\tilde{\beta} - \beta_0\right) - \hat{g}(\beta_0) - \Omega\tilde{\lambda} = 0, \tag{A.3}$$

which may be stacked as

$$\begin{pmatrix} 0 \\ -\sqrt{n}\hat{g}(\beta_0) \end{pmatrix} + M\sqrt{n}\left(\tilde{\theta} - \theta\right) = 0$$

where

$$M = -\begin{pmatrix} 0 & G' \\ G & \Omega \end{pmatrix}.$$

Solving

$$\begin{aligned} \sqrt{n}\left(\tilde{\theta} - \theta\right) &= M^{-1}\begin{pmatrix} 0 \\ \sqrt{n}\hat{g}(\beta_0) \end{pmatrix} \\ &= -\begin{pmatrix} \Sigma G'\Omega^{-1} \\ P \end{pmatrix}\sqrt{n}\hat{g}(\beta_0). \end{aligned}$$

Consequently by CLT $\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, \mathtt{diag}(\Sigma, P))$.

The final step of the proof requires $\sqrt{n}(\hat{\theta} - \tilde{\theta}) = o_p(1)$. First, we prove $\left\| g\left(\tilde{\beta}\right) \right\| = O_p(n^{-1/2})$. By the differentiability of $g(\beta)$ at $\beta_0$

$$\left\| g\left(\tilde{\beta}\right) \right\| \leq \left\| G\left(\tilde{\beta} - \beta_0\right) \right\| + o_p(\left\| \tilde{\beta} - \beta_0 \right\|) = O_p(n^{-1/2}).$$

Next, by Assumption 2.2 (d)

$$\begin{aligned} &\left\| \hat{g}\left(\tilde{\beta}\right) \right\| \leq \left\| \hat{g}\left(\tilde{\beta}\right) - \hat{g}(\beta_0) - g\left(\tilde{\beta}\right) \right\| + \left\| \hat{g}(\beta_0) \right\| + \left\| g\left(\tilde{\beta}\right) \right\| \\ &\leq (1 + n^{1/2}\left\| \tilde{\beta} - \beta_0 \right\|)o_p(n^{-1/2}) + O_p(n^{-1/2}) = O_p(n^{-1/2}). \end{aligned}$$

From these results, by the same arguments used above for $\hat{\theta}$, $\left|\hat{P}_n(\tilde{\beta},\hat{\lambda})-\hat{L}_n(\tilde{\beta},\hat{\lambda})\right| = o_p\left(n^{-1}\right)$. It then follows that

$$\begin{aligned}\hat{L}_n(\hat{\beta},\hat{\lambda})-o_p\left(n^{-1}\right) &\leq \hat{P}_n(\hat{\beta},\hat{\lambda}) \\ &\leq \hat{P}_n(\tilde{\beta},\hat{\lambda})+o_p\left(n^{-1}\right) \\ &\leq \hat{L}_n(\tilde{\beta},\hat{\lambda})+o_p\left(n^{-1}\right).\end{aligned} \tag{A.4}$$

Thus, $\hat{L}_n(\hat{\beta},\hat{\lambda})=\hat{L}_n(\tilde{\beta},\hat{\lambda})+o_p\left(n^{-1}\right)$. Hence, since $\hat{L}_n(\hat{\beta},\hat{\lambda})-\hat{L}_n(\tilde{\beta},\hat{\lambda})=[-G(\hat{\beta}-\tilde{\beta})]'\hat{\lambda}$, $[-G(\hat{\beta}-\tilde{\beta})]'\hat{\lambda}=o_p\left(n^{-1}\right)$. Now $\hat{\lambda}=O_p(n^{-1/2})$ so $G(\hat{\beta}-\tilde{\beta})=o_p(n^{-1/2})$. Therefore $\sqrt{n}(\hat{\beta}-\tilde{\beta})=o_p(1)$ since $G$ is full rank by Assumption 2.2 (c). It remains to show that $\hat{\lambda}-\tilde{\lambda}=o_p(n^{-1/2})$. To show this, notice that $\hat{L}_n\left(\tilde{\beta},\tilde{\lambda}\right)\leq\hat{L}_n(\hat{\beta},\hat{\lambda})$ and, thus, from eq. (A.4), $\hat{L}_n\left(\tilde{\beta},\tilde{\lambda}\right)=\hat{L}_n(\tilde{\beta},\hat{\lambda})+o_p\left(n^{-1}\right)$. But, from the first order conditions eq. (A.3), $\hat{L}_n\left(\tilde{\beta},\tilde{\lambda}\right)-\hat{L}_n(\tilde{\beta},\hat{\lambda})=(\hat{\lambda}-\tilde{\lambda})'\Omega(\hat{\lambda}-\tilde{\lambda})/2$, which implies that $\hat{\lambda}-\tilde{\lambda}=o_p(n^{-1/2})$. ■

**Proof of Theorem 3.1:** Again let $\hat{\theta}=(\hat{\beta}',\hat{\lambda}')'$ and $\tilde{\theta}=(\tilde{\beta}',\tilde{\lambda}')'$.

First we show that $\mathcal{LR}\xrightarrow{d}\chi^2_{m-p}$. From (A.2) that

$$2n\left|\hat{P}_n(\hat{\beta},\hat{\lambda})-\hat{L}_n(\hat{\beta},\hat{\lambda})\right|=o_p(1). \tag{A.5}$$

However, since $\hat{\theta}-\tilde{\theta}=o_p(n^{-1/2})$ from the Proof of Theorem 2.2,

$$2n\hat{L}_n(\hat{\beta},\hat{\lambda})=2n\hat{L}_n\left(\tilde{\beta},\tilde{\lambda}\right)+o_p(1), \tag{A.6}$$

where

$$\hat{L}_n\left(\tilde{\beta},\tilde{\lambda}\right)=\left[-G\left(\tilde{\beta}-\beta_0\right)\right]'\tilde{\lambda}-\hat{g}\left(\beta_0\right)'\tilde{\lambda}-\frac{1}{2}\tilde{\lambda}'\Omega\tilde{\lambda}.$$

Using the first order conditions (A.3)

$$2n\hat{L}_n\left(\tilde{\beta},\tilde{\lambda}\right)=n\tilde{\lambda}'\Omega\tilde{\lambda}. \tag{A.7}$$

Now $\Omega P\Omega P\Omega=\Omega P\Omega$ and $\sqrt{n}\tilde{\lambda}\xrightarrow{d}N(0,P)$ by Theorem 2.2. Therefore, using Theorem 9.2.1 of Rao and Mitra (1971),

$$n\tilde{\lambda}'\Omega\tilde{\lambda}\xrightarrow{D}\chi^2_{m-p}. \tag{A.8}$$

The result for $\mathcal{LR}$ follows from (A.5), (A.6) and (A.7).

The result for the LM statistic follows immediately from (A.8), $\hat{\lambda}-\tilde{\lambda}=o_p(n^{-1/2})$ and $\hat{\Omega}=\Omega+o_p(1)$.

For the score statistic, Assumption 2.2 (d) implies

$$\begin{aligned} n^{1/2}\left\|\hat{g}(\hat{\beta})-g(\hat{\beta})-\hat{g}(\beta_0)\right\| &\le (1+n^{1/2}\left\|\hat{\beta}-\beta_0\right\|)o_p(1) \\ &= o_p(1). \end{aligned}$$

Consequently

$$\sqrt{n}\hat{g}(\hat{\beta})=\sqrt{n}\hat{g}(\beta_0)+\sqrt{n}g(\hat{\beta})+o_p(1). \tag{A.9}$$

In addition, since $g(\hat{\beta})=G(\hat{\beta}-\beta_0)+o_p(n^{-1/2})$ and $G(\hat{\beta}-\tilde{\beta})=o_p(n^{-1/2})$, $\sqrt{n}g(\hat{\beta})=\sqrt{n}G\left(\tilde{\beta}-\beta_0\right)+o_p(1)$. Now, from the first order conditions (A.3), $-G\left(\tilde{\beta}-\beta_0\right)-\hat{g}(\beta_0)-\Omega\tilde{\lambda}=0$. Hence, $\sqrt{n}g(\hat{\beta})=-\sqrt{n}\hat{g}(\beta_0)-\sqrt{n}\Omega\tilde{\lambda}+o_p(1)$. Therefore, substituting into (A.9),

$$\sqrt{n}\hat{g}(\hat{\beta})=-\sqrt{n}\Omega\tilde{\lambda}+o_p(1).$$

Therefore, from and Assumption 2.1 (e),

$$\begin{aligned} \mathcal{S} &= n\tilde{\lambda}'\Omega\hat{\Omega}^{-1}\Omega\tilde{\lambda}+o_p(1) \\ &= n\tilde{\lambda}'\Omega\tilde{\lambda}+o_p(1) \\ &= \mathcal{LM}+o_p(1) \end{aligned}$$

because $\hat{\lambda}-\tilde{\lambda}=o_p(n^{-1/2})$ and $\hat{\Omega}=\Omega+o_p(1)$. Hence $\mathcal{LR}$, $\mathcal{LM}$ and $\mathcal{S}$ are asymptotically equivalent. ■

**Proof of Theorem 3.2:** The proof is identical to Theorem 3.1 of Ramalho and Smith (2004). Indeed, they showed that

$$\mathcal{P}_n^a=\mathcal{LM}+o_p(1)$$

using only Assumption 2.1. Consequently, by Theorem 3.1, the asymptotic distribution of $\mathcal{P}_n^a$ is the same as that given for $\mathcal{LM}$. The proof for $\mathcal{P}_n^b$ follows immediately from this result and Lemma A.4 as does asymptotic equivalence. ■

**Proof of Theorem 4.1:** Proofs for the consistency and asymptotic normality of the efficient GMM estimator with non-smooth moment conditions may be found in Pakes and Pollard (1989) and Newey and McFadden (1994).

Consider the function $\hat{M}_n(\beta) = \|G(\beta - \beta_0) + \hat{g}(\beta_0)\|^2_{\Omega^{-1}}$. Let

$$\tilde{\beta}_{GMM} = \arg\min_{\beta \in \mathcal{B}} \hat{M}_n(\beta).$$

The first order conditions for this problem are

$$G'\Omega^{-1}G(\tilde{\beta}_{GMM} - \beta_0) + G'\Omega^{-1}\hat{g}(\beta_0) = 0. \tag{A.10}$$

We now show that under Assumptions 2.1 and 2.2

$$\left|\hat{Q}_n(\hat{\beta}_{GMM}) - M_n(\hat{\beta}_{GMM})\right| = o_p\left(n^{-1}\right).$$

First notice that $\hat{Q}_n(\hat{\beta}_{GMM}) \leq \hat{Q}_n(\beta_0)$. The right hand side is $O_p(n^{-1})$ since $\hat{\Omega}^{-1} = \Omega^{-1} + o_p(1)$ and $n\|\hat{g}(\beta_0)\|^2_{\Omega^{-1}} \xrightarrow{d} \chi^2_p$. Hence $\left\|\hat{g}(\hat{\beta}_{GMM})\right\| = O_p(n^{-1/2})$. Now

$$\begin{aligned}
\hat{Q}_n(\hat{\beta}_{GMM}) &= \left\|\hat{g}(\hat{\beta}_{GMM})\right\|^2_{\Omega^{-1}} + o_p\left(n^{-1}\right) \\
&= \hat{M}_n(\hat{\beta}_{GMM}) + \left\|\hat{g}(\hat{\beta}_{GMM}) - G(\hat{\beta}_{GMM} - \beta_0) - \hat{g}(\beta_0)\right\|^2_{\Omega^{-1}} \\
&\quad +2(\hat{g}(\hat{\beta}_{GMM}) - G(\hat{\beta}_{GMM} - \beta_0) - \hat{g}(\beta_0))\Omega^{-1}(G(\hat{\beta}_{GMM} - \beta_0) - \hat{g}(\beta_0)) + o_p(n^{-1}).
\end{aligned}$$

Consequently by T

$$\begin{aligned}
\left|\hat{Q}_n(\hat{\beta}_{GMM}) - \hat{M}_n(\hat{\beta}_{GMM})\right| &\leq \left\|\hat{g}(\hat{\beta}_{GMM}) - G(\hat{\beta}_{GMM} - \beta_0) - \hat{g}(\beta_0)\right\|^2_{\Omega^{-1}} \\
+2\left|(\hat{g}(\hat{\beta}_{GMM}) - G(\hat{\beta}_{GMM} - \beta_0) - \hat{g}(\beta_0))\Omega^{-1}(G(\hat{\beta}_{GMM} - \beta_0) - \hat{g}(\beta_0))\right| &+ o_p(n^{-1}).
\end{aligned}$$

Moreover, by T and Assumption 2.2 (d),

$$\begin{aligned}
\left\|\hat{g}(\hat{\beta}_{GMM}) - G(\hat{\beta}_{GMM} - \beta_0) - \hat{g}(\beta_0)\right\| &\leq \left\|\hat{g}(\hat{\beta}_{GMM}) - g(\hat{\beta}_{GMM}) - \hat{g}(\beta_0)\right\| \\
&\quad + \left\|G(\hat{\beta}_{GMM} - \beta_0) - g(\hat{\beta}_{GMM})\right\| \\
&\leq (1 + \sqrt{n}\left\|\hat{\beta}_{GMM} - \beta_0\right\|)o_p(n^{-1/2}) + o_p(\left\|\hat{\beta}_{GMM} - \beta_0\right\|) \\
&= o_p(n^{-1/2})
\end{aligned}$$

since $\sqrt{n}\left\|\hat{\beta}_{GMM}-\beta_0\right\| = O_p(1)$. Hence, by CS,

$$\left\|\hat{g}(\hat{\beta}_{GMM})-G(\hat{\beta}_{GMM}-\beta_0)-\hat{g}(\beta_0)\right\|^2_{\Omega^{-1}} \leq \left\|\hat{g}(\hat{\beta}_{GMM})-G(\hat{\beta}_{GMM}-\beta_0)-\hat{g}(\beta_0)\right\|^2 \left\|\Omega^{-1}\right\| = o_p(n^{-1}).$$

Similarly, by T and CS,

$$\begin{aligned}
& \left|(\hat{g}(\hat{\beta}_{GMM})-G(\hat{\beta}_{GMM}-\beta_0)-\hat{g}(\beta_0))\Omega^{-1}(G(\hat{\beta}_{GMM}-\beta_0)-\hat{g}(\beta_0))\right| \\
\leq\ & \left\|\hat{g}(\hat{\beta}_{GMM})-G(\hat{\beta}_{GMM}-\beta_0)-\hat{g}(\beta_0)\right\| \left\|\Omega^{-1}\right\| \left\|G(\hat{\beta}_{GMM}-\beta_0)\right\| \\
& +\left\|\hat{g}(\hat{\beta}_{GMM})-G(\hat{\beta}_{GMM}-\beta_0)-\hat{g}(\beta_0)\right\| \left\|\Omega^{-1}\right\| \left\|\hat{g}(\beta_0)\right\| \\
=\ & o_p(n^{-1/2})O_p(n^{-1/2}) = o_p(n^{-1}).
\end{aligned}$$

Therefore

$$\left|\hat{Q}_n(\hat{\beta}_{GMM})-\hat{M}_n(\hat{\beta}_{GMM})\right| = o_p(n^{-1}).$$

Now, from (A.3) and (A.10), $\tilde{\beta}_{GMM}-\tilde{\beta} = o_p(n^{-1/2})$. Therefore, as in the Proof of Theorem 2.2, $g(\tilde{\beta}_{GMM}) = O_p(n^{-1/2})$ and $\hat{g}(\tilde{\beta}_{GMM}) = O_p(n^{-1/2})$. Hence, $\left|\hat{Q}_n(\tilde{\beta}_{GMM})-\hat{M}_n(\tilde{\beta}_{GMM})\right|$ $= o_p(n^{-1})$. Therefore

$$\begin{aligned}
\hat{M}_n(\hat{\beta}_{GMM})-o_p\left(n^{-1}\right) &\leq \hat{Q}_n(\hat{\beta}_{GMM}) \\
&\leq \hat{Q}_n(\tilde{\beta}_{GMM})+o_p(n^{-1}) \leq \hat{M}_n(\tilde{\beta}_{GMM})+o_p(n^{-1}).
\end{aligned}$$

Consequently $\hat{M}_n(\hat{\beta}_{GMM}) = \hat{M}_n(\tilde{\beta}_{GMM})+o_p(n^{-1})$. But, from (A.10), $\hat{M}_n(\hat{\beta}_{GMM}) - \hat{M}_n(\tilde{\beta}_{GMM}) = (\hat{\beta}-\tilde{\beta})'G'\Omega^{-1}G(\hat{\beta}-\tilde{\beta})$. Therefore

$$\hat{\beta}_{GMM}-\tilde{\beta}_{GMM} = o_p(n^{-1/2}).$$

Newey and McFadden (1994, Theorem 2.6, p.2132) is also valid for the restricted efficient GMM estimator $\hat{\beta}^r_{GMM}$. This follows from Assumption 2.1 because the restricted parameter space $\mathcal{B}^r$ is compact. To see this $r(\beta)$ is continuous in $\beta$ by Assumption 4.1 (c) and thus $\mathcal{B}^r$ is closed. Since $\mathcal{B}^r$ is a subset of the compact set $\mathcal{B}$, $\mathcal{B}^r$ is also compact. The hypotheses of Newey and McFadden (1994, Theorem 7.2, p.2186) hold for $\hat{\beta}^r_{GMM}$. Assumptions 2.1, 2.2 and 4.1 guarantee that the conditions required for asymptotic normality are satisfied apart from $\beta_0 \in int(\mathcal{B}^r)$. This holds since, if $r(\beta_0) = 0$, the continuously differentiability of $r(\beta)$ in $\beta$ and $rank(R) = r$ from Assumptions 4.1 (c) and (d) ensure that there is always a neighbourhood of $\beta_0$ within which the condition $r(\beta) = 0$ holds.

Let $\hat{M}_n^r(\beta) = \left\| H(\beta - \beta_0) + \hat{h}(\beta_0) \right\|_{\Xi^{-1}}^2$ and

$$\tilde{\beta}_{GMM}^r = \arg\min_{\beta \in \mathcal{B}^r} \hat{M}_n^r(\beta). \tag{A.11}$$

As above $\left|\hat{Q}_n^r(\hat{\beta}_{GMM}^r) - \hat{M}_n^r(\hat{\beta}_{GMM}^r)\right| = o_p(n^{-1})$, $\left|\hat{Q}_n^r(\tilde{\beta}_{GMM}^r) - \hat{M}_n^r(\tilde{\beta}_{GMM}^r)\right| = o_p(n^{-1})$, $\hat{M}_n^r(\hat{\beta}_{GMM}^r) = \hat{M}_n^r(\tilde{\beta}_{GMM}^r) + o_p(n^{-1})$ and $\hat{\beta}_{GMM}^r - \tilde{\beta}_{GMM}^r = o_p(n^{-1/2})$. Now

$$\begin{aligned}\mathcal{LR}^{GMM} &= n[\hat{Q}_n^r(\hat{\beta}_{GMM}^r) - \hat{M}_n^r(\hat{\beta}_{GMM}^r)] - n[\hat{Q}_n(\hat{\beta}_{GMM}) - \hat{M}_n(\hat{\beta}_{GMM})] \\ &\quad + n[\hat{M}_n^r(\hat{\beta}_{GMM}^r) - \hat{M}_n^r(\tilde{\beta}_{GMM}^r)] - n[\hat{M}_n(\hat{\beta}_{GMM}) - \hat{M}_n(\tilde{\beta}_{GMM})] \\ &\quad + n[\hat{M}_n^r(\tilde{\beta}_{GMM}^r) - \hat{M}_n(\tilde{\beta}_{GMM})].\end{aligned}$$

The first four terms are all $o_p(1)$ as noted above. Thus

$$\mathcal{LR}^{GMM} = n[\hat{M}_n^r(\tilde{\beta}_{GMM}^r) - \hat{M}_n(\tilde{\beta}_{GMM})] + o_p(1). \tag{A.12}$$

The Lagrangean of the program (A.11) is $\mathcal{L}^{GMM}(\beta, \mu) = \hat{M}_n^r(\beta) + \mu' r(\beta)$ where $\mu$ is an $r$-vector of Lagrange mulitpliers. As $\tilde{\beta}_{GMM}^r = \beta_0 + O_p(n^{-1/2})$, the first order conditions associated with $\mathcal{L}^{GMM}(\beta, \mu)$ together with Assumptions 4.1 (c) and (d) imply that

$$\sqrt{n}(\tilde{\beta}_{GMM}^r - \beta_0) = -(\Sigma^r - \Sigma^r R(R'\Sigma^r R)^{-1} R'\Sigma^r) H'\Xi^{-1}\hat{h}(\beta_0) + o_p(1).$$

Denote the $(m+s) \times m$ selection matrix $S_g = (I_m, 0)'$, so $S_g'\hat{h}(\beta_0) = \hat{g}(\beta_0)$, $S_g'\Xi S_g = \Omega$ and $S_g' H = G$. Therefore, as $\sqrt{n}(\tilde{\beta}_{GMM} - \beta_0) = -(G'\Omega^{-1}G)^{-1} G'\Omega^{-1}\sqrt{n}\hat{g}(\beta_0)$ from (A.10), (A.12) may be written as

$$\mathcal{LR}^{GMM} = n\hat{h}(\beta_0)'(P^r - S_g P S_g')\hat{h}(\beta_0) + o_p(1).$$

The result then follows from Theorem 9.2.1 of Rao and Mitra (1971). ■

Similarly to $\Sigma$ and $P$, define $\Sigma^r = (H'\Xi^{-1}H)^{-1}$ and $P^r = \Xi^{-1} - \Xi^{-1}H(\Sigma^r - \Sigma^r R'(R\Sigma^r R')^{-1} R\Sigma^r)H'\Xi^{-1}$.

**Lemma A.5** *Let Assumptions 2.1, 2.2 and 4.1 hold. Then*

$$\sqrt{n}\begin{pmatrix} \hat{\beta}^r - \beta_0 \\ \hat{\lambda}^r \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \texttt{diag}\left(\Sigma^r - \Sigma^r R'(R\Sigma^r R')^{-1} R\Sigma^r, P^r\right)\right).$$

**Proof:** First we need to show that $\hat{\beta}^r$ is consistent, that $\hat{h}(\hat{\beta}^r) = O_p(n^{-1/2})$ and $\hat{\eta}^r = O_p(n^{-1/2})$. To do this we verify Assumption 2.1 for $h(z,\beta)$ rather than $g(z,\beta)$. Assumptions 2.1 (a) and (c) hold for $h(z,\beta)$ by Assumptions 4.1 (a) and (b). Assumption 2.1 (b) holds since $\mathcal{B}^r$ is compact from the Proof of Theorem 4.1. To prove Assumption 2.1 (d), by Assumption 4.1 (b),

$$\begin{aligned} E[\sup_{\beta\in\mathcal{B}} \|h(z,\beta)\|^2] &= E[\sup_{\beta\in\mathcal{B}}(\|g(z,\beta)\|^2 + \|q(z,\beta)\|^2)] \\ &\leq E[\sup_{\beta\in\mathcal{B}} \|g(z,\beta)\|^2] + E[\sup_{\beta\in\mathcal{B}} \|q(z,\beta)\|^2] < \infty. \end{aligned}$$

Lastly, Assumption 4.1 (e) guarantees Assumption 2.1 (e) for $h(z,\beta)$.

Secondly, for asymptotic normality, we verify Assumption 2.2. For Assumption 2.2 (a), $\beta_0 \in int(\mathcal{B}^r)$ was shown in the Proof of Theorem 4.1. Assumption 2.2 (b) holds by Assumption 4.1 (c). Assumption 2.2 (c) is immediate for $H = (G', Q')'$. Write $s_n(\beta) = \sqrt{n}(\hat{h}(\beta) - h(\beta))$ where $h(\beta) = E[\hat{h}(\beta)]$. Hence $\|s_n(\beta) - s(\beta_0)\|^2 = \|v_n(\beta) - v(\beta_0)\|^2 + \|w_n(\beta) - w(\beta_0)\|^2$. Consequently, by Assumptions 2.2 (d) and 5.1 (f),

$$\begin{aligned} \sup_{\|\beta-\beta_0\|\leq\delta_n} \frac{\|s_n(\beta) - s(\beta_0)\|}{1+\sqrt{n}\|\beta-\beta_0\|} &= \sup_{\|\beta-\beta_0\|\leq\delta_n} \sqrt{\left(\frac{\|v_n(\beta) - v(\beta_0)\|}{1+\sqrt{n}\|\beta-\beta_0\|}\right)^2 + \left(\frac{\|w_n(\beta) - w(\beta_0)\|}{1+\sqrt{n}\|\beta-\beta_0\|}\right)^2} \\ &\leq \sqrt{\left(\sup_{\|\beta-\beta_0\|\leq\delta_n} \frac{\|v_n(\beta) - v(\beta_0)\|}{1+\sqrt{n}\|\beta-\beta_0\|}\right)^2 + \left(\sup_{\|\beta-\beta_0\|\leq\delta_n} \frac{\|w_n(\beta) - w(\beta_0)\|}{1+\sqrt{n}\|\beta-\beta_0\|}\right)^2} \\ &= o_p(1). \end{aligned}$$

Then, similarly to $\hat{L}_n(\beta,\lambda)$, define

$$\hat{L}_n^r(\beta,\eta) = [-H(\beta-\beta_0)]'\eta - \hat{h}(\beta_0)'\eta - \frac{1}{2}\eta'\Omega\eta.$$

An identical development to that in the Proof of Theorem 2.2 allows a corresponding result to eq. (A.5)

$$2n\left|\hat{P}_n^r(\hat{\beta}^r, \hat{\eta}^r) - \hat{L}_n^r(\hat{\beta}^r, \hat{\eta}^r)\right| = o_p(1); \tag{A.13}$$

cf. eq. (A.2). Denote by $\tilde{\beta}^r$ and $\tilde{\eta}^r$ the optimizers of

$$\min_{\beta \in \mathcal{B}^r} \sup_{\eta \in \mathcal{R}^{m+s}} \hat{L}_n^r(\beta, \eta). \tag{A.14}$$

The Lagrangean corresponding to the program (A.14) is $\mathcal{L}^{GEL}(\beta, \eta, \mu) = \hat{L}_n(\beta, \eta) + \mu' r(\beta)$ where $\mu$ is an $r$-vector of Lagrange multipliers. The first order conditions for an interior global maximum are satisfied as

$$\begin{aligned} -H'\tilde{\eta}^r + R(\tilde{\beta}^r)'\tilde{\mu}^r &= 0 \\ -H(\tilde{\beta}^r - \beta_0) - \hat{h}(\beta_0) - \Xi\tilde{\eta}^r &= 0 \\ r(\tilde{\beta}^r) &= 0; \end{aligned}$$

cf. eq. (A.3). By Assumption 4.1 (c) and $\sqrt{n}\left\|\tilde{\beta}^r - \beta_0\right\| = O_p(1)$ as in the Proof of Theorem 2.2, $R(\tilde{\beta}^r) = R + O_p(n^{-1/2})$ and $r(\tilde{\beta}^r) = R(\tilde{\beta}^r - \beta_0) + o_p(n^{-1/2})$. Moreover, since $\tilde{\eta}^r = O_p(n^{-1/2})$, $\tilde{\mu}^r = O_p(n^{-1/2})$ as $R$ has full rank $r$. Thus

$$\begin{aligned} -H'\tilde{\eta}^r + R'\tilde{\mu}^r &= o_p(n^{-1/2}) \\ -H(\tilde{\beta}^r - \beta_0) - \hat{h}(\beta_0) - \Xi\tilde{\eta}^r &= 0 \\ R(\tilde{\beta}^r - \beta_0) &= o_p(n^{-1/2}). \end{aligned} \tag{A.15}$$

Hence, $\tilde{\mu}^r = (R\Sigma^r R')^{-1}R\Sigma^r H'\tilde{\eta}^r + o_p(n^{-1/2})$ and, thus, $[\Sigma^r - \Sigma^r R'(R\Sigma^r R')^{-1}R\Sigma^r]H'\tilde{\eta}^r = o_p(n^{-1/2})$. Premultiplying by $[\Sigma^r - \Sigma^r R'(R\Sigma^r R')^{-1}R\Sigma^r]H'\Xi^{-1}$ and solving,

$$\sqrt{n}(\tilde{\beta}^r - \beta_0) = -[\Sigma^r - \Sigma^r R'(R\Sigma^r R')^{-1}R\Sigma^r]H'\Xi^{-1}\sqrt{n}\hat{h}(\beta_0) + o_p(1). \tag{A.16}$$

Thus,

$$\sqrt{n}\tilde{\eta}^r = -(\Xi^{-1} - \Xi^{-1}H[\Sigma^r - \Sigma^r R'(R\Sigma^r R')^{-1}R\Sigma^r]H'\Xi^{-1})\sqrt{n}\hat{h}(\beta_0) + o_p(1). \tag{A.17}$$

Therefore,

$$\sqrt{n}\begin{pmatrix} \tilde{\beta}^r - \beta_0 \\ \tilde{\eta}^r \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \texttt{diag}\left(\Sigma^r - \Sigma^r R'(R\Sigma^r R')^{-1}R\Sigma^r, P^r\right)\right).$$

An identical argument to that in the final part of the Proof of Theorem 2.2 shows that $\sqrt{n}\left\|\hat{\beta}^{r}-\tilde{\beta}^{r}\right\|=O_{p}(1)$ and $\sqrt{n}\left\|\hat{\eta}^{r}-\tilde{\eta}^{r}\right\|=O_{p}(1)$. Note that $\sqrt{n}\left\|\hat{\mu}^{r}-\tilde{\mu}^{r}\right\|=O_{p}(1)$. The conclusion of the Lemma then follows directly. ■

**Proof of Theorem 4.2:** First, we consider the LR statistic

$$\mathcal{LR}^{r}= 2n[\hat{P}_{n}^{r}(\hat{\beta}^{r},\hat{\eta}^{r})-\hat{P}_{n}(\hat{\beta},\hat{\lambda})].$$

We follow an approach close to that in the Proof of Theorem 5.2 in Smith (2001). Rewrite

$$\begin{aligned}\mathcal{LR}^{r} &= 2n[\hat{P}_{n}^{r}(\hat{\beta}^{r},\hat{\eta}^{r})-\hat{L}_{n}^{r}(\hat{\beta}^{r},\hat{\eta}^{r})]\\ &\quad -2n[\hat{P}_{n}(\hat{\beta},\hat{\lambda})-\hat{L}_{n}(\hat{\beta},\hat{\lambda})]\\ &\quad +2n[\hat{L}_{n}^{r}(\hat{\beta}^{r},\hat{\eta}^{r})-\hat{L}_{n}(\hat{\beta},\hat{\lambda})];\end{aligned}$$

$\hat{L}_{n}^{r}(\beta,\eta)$ is defined in the Proof of Lemma A.5. Now, by eqs. (A.5) and (A.13), $2n[\hat{P}_{n}(\hat{\beta},\hat{\lambda})-\hat{L}_{n}(\hat{\beta},\hat{\lambda})]=o_{p}(1)$ and $2n[\hat{P}_{n}^{r}(\hat{\beta}^{r},\hat{\eta}^{r})-\hat{L}_{n}^{r}(\hat{\beta}^{r},\hat{\eta}^{r})]=o_{p}(1)$. Therefore,

$$\mathcal{LR}^{r}=2n[\hat{L}_{n}^{r}(\hat{\beta}^{r},\hat{\eta}^{r})-\hat{L}_{n}(\hat{\beta},\hat{\lambda})]+o_{p}(1). \tag{A.18}$$

Similarly to (A.6), $2n\hat{L}_{n}(\hat{\beta},\hat{\lambda})=2n\hat{L}_{n}(\tilde{\beta},\tilde{\lambda})+o_{p}(1)$, $2n\hat{L}_{n}^{r}(\hat{\beta}^{r},\hat{\eta}^{r})=2n\hat{L}_{n}^{r}(\tilde{\beta}^{r},\tilde{\eta}^{r})+o_{p}(1)$ where $\tilde{\beta}^{r}$ and $\tilde{\eta}^{r}$ are the optimisers of $\min_{\beta\in\mathcal{B}^{r}}\sup_{\eta\in\mathcal{R}^{m+s}}\hat{L}_{n}^{r}(\beta,\eta)$. From the first order conditions eq. (A.15) and noting that $\sqrt{n}\tilde{\eta}^{r}=-P^{r}\sqrt{n}\hat{h}(\beta_{0})+o_{p}(1)$,

$$\begin{aligned}2n\hat{L}_{n}^{r}\left(\tilde{\beta}^{r},\tilde{\eta}^{r}\right) &= n(\tilde{\eta}^{r})'\Xi\tilde{\eta}^{r}\\ &= n\hat{h}(\beta_{0})'P^{r}\Xi P^{r}\hat{h}(\beta_{0})+o_{p}(1).\end{aligned}$$

Now $P\Omega P=P$ and $P^{r}\Xi P^{r}=P^{r}$. Define the $(m+s)\times m$ selection matrix $S_{g}=(I_{m},0)'$. Hence, $S_{g}'\hat{h}(\beta_{0})=\hat{g}(\beta_{0})$, $S_{g}'\Xi S_{g}=\Omega$ and $S_{g}'H=G$. Thus, using eqs. (A.3) and (A.7), $2n\hat{L}_{n}\left(\tilde{\beta},\tilde{\lambda}\right)=n\hat{h}(\beta_{0})'S_{g}PS_{g}'\hat{h}(\beta_{0})$. Therefore,

$$2n[\hat{L}_{n}^{r}(\tilde{\beta}^{r},\tilde{\eta}^{r})-\hat{L}_{n}(\tilde{\beta},\tilde{\lambda})]=n\hat{h}(\beta_{0})'(P^{r}-S_{g}PS_{g}')\hat{h}(\beta_{0})+o_{p}(1). \tag{A.19}$$

As $PS_{g}'\Xi P^{r}=PS_{g}'$ and, thus, $PS_{g}'\Xi(P^{r}-S_{g}PS_{g}')=0$, it is straightforward to show that

$$\Xi(P^{r}-S_{g}PS_{g}')\Xi(P^{r}-S_{g}PS_{g}')\Xi=\Xi(P^{r}-S_{g}PS_{g}')\Xi.$$

Additionally,

$$\begin{aligned} tr\left(\Xi(P^r - S_g P S_g')\right) &= tr\left(\Xi P^r\right) - tr\left(P S_g' \Xi S_g\right) \\ &= tr\left(\Xi P^r\right) - tr\left(\Omega P\right) = m + s - p + r - (m - p) = r + s. \end{aligned}$$

Consequently, by Theorem 9.2.1, p.171, of Rao and Mitra (1971), from eqs. (A.18) and (A.19),

$$\begin{aligned} \mathcal{LR}^r &= n\hat{h}\left(\beta_0\right)'\left(P^r - S_g P S_g'\right)\hat{h}\left(\beta_0\right) + o_p(1) \\ &\xrightarrow{d} \chi^2_{r+s}. \end{aligned}$$

To prove the result for the LM statistic, recall that $\sqrt{n}(\hat{\lambda}-\tilde{\lambda}) = o_p(1)$ and $\sqrt{n}\left(\hat{\eta}^r - \tilde{\eta}^r\right) = o_p\left(1\right)$. Write $\tilde{\eta} = (\tilde{\lambda}', \tilde{\mu}')'$ and $\tilde{\mu} = 0$. Thus, as $\hat{\Xi}^r = \Xi + o_p\left(1\right)$,

$$\begin{aligned} \mathcal{LM}^r &= n(\tilde{\eta}^r - \tilde{\eta})'\Xi(\tilde{\eta}^r - \tilde{\eta}) + o_p(1) \\ &= n\hat{h}(\beta_0)'(P^r - S_g P S_g')\Xi(P^r - S_g P S_g')\hat{h}(\beta_0) + o_p(1) \\ &= n\hat{h}(\beta_0)'(P^r - S_g P S_g')\hat{h}(\beta_0) + o_p(1) \\ &= \mathcal{LR}^r + o_p(1). \end{aligned}$$

To show the Wald and the score statistics are asymptotically equivalent to $\mathcal{LM}^r$ and $\mathcal{LR}^r$ we use two steps. First we prove that $\mathcal{W}^r$ is asymptotically equivalent to $\mathcal{S}^r$. Secondly, we demonstrate that $\mathcal{W}^r$ is asymptotically equivalent to $\mathcal{LR}^r$.

Recall $\sqrt{n}(\hat{\beta}^r - \tilde{\beta}^r) = o_p\left(1\right)$. Let $\vartheta = (\beta', \eta', \mu')'$. Since $\hat{\Psi}^r = \Psi + o_p\left(1\right)$,

$$\begin{aligned} \mathcal{W}^r &= n(\tilde{\psi}^{r\prime}, \tilde{\mu}^{r\prime})(S_{\psi,\mu}'\Psi^{-1}S_{\psi,\mu})^{-1}(\tilde{\psi}^{r\prime}, \tilde{\mu}^{r\prime})' + o_p\left(1\right) \\ &= n(\tilde{\vartheta}^r - \tilde{\vartheta})'S_{\psi,\mu}(S_{\psi,\mu}'\Psi^{-1}S_{\psi,\mu})^{-1}S_{\psi,\mu}'(\tilde{\vartheta}^r - \tilde{\vartheta}) + o_p\left(1\right). \end{aligned} \tag{A.20}$$

However, from eq. (A.15),

$$-\Psi(\tilde{\vartheta}^r - \vartheta_0) = \begin{pmatrix} 0 \\ \hat{h}(\beta_0) \\ 0 \end{pmatrix} + o_p(n^{-1/2}).$$

Hence,

$$-\Psi(\tilde{\vartheta}^r - \tilde{\vartheta}) = \begin{pmatrix} 0 \\ \hat{h}(\beta_0) \\ 0 \end{pmatrix} + \Psi(\tilde{\vartheta} - \vartheta_0) + o_p(n^{-1/2}).$$

Consequently

$$(\tilde{\vartheta}^r - \tilde{\vartheta}) = -\Psi^{-1}\begin{pmatrix} 0 \\ \hat{h}(\beta_0) \\ 0 \end{pmatrix} - (\tilde{\vartheta} - \vartheta_0) + o_p(n^{-1/2}). \tag{A.21}$$

Write $\hat{h}_i = h_i(\hat{\beta})$, $(i = 1, ..., n)$. Then, by Lemma A.4,

$$\sqrt{n}\sum_{i=1}^{n}\hat{p}_i h_i(\hat{\beta}) = \sqrt{n}\hat{h}(\hat{\beta}) + (\sum_{i=1}^{n}\hat{h}_i\hat{g}_i'/n)\sqrt{n}\hat{\lambda} + o_p(1)$$

as $\hat{h}(\hat{\beta}) = O_p(n^{-1/2})$ and $\sqrt{n}\hat{\lambda} = O_p(1)$. Now, by UWL, $\sum_{i=1}^{n}\hat{h}_i\hat{g}_i'/n = \Xi S_g + o_p(1)$. In addition by Assumption 4.1 (f), and similarly to eq. (A.9), $\sqrt{n}\hat{h}(\hat{\beta}) = \sqrt{n}\hat{h}(\beta_0) + \sqrt{n}h(\hat{\beta}) + o_p(1)$. Notice also that, as $\sqrt{n}(\hat{\beta} - \tilde{\beta}) = o_p(1)$, $h(\hat{\beta}) = H(\tilde{\beta} - \beta_0) + o_p(n^{-1/2})$ and $\sqrt{n}(\hat{\lambda} - \tilde{\lambda}) = o_p(1)$. Thus

$$\sqrt{n}\sum_{i=1}^{n}\hat{p}_i h_i(\hat{\beta}) = \sqrt{n}\hat{h}(\beta_0) + H\sqrt{n}\left(\tilde{\beta} - \beta_0\right) + \Xi S_g\sqrt{n}\tilde{\lambda} + o_p(1).$$

Write $\tilde{\eta} = (\tilde{\lambda}', 0')'$ and $\tilde{\mu} = 0$. Thus, $H'\tilde{\eta} = 0$. From the first order conditions of the unrestricted GEL problem, i.e., $\sum_{i=1}^{n}\hat{p}_i g_i(\hat{\beta}) = 0$,

$$\begin{aligned} S_{\psi,\mu}\sqrt{n}\sum_{i=1}^{n}\hat{p}_i\begin{pmatrix} q_i(\hat{\beta}) \\ r(\hat{\beta}) \end{pmatrix} &= \sqrt{n}\sum_{i=1}^{n}\hat{p}_i\begin{pmatrix} 0 \\ h_i(\hat{\beta}) \\ r(\hat{\beta}) \end{pmatrix} \\ &= \sqrt{n}\begin{pmatrix} 0 \\ \hat{h}(\beta_0) \\ 0 \end{pmatrix} + \Psi\sqrt{n}(\tilde{\vartheta}^r - \vartheta_0) + o_p(1). \end{aligned}$$

Therefore, by eq. (A.21),

$$\begin{aligned} -S_{\psi,\mu}'\Psi^{-1}S_{\psi,\mu}\sqrt{n}\sum_{i=1}^{n}\hat{p}_i\begin{pmatrix} q_i(\hat{\beta}) \\ r(\hat{\beta}) \end{pmatrix} &= -S_{\psi,\mu}'\Psi^{-1}\sqrt{n}\begin{pmatrix} 0 \\ \hat{h}(\beta_0) \\ 0 \end{pmatrix} - S_{\psi,\mu}'\sqrt{n}(\tilde{\vartheta} - \vartheta_0) + o_p(1) \\ &= S_{\psi,\mu}'\sqrt{n}(\tilde{\vartheta}^r - \tilde{\vartheta}) + o_p(1) \\ &= \sqrt{n}\begin{pmatrix} \tilde{\psi}^r \\ \tilde{\mu}^r \end{pmatrix} + o_p(1). \end{aligned}$$

Substituting this expression in eq. (A.20)

$$\begin{aligned} \mathcal{W}^r &= n\sum_{i=1}^{n}\hat{p}_i\begin{pmatrix} q_i(\hat{\beta}) \\ r(\hat{\beta}) \end{pmatrix}' S_{\psi,\mu}'\Psi^{-1}S_{\psi,\mu}\sum_{i=1}^{n}\hat{p}_i\begin{pmatrix} q_i(\hat{\beta}) \\ r(\hat{\beta}) \end{pmatrix} + o_p(1) \\ &= \mathcal{S}^r + o_p(1) \end{aligned} \tag{A.22}$$

since $\hat{\Psi}^r = \Psi + o_p(1)$ and, by Lemma A.4 and above, $\sqrt{n}\sum_{i=1}^n \hat{p}_i q_i(\hat{\beta}) = O_p(1)$ and $\sqrt{n} r(\hat{\beta}) = O_p(1)$.

To conclude the proof, note that $S_g' \sum_{i=1}^n \hat{p}_i h_i(\hat{\beta}) = 0$. Hence, from eq. (A.22), substituting eqs. (A.21) and (A.22),

$$\begin{aligned}\mathcal{W}^r &= n\sum_{i=1}^n \hat{p}_i(0', h_i(\hat{\beta})', r(\hat{\beta})')\Psi^{-1}\sum_{i=1}^n \hat{p}_i(0', h_i(\hat{\beta})', r(\hat{\beta})')' + o_p(1) \\ &= n(\tilde{\vartheta}^r - \tilde{\vartheta})'\Psi(\tilde{\vartheta}^r - \tilde{\vartheta}) + o_p(1).\end{aligned}$$

Now, since $H'\tilde{\eta} = 0$ and $\tilde{\mu} = 0$, from eqs. (A.15) and (A.17),

$$\begin{aligned}\mathcal{W}^r &= n\hat{h}(\beta_0)'(P^r - S_g P S_g')\hat{h}(\beta_0) + o_p(1) \\ &= \mathcal{LR}^r + o_p(1).\end{aligned}$$

■

**Proof of Theorem 4.3:** We examine $\mathcal{P}_r^c$. The results for $\mathcal{P}_r^a$ and $\mathcal{P}_r^b$ following directly from Lemma A.4.

Let $\hat{h}_i^r = h_i(\hat{\beta}^r)$, $(i = 1, ..., n)$. By Lemma A.4

$$\begin{aligned}n\hat{p}_i^r - n\hat{p}_i &= \hat{h}_i^{r\prime}\hat{\eta}^r(1 + o_p(1)) - \hat{h}_i{}'\hat{\eta}(1 + o_p(1)) + O_p\left(n^{-1}\right) \\ &= \hat{h}_i^{r\prime}(\hat{\eta}^r - \hat{\eta})(1 + o_p(1)) + (\hat{h}_i^r - \hat{h}_i)'\hat{\eta}(1 + o_p(1)) + O_p\left(n^{-1}\right).\end{aligned}$$

Thus,

$$\begin{aligned}\mathcal{P}_r^c &= \sum_{i=1}^n(\hat{h}_i^{r\prime}(\hat{\eta}^r - \hat{\eta})(1 + o_p(1)) + (\hat{h}_i^r - \hat{h}_i)'\hat{\eta}(1 + o_p(1)) + O_p\left(n^{-1}\right))^2 \\ &= n(\hat{\eta}^r - \hat{\eta})'(\sum_{i=1}^n \hat{h}_i^r\hat{h}_i^{r\prime}/n)(\hat{\eta}^r - \hat{\eta})(1 + o_p(1)) \\ &\quad +2n(\hat{\eta}^r - \hat{\eta})'(\sum_{i=1}^n \hat{h}_i^r(\hat{h}_i^r - \hat{h}_i)'/n)\hat{\eta}(1 + o_p(1)) \\ &\quad +n\hat{\eta}'(\sum_{i=1}^n(\hat{h}_i^r - \hat{h}_i)(\hat{h}_i^r - \hat{h}_i)'/n)\hat{\eta}(1 + o_p(1)) + O_p\left(n^{-1}\right),\end{aligned}$$

where the second equality follows from $\hat{h}(\hat{\beta}^r) = O_p(n^{-1/2})$, $\hat{h}(\hat{\beta}) = O_p(n^{-1/2})$, $\hat{\eta}^r - \hat{\eta} = O_p(n^{-1/2})$ and $\hat{\eta} = O_p(n^{-1/2})$.

For the second term, by T and CS,

$$\begin{aligned}\left|(\hat{\eta}^r - \hat{\eta})'(\sum_{i=1}^n \hat{h}_i^r(\hat{h}_i^r - \hat{h}_i)'/n)\hat{\eta}\right| &\leq \|\hat{\eta}^r - \hat{\eta}\|\|\hat{\eta}\|\left\|\sum_{i=1}^n \hat{h}_i^r(\hat{h}_i^r - \hat{h}_i)'/n\right\| \\ &\leq \|\hat{\eta}^r - \hat{\eta}\|\|\hat{\eta}\|(\sum_{i=1}^n\left\|\hat{h}_i^r\right\|^2/n)^{1/2}(\sum_{i=1}^n\left\|\hat{h}_i^r - \hat{h}_i\right\|^2/n)^{1/2} \\ &= o_p(n^{-1}).\end{aligned}$$

The last equality follows as $\|\hat{\eta}^r - \hat{\eta}\| = O_p(n^{-1/2})$, $\|\hat{\eta}\| = O_p(n^{-1/2})$, $\sum_{i=1}^n \left\|\hat{h}_i^r\right\|^2 /n = O_p(1)$, by UWL and Assumption 2.1, and, defining $h_{i0} = h_i(\beta_0)$, $(i = 1, ..., n)$, by T and M,

$$\begin{aligned}\sum\nolimits_{i=1}^n \left\|\hat{h}_i^r - \hat{h}_i\right\|^2 /n &= \sum\nolimits_{i=1}^n \left\|\hat{h}_i^r - h_{i0} + h_{i0} - \hat{h}_i\right\|^2 /n \\ &\leq \sum\nolimits_{i=1}^n (\left\|\hat{h}_i^r - h_{i0}\right\| + \left\|\hat{h}_i - h_{i0}\right\|)^2/n \\ &\leq ((\sum\nolimits_{i=1}^n \left\|\hat{h}_i^r - h_{i0}\right\|^2 /n)^{1/2} + (\sum\nolimits_{i=1}^n \left\|\hat{h}_i - h_{i0}\right\|^2 /n)^{1/2})^2 = o_p(1)\end{aligned}$$

by Newey and McFadden (1994, Lemma 4.3, p.2156).

Finally, by CS,

$$\begin{aligned}\hat{\eta}'(\sum\nolimits_{i=1}^n (\hat{h}_i^r - \hat{h}_i)(\hat{h}_i^r - \hat{h}_i)'/n)\hat{\eta} &\leq \|\hat{\eta}\|^2 (\sum\nolimits_{i=1}^n \left\|\hat{h}_i^r - \hat{h}_i\right\|^2 /n) \\ &= o_p(n^{-1}).\end{aligned}$$

Therefore, as $\sum_{i=1}^n \hat{h}_i^r \hat{h}_i^{r\prime}/n \xrightarrow{p} \Xi$ by UWL from Assumptions 2.1 and 5.1,

$$\begin{aligned}\mathcal{P}_r^c &= n\left(\hat{\eta}^r - \hat{\eta}\right)' \Xi \left(\hat{\eta}^r - \hat{\eta}\right) + o_p(1) \\ &= \mathcal{LM}^r + o_p\left(1\right)\end{aligned}$$

and the result follows from Theorem 4.2. ■

■

**Proof of Theorem E.1:** To prove this theorem we establish that Assumption E.1 implies Assumption 2.1.

First notice that Assumption E.1 (a) implies $E[sgn_\theta(\xi)x] = 0$. Secondly Assumption 2.1 (b) is ensured by Assumption E.1 (b). Thirdly the moment indicator $sgn_\theta(y - w'\beta)x$ is continuous at each $\beta \in \mathcal{B}$ w.p.a.1 and therefore Assumption 2.1 (c) holds. Fourthly

$$\begin{aligned}E[\sup_{\beta\in\mathcal{B}} \|sgn_\theta(y - w'\beta)x\|^2] &\leq E[\|x\|^2 \left|sgn_\theta(y - w'\beta)\right|^2] \\ &\leq (1+\theta)^2 E[\|x\|^2]\end{aligned}$$

which is finite by Assumption E.1 (c). Consequently Assumption 2.1 (d) is verified. Fifthly $\Omega = \theta(1-\theta)E[xx']$ which is nonsingular by Assumption E.1 (d). Finally as

Assumption E.1 (e) is identical to Assumption 2.1 (f) the result is proven from Theorem 2.1. ■

**Proof of Theorem E.2:** First the derivative matrix of $g(\beta) = E[sgn_\theta(y - w'\beta)x]$ is given by $G = E[f_\xi(0|x,w)xw']$ which exists because $f_\xi\,(0|x,w)$ exists by Assumption E.2 (a) and is full rank by Assumption E.2 (b). Consequently Assumptions 2.2 (b) and (c) hold. To prove Assumption 2.2 (d) note that

$$\sup_{\|\beta-\beta_0\|\le\delta_n} \frac{\sqrt{n}\,\|\hat{g}\,(\beta) - \hat{g}\,(\beta_0) - g\,(\beta)\|}{1+\sqrt{n}\,\|\beta-\beta_0\|} \le \sup_{\|\beta-\beta_0\|\le\delta_n} \sqrt{n}\,\|\hat{g}\,(\beta) - \hat{g}\,(\beta_0) - g\,(\beta)\|$$

The right hand side of this expression is $o_p(1)$ if $\{\hat{g}\,(\beta) - g\,(\beta)\,, n = 1, 2, ...\}$ is stochastically equicontinuous where

$$\hat{g}\,(\beta) - g\,(\beta) = \sum\nolimits_{i=1}^{n} (x_i sgn_\theta\,(\xi_i - w_i'\delta) - E[x(\theta - F_\xi\,(w'\delta|x)])/n$$

where $\delta = \beta - \beta_\theta$ and $F_\xi\,(\cdot|x)$ is the distribution function of $\xi$ conditional on $x$. Let

$$g^*\,(z,\beta) = xsgn_\theta\,(\xi - w'\delta) - E[x(\theta - F_\xi\,(w'\delta|x)].$$

Notice that $sgn_\theta\,(\xi - w'\delta)$ and $E[x(\theta - F_\xi\,(w'\delta|x)]$ are functions of bounded variation and thus are Euclidean by Lemma 22, p.797, of Nolan and Pollard (1987). Also $x$ is Euclidean due to the fact that $E[x]$ is finite and Lemmata II.28 and II.25 of Pollard (1984). It follows by Lemma 2.14, p.1035, of Pakes and Pollard (1989) that $g^*\,(z,\beta)$ is Euclidean with envelope $(1+\theta)\,(\|x\| + \|E[x]\|)$. Moreover, this envelope is square integrable as $E[\|x\|^2]$ is finite. In addition, by T and the $c_r$ inequality

$$\begin{aligned} E[\|g^*\,(z,\beta) - g^*\,(z,\beta_0)\|^2] &= E[\|xsgn_\theta\,(\xi - w'\delta) - E[x(\theta - F_\xi\,(w'\delta|x)] - xsgn_\theta\,(\xi)\|^2] \\ &= E[\|x\,[I(\xi<0) - I(\xi<w'\delta)] - E[x(\theta - F_\xi\,(w'\delta|x)]\|^2] \\ &\le E[(\|x\,[I(\xi<0) - I(\xi<w'\delta)]\| + \|E[x(\theta - F_\xi\,(w'\delta|x)]\|)^2] \\ &\le 2(E[\|x\,[I(\xi<0) - I(\xi<w'\delta)]\|^2] + \|E[x(\theta - F_\xi\,(w'\delta|x)]\|^2). \end{aligned}$$

Since $E[\|x\|^2] < \infty$ it follows that the second term of the last expression converges to zero as $\delta \to 0$ by dominated convergence. For the first term, notice that by iterated expectations

$$E[\|x\,[I(\xi<0) - I(\xi<w'\delta)]\|^2] = E[\|x\|^2\,|\theta - 2\min[\theta, F_\xi\,(w'\delta|x)] + F_\xi\,(w'\delta|x)|].$$

Again this term goes to zero as $\delta \to 0$ by dominated convergence. It follows that $\{\hat{g}(\beta) - g(\beta), n = 1, 2, ...\}$ is stochastically equicontinuous by Pakes and Pollard (1989, Theorem 2.17). Therefore, the conclusion follows by Theorem 2.2. ■

# References


Andrews, D.W.K. (1994): "Empirical Process Methods in Econometrics," in Engle, R.F., and D.L. McFadden, eds., *Handbook of Econometrics,* Vol. 4*, 2247-2294.* New York: North Holland.

Andrews, D.W.K. (1997): "A Stopping Rule for the Computation of Generalized Method of Moments Estimators," *Econometrica*, 65 , 913-931.

Bowman, A.W. and A. Azzalini (1997): *Applied Smoothing Techniques for Data Analysis.* Oxford: Oxford University Press.

Brown, B.W., and W.K. Newey (2002): "Generalized Method of Moments, Efficient Bootstrapping and Improved Inference", *Journal of Business Economics and Statistics*, 20, 507-517.

Buchinsky, M. (1995): "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models. A Monte Carlo Study," *Journal of Econometrics*, 68, 303-338.

Chernozhukov, V., and H. Hong (2003): "A MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346.

Christoffersen, P., Hahn, J., and A. Inoue (1999): "Testing, Comparing and Combining Value at Risk Measures", working paper, Wharton School, University of Pennsylvania.

Corcoran, S. (1998): "Bartlett Adjustment of Empirical Discrepancy Statistics", *Biometrika*, 85, 965-972.

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

Daniels, H. (1961): "The Asymptotic Efficiency of a Maximum Likelihood Estimator", in *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 151-163. Berkeley: University of California Press.

Davidson, R., and MacKinnon, J. (1983): "Small Sample Properties of Alternative forms of the Lagrange Multiplier Test", *Economics Letters*, 12, 269-275.

Dorsey, R., and Mayer, W. (1995): "Genetic Algorithms for estimation Problems with Multiple Optima, Non-differentiability and Other Irregular features", *Journal of Business, Economics and Statistics*, 13, 53-66;

Fitzenberger, B. (1997): "The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions", *Journal of Econometrics*, 82, 235-287.

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

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

Hansen, L., J. Heaton, and A. Yaron (1996), "Finite-Sample Properties of Some Alternative GMM Estimators", *Journal of Business and Economic Statistics*, 14, 262-280.

Hoeffding, W. (1965): "Asymptotically Optimal Tests for Multinomial Distributions", (with Discussion), *Annals of Mathematical Statistics*, 36, 369-408.

Hogg, R. (1979): "Statistical Robustness: One View of Its Use in Applications Today", *American Statistician*, 33, 108-116.

Honore, B., and L. Hu (2004): "On the Performance of Some Robust Instrumental Variables Estimators", *Journal of Business Economics and Statistics*, 22, 30-39

Horowitz, J. (1992): "A Smooth Maximum Score Estimator for the Binary Response Model", *Econometrica*, 60, 505-531.

Horowitz, J. (1998): "Bootstrap Methods for Median Regression Models", *Econometrica*, 66, 1327-1351.

Houck, C., J. Joines, and M. Kay (1995): "A Genetic Algorithm for Function Optimization: A Matlab Implementation", NSCU IE TR 95-09, North Carolina State University. http://www.ie.ncsu.edu/mirage/GAToolBox/gaot

Huber, P. (1967): "The Behavior of Maximum Likelihood Estimates Under Nonstandard Condition", in *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221-233. Berkeley: University of California Press.

Huber, P. (1974): *Robust Statistics*. New York: Wiley.

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

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

Johnson, N., and S. Kotz (1972): *Distributions in Statistics: Continuous Univariate Distributions-2*. New York: Wiley.

Kemp, G.C.R. (2005): "GEL Estimation and Inference with Non-Smooth Moment Indicators and Dynamic Data", working paper, University of Essex.

Kitamura, Y. (2001): "Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions", *Econometrica*, 69, 1661-1672.

Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation", *Econometrica*, 65, 861-874.

Koenker, R. (1997): "Rank Tests for Linear Models", in Rao, C.R., and G.S. Maddala, eds., *Handbook of Statistics*, Vol. 15. New York: North Holland.

Koenker, R. (2005): *Quantile Regression.* New York: Cambridge University Press.

Koenker, R., and G. Bassett (1978): "Regression Quantiles", *Econometrica*, 46, 33-50.

Koenker, R., and G. Bassett (1982): "Tests for Linear Hypotheses and L1 Estimation", *Econometrica*, 50, 1577-1583.

Newey, W.K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica*, 53, 1047-1070.

Newey, W.K., and D.L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing", in Engle, R.F., and D.L. McFadden, eds., *Handbook of Econometrics,* Vol. 4*, 2111-2245.* New York: North Holland.

Newey, W.K., and J.L. Powell (1987): "Asymmetric Least Squares Estimation and Testing", *Econometrica*, 55, 819-847.

Newey, W.K., Ramalho, J.J.S., and R.J. Smith (2005): "Asymptotic Bias for GMM and GEL Estimators with Estimated Nuisance Parameters", in Andrews, D.W.K., and J.H. Stock, eds., *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, 245-281. Cambridge: Cambridge University Press.

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

Newey, W.K., and K. West (1987): "Hypothesis Testing with Efficient Method of Moments Estimation", *International Economic Review*, 28, 777-787.

Nolan, D., and D. Pollard (1987): "U-Processes: Rates of Convergence", *Annals of Statistics*, 15, 780-799.

Otsu, T. (2007): "Conditional Empirical Likelihood Estimation and Inference for Quantile Regression Models". Forthcoming in *Journal of Econometrics*.

Owen, A. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional", *Biometrika*, 75, 237-249.

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

Pakes, A., and D. Pollard, D. (1989): "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57, 1027-1057.

Pollard, D. (1984): *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Pollard, D. (1985): "New Ways to Prove Central Limit Theorems", *Econometric Theory*, 1, 295-314.

Powell, J.L. (1984): "Least Absolute Deviation Estimation for Censored Regression Model", *Journal of Econometrics*, 25, 303-325.

Powell, J.L. (1986): "Censored Regression Quantiles", *Journal of Econometrics*, 32, 143-155.

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

Ramalho, J.J.S. (2001): *Alternative Estimation Methods and Specification Tests for Moment Condition Models*. Unpublished Ph.D. thesis, Department of Economics, University of Bristol.

Ramalho, J.J.S., and R.J. Smith (2004): "Goodness of Fit Tests for Moment Conditions Models", working paper, University of Warwick.

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

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

Smith, R.J. (2000): "Empirical Likelihood Estimation and Inference", Marriott, P., and M. Salmon, eds., *Applications of Differential Geometry to Econometrics*, 119-150. Cambridge: Cambridge University Press.

Smith, R.J. (2001): "GEL Method for Moment Condition Models", working paper, University of Bristol. Revised version CWP 19/04, c*e*mmap, I.F.S. and U.C.L. http://cemmap.ifs.org.uk/wps/cwp0419.pdf

Tauchen G. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models", *Journal of Econometrics*, 30, 415-443.

Van der Vaart, A. (1998): *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Weiss, A. (1991): "Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation", *Econometric Theory*, 7, 46-68.

Whang, Y. (2003): "Smoothed Empirical Likelihood Methods for Quantile Regression Models", working paper, Korea University.

Zhang, J., and I. Gijbels (2003): "Sieve Empirical Likelihood and Extensions of the Generalized Least Squares", *Scandinavian Journal of Statistics*, 30, 1-24.

**Table 1. Estimator Performance under Homoskedasticity**

| | $\beta_0$ | | | | $\beta_1$ | | | |
|---|---|---|---|---|---|---|---|---|
| | MeanB | RMSE | MedB | MAE | MeanB | RMSE | MedB | MAE |
| $N(0,1)$ | | | | | | | | |
| LS | 0.1501 | 0.1769 | 0.146 | 0.1544 | 0.3775 | 0.3811 | 0.3766 | 0.3775 |
| QR | −0.2454 | 0.2733 | −0.2524 | 0.2470 | 0.3691 | 0.3752 | 0.3720 | 0.3691 |
| 2SLS | 0.5324 | 0.5578 | 0.5136 | 0.5324 | −0.0049 | 0.1367 | 0.0093 | 0.1024 |
| LIML | 0.5501 | 0.5767 | 0.5284 | 0.5501 | −0.0226 | 0.1459 | −0.0057 | 0.1064 |
| GMM | 0.067 | 0.2921 | 0.0350 | 0.2021 | −0.1182 | 0.3908 | −0.0327 | 0.2425 |
| 2S-GMM | 0.0153 | 0.1751 | 0.0030 | 0.1355 | −0.0466 | 0.2521 | 0.0004 | 0.1651 |
| CUE | 0.0154 | 0.1719 | −0.0014 | 0.1319 | −0.0309 | 0.2387 | 0.0136 | 0.1590 |
| EL | 0.0122 | 0.1706 | −0.0039 | 0.1317 | −0.0384 | 0.2397 | 0.0070 | 0.1594 |
| ET | 0.0135 | 0.1718 | −0.0001 | 0.1328 | −0.0379 | 0.2502 | 0.0102 | 0.1642 |
| $t_3$ | | | | | | | | |
| LS | −0.0299 | 0.1406 | −0.0159 | 0.1079 | 0.6068 | 0.6141 | 0.5973 | 0.6068 |
| QR | −0.4140 | 0.4481 | −0.4038 | 0.4146 | 0.5385 | 0.5468 | 0.5370 | 0.5385 |
| 2SLS | 0.5813 | 0.6481 | 0.5325 | 0.5815 | 0.0035 | 0.2236 | 0.0285 | 0.1680 |
| LIML | 0.6434 | 0.7479 | 0.5780 | 0.6435 | −0.0621 | 0.3896 | −0.0044 | 0.1955 |
| GMM | 0.0781 | 0.3853 | 0.0281 | 0.2589 | −0.1783 | 0.5208 | −0.0325 | 0.3258 |
| 2S-GMM | 0.0227 | 0.2258 | 0.0092 | 0.1689 | −0.1049 | 0.3733 | −0.0173 | 0.2344 |
| CUE | 0.0275 | 0.2250 | 0.0043 | 0.1655 | −0.0787 | 0.3493 | 0.0057 | 0.2218 |
| EL | 0.0259 | 0.2266 | 0.0088 | 0.1659 | −0.0902 | 0.3569 | −0.0034 | 0.2287 |
| ET | 0.0226 | 0.2238 | 0.0054 | 0.1649 | −0.0766 | 0.3390 | 0.0047 | 0.2199 |
| $\chi^2_1$ | | | | | | | | |
| LS | −0.2231 | 0.2753 | −0.2164 | 0.2318 | 0.5341 | 0.5429 | 0.5375 | 0.5341 |
| QR | −0.1942 | 0.2211 | −0.1681 | 0.1942 | 0.2111 | 0.2324 | 0.1955 | 0.2111 |
| 2SLS | 0.8506 | 0.9538 | 0.794 | 0.8513 | 0.0019 | 0.2044 | 0.0229 | 0.1497 |
| LIML | 0.9178 | 1.0315 | 0.8392 | 0.9183 | −0.0312 | 0.2222 | −0.0080 | 0.1555 |
| GMM | 0.1261 | 0.418 | 0.0286 | 0.2307 | −0.0646 | 0.2797 | −0.0134 | 0.1482 |
| 2S-GMM | 0.0306 | 0.1641 | −0.0026 | 0.0886 | −0.0053 | 0.1286 | −0.0001 | 0.0672 |
| CUE | 0.0173 | 0.1333 | −0.0075 | 0.0819 | 0.0092 | 0.1002 | 0.0088 | 0.0628 |
| EL | 0.0211 | 0.1376 | −0.0041 | 0.0832 | 0.0034 | 0.1021 | 0.0038 | 0.0627 |
| ET | 0.0198 | 0.1385 | −0.0043 | 0.0827 | 0.004 | 0.1095 | 0.0031 | 0.0634 |

**Table 2. Estimator Performance under Heteroskedasticity**

| | $\beta_0$ | | | | $\beta_1$ | | | |
|---|---|---|---|---|---|---|---|---|
| | MeanB | RMSE | MedB | MAE | MeanB | RMSE | MedB | MAE |
| $N(0,1)$ | | | | | | | | |
| LS | 0.0780 | 0.1133 | 0.0809 | 0.0947 | 0.4048 | 0.4093 | 0.4061 | 0.4048 |
| QR | −0.1670 | 0.1954 | −0.1616 | 0.1703 | 0.3512 | 0.3566 | 0.3524 | 0.3512 |
| 2SLS | 0.2433 | 0.2914 | 0.2319 | 0.2488 | 0.2372 | 0.3034 | 0.2519 | 0.2652 |
| LIML | 0.2609 | 0.3406 | 0.2476 | 0.2746 | 0.2189 | 0.3266 | 0.2241 | 0.2595 |
| GMM | −0.0080 | 0.3334 | −0.0114 | 0.1973 | −0.0676 | 0.4385 | −0.0185 | 0.2956 |
| 2S-GMM | −0.0161 | 0.1621 | −0.0210 | 0.1161 | −0.0479 | 0.3394 | 0.0056 | 0.2278 |
| CUE | −0.0143 | 0.1589 | −0.0185 | 0.1147 | −0.0219 | 0.3146 | 0.0202 | 0.2189 |
| EL | −0.0150 | 0.1632 | −0.0202 | 0.1167 | −0.0397 | 0.3256 | 0.0045 | 0.2232 |
| ET | −0.0199 | 0.1616 | −0.0219 | 0.1146 | −0.0270 | 0.3134 | 0.0107 | 0.2198 |
| $t_3$ | | | | | | | | |
| LS | −0.0691 | 0.1583 | −0.0557 | 0.1192 | 0.6045 | 0.6141 | 0.5928 | 0.6045 |
| QR | −0.2888 | 0.3211 | −0.2831 | 0.2892 | 0.4969 | 0.5032 | 0.4927 | 0.4969 |
| 2SLS | 0.2510 | 0.3639 | 0.2277 | 0.2868 | 0.2804 | 0.4145 | 0.3094 | 0.3472 |
| LIML | −0.1666 | 14.4676 | 0.2843 | 0.8728 | 0.6523 | 12.7328 | 0.2560 | 0.8198 |
| GMM | −0.0620 | 0.4067 | −0.0335 | 0.2381 | −0.0871 | 0.5685 | −0.0020 | 0.3807 |
| 2S-GMM | −0.0309 | 0.2228 | −0.0190 | 0.1407 | −0.0797 | 0.4161 | 0.0030 | 0.2811 |
| CUE | −0.0365 | 0.2317 | −0.0223 | 0.1442 | −0.0469 | 0.4268 | 0.0271 | 0.2882 |
| EL | −0.0319 | 0.2298 | −0.0206 | 0.1416 | −0.0610 | 0.4182 | 0.0164 | 0.2794 |
| ET | −0.0313 | 0.2251 | −0.0178 | 0.1429 | −0.0592 | 0.4170 | 0.0126 | 0.2816 |
| $\chi^2_1$ | | | | | | | | |
| LS | −0.3496 | 0.4077 | −0.3211 | 0.3507 | 0.5649 | 0.5791 | 0.5581 | 0.5649 |
| QR | −0.2058 | 0.2428 | −0.1720 | 0.2058 | 0.2288 | 0.2536 | 0.2140 | 0.2288 |
| 2SLS | −0.0264 | 0.4395 | −0.0014 | 0.3367 | 0.4034 | 0.4727 | 0.3971 | 0.4090 |
| LIML | 0.0016 | 0.7130 | 0.0529 | 0.3939 | 0.3906 | 0.5512 | 0.3745 | 0.4044 |
| GMM | −0.1446 | 0.6097 | −0.0044 | 0.3333 | 0.0924 | 0.3622 | 0.0071 | 0.2192 |
| 2S-GMM | −0.0634 | 0.3550 | −0.0046 | 0.1536 | 0.0548 | 0.2555 | 0.0041 | 0.1297 |
| CUE | −0.0754 | 0.3494 | −0.0084 | 0.1507 | 0.0727 | 0.2386 | 0.0146 | 0.1278 |
| EL | −0.0768 | 0.3737 | −0.0067 | 0.1560 | 0.0674 | 0.2507 | 0.0100 | 0.1291 |
| ET | −0.0791 | 0.3648 | −0.0063 | 0.1549 | 0.0722 | 0.2426 | 0.0116 | 0.1282 |

**Table 3. Rejection Frequencies: Homoskedastic $N(0,1)$ Errors**

| $\beta_1$ | 0.0 | 0.25 | 0.50 | 0.75 | 1.00 |
|---|---|---|---|---|---|
| $\mathcal{LR}^{\text{2S-GMM}}$ | 0.0330 | 0.1470 | 0.3460 | 0.4920 | 0.5890 |
| $\mathcal{W}_{11}^{\text{2S-GMM}}$ | 0.1060 | 0.2980 | 0.5700 | 0.7400 | 0.8360 |
| $\mathcal{W}_{12}^{\text{2S-GMM}}$ | 0.0970 | 0.2830 | 0.5520 | 0.7250 | 0.8280 |
| $\mathcal{S}_{11}^{\text{2S-GMM}}$ | 0.0250 | 0.2400 | 0.4470 | 0.6090 | 0.7060 |
| $\mathcal{S}_{12}^{\text{2S-GMM}}$ | 0.0220 | 0.1600 | 0.3580 | 0.5140 | 0.5950 |
| $\mathcal{LR}^{\text{CUE}}$ | 0.0370 | 0.2340 | 0.4650 | 0.6180 | 0.7170 |
| $\mathcal{W}_{11}^{\text{CUE}}$ | 0.1120 | 0.2960 | 0.5830 | 0.7360 | 0.8360 |
| $\mathcal{W}_{12}^{\text{CUE}}$ | 0.1110 | 0.3070 | 0.5910 | 0.7360 | 0.8410 |
| $\mathcal{S}_{11}^{\text{CUE}}$ | 0.0250 | 0.2110 | 0.4250 | 0.5890 | 0.6940 |
| $\mathcal{S}_{12}^{\text{CUE}}$ | 0.0250 | 0.1700 | 0.3680 | 0.5190 | 0.6250 |
| $\mathcal{LR}^{\text{EL}}$ | 0.0530 | 0.1900 | 0.4010 | 0.5490 | 0.6450 |
| $\mathcal{W}_{11}^{\text{EL}}$ | 0.1040 | 0.3030 | 0.5880 | 0.7360 | 0.8400 |
| $\mathcal{W}_{12}^{\text{EL}}$ | 0.1040 | 0.2880 | 0.5740 | 0.7290 | 0.8340 |
| $\mathcal{W}_{23}^{\text{EL}}$ | 0.0830 | 0.2900 | 0.5680 | 0.7120 | 0.8200 |
| $\mathcal{W}_{24}^{\text{EL}}$ | 0.0730 | 0.2990 | 0.5730 | 0.7370 | 0.8280 |
| $\mathcal{S}_{11}^{\text{EL}}$ | 0.0210 | 0.2360 | 0.4560 | 0.6150 | 0.7100 |
| $\mathcal{S}_{12}^{\text{EL}}$ | 0.0230 | 0.1660 | 0.3530 | 0.4990 | 0.6080 |
| $\mathcal{S}_{23}^{\text{EL}}$ | 0.0280 | 0.2100 | 0.3960 | 0.5580 | 0.6690 |
| $\mathcal{S}_{24}^{\text{EL}}$ | 0.0290 | 0.2130 | 0.4070 | 0.5660 | 0.6730 |
| $\mathcal{LR}^{\text{ET}}$ | 0.0460 | 0.2170 | 0.4340 | 0.5900 | 0.6890 |
| $\mathcal{W}_{11}^{\text{ET}}$ | 0.0890 | 0.3080 | 0.6030 | 0.7690 | 0.8510 |
| $\mathcal{W}_{12}^{\text{ET}}$ | 0.0750 | 0.3070 | 0.6020 | 0.7610 | 0.8450 |
| $\mathcal{W}_{23}^{\text{ET}}$ | 0.0580 | 0.2830 | 0.5700 | 0.7350 | 0.8260 |
| $\mathcal{W}_{24}^{\text{ET}}$ | 0.0720 | 0.3040 | 0.5820 | 0.7470 | 0.8320 |
| $\mathcal{S}_{11}^{\text{ET}}$ | 0.0200 | 0.2230 | 0.4540 | 0.6200 | 0.7200 |
| $\mathcal{S}_{12}^{\text{ET}}$ | 0.0240 | 0.1820 | 0.3990 | 0.5460 | 0.6540 |
| $\mathcal{S}_{23}^{\text{ET}}$ | 0.0320 | 0.2110 | 0.4090 | 0.5910 | 0.6760 |
| $\mathcal{S}_{24}^{\text{ET}}$ | 0.0340 | 0.2050 | 0.3930 | 0.5770 | 0.6680 |

**Table 4. Rejection Frequencies: Homoskedastic $t_3$ Errors**

| $\beta_1$ | 0.0 | 0.25 | 0.50 | 0.75 | 1.00 |
|---|---|---|---|---|---|
| $\mathcal{LR}^{\text{2S-GMM}}$ | 0.0340 | 0.1050 | 0.2310 | 0.3470 | 0.4190 |
| $\mathcal{W}_{11}^{\text{2S-GMM}}$ | 0.1160 | 0.2220 | 0.4590 | 0.6150 | 0.7120 |
| $\mathcal{W}_{12}^{\text{2S-GMM}}$ | 0.1050 | 0.2240 | 0.4530 | 0.6130 | 0.7130 |
| $\mathcal{S}_{11}^{\text{2S-GMM}}$ | 0.0200 | 0.2080 | 0.3260 | 0.4430 | 0.4970 |
| $\mathcal{S}_{12}^{\text{2S-GMM}}$ | 0.0190 | 0.1260 | 0.2360 | 0.3450 | 0.4150 |
| $\mathcal{LR}^{\text{CUE}}$ | 0.0450 | 0.1720 | 0.3180 | 0.4500 | 0.5230 |
| $\mathcal{W}_{11}^{\text{CUE}}$ | 0.1280 | 0.2230 | 0.4430 | 0.6120 | 0.7160 |
| $\mathcal{W}_{12}^{\text{CUE}}$ | 0.1160 | 0.2450 | 0.4670 | 0.6310 | 0.7260 |
| $\mathcal{S}_{11}^{\text{CUE}}$ | 0.0220 | 0.1890 | 0.3250 | 0.4310 | 0.4950 |
| $\mathcal{S}_{12}^{\text{CUE}}$ | 0.0280 | 0.1200 | 0.2370 | 0.3330 | 0.4060 |
| $\mathcal{LR}^{\text{EL}}$ | 0.0570 | 0.1730 | 0.2960 | 0.4120 | 0.4930 |
| $\mathcal{W}_{11}^{\text{EL}}$ | 0.1240 | 0.2360 | 0.4590 | 0.6080 | 0.6990 |
| $\mathcal{W}_{12}^{\text{EL}}$ | 0.1110 | 0.2450 | 0.4750 | 0.6200 | 0.7130 |
| $\mathcal{W}_{23}^{\text{EL}}$ | 0.0980 | 0.2250 | 0.4530 | 0.6050 | 0.6930 |
| $\mathcal{W}_{24}^{\text{EL}}$ | 0.0900 | 0.2220 | 0.4620 | 0.6190 | 0.6970 |
| $\mathcal{S}_{11}^{\text{EL}}$ | 0.0180 | 0.1840 | 0.3180 | 0.4230 | 0.4920 |
| $\mathcal{S}_{12}^{\text{EL}}$ | 0.0210 | 0.1050 | 0.2220 | 0.3220 | 0.3940 |
| $\mathcal{S}_{23}^{\text{EL}}$ | 0.0300 | 0.1630 | 0.2920 | 0.3860 | 0.4510 |
| $\mathcal{S}_{24}^{\text{EL}}$ | 0.0290 | 0.1640 | 0.3000 | 0.3940 | 0.4680 |
| $\mathcal{LR}^{\text{ET}}$ | 0.0570 | 0.1720 | 0.3020 | 0.4190 | 0.5100 |
| $\mathcal{W}_{11}^{\text{ET}}$ | 0.1010 | 0.2690 | 0.4910 | 0.6560 | 0.7520 |
| $\mathcal{W}_{12}^{\text{ET}}$ | 0.0770 | 0.2620 | 0.4920 | 0.6690 | 0.7580 |
| $\mathcal{W}_{23}^{\text{ET}}$ | 0.0660 | 0.2520 | 0.4730 | 0.6490 | 0.7380 |
| $\mathcal{W}_{24}^{\text{ET}}$ | 0.0750 | 0.2490 | 0.4830 | 0.6450 | 0.7360 |
| $\mathcal{S}_{11}^{\text{ET}}$ | 0.0210 | 0.2040 | 0.3530 | 0.4680 | 0.5480 |
| $\mathcal{S}_{12}^{\text{ET}}$ | 0.0220 | 0.1350 | 0.2740 | 0.3890 | 0.4490 |
| $\mathcal{S}_{23}^{\text{ET}}$ | 0.0360 | 0.1600 | 0.2950 | 0.3990 | 0.4800 |
| $\mathcal{S}_{24}^{\text{ET}}$ | 0.0420 | 0.1420 | 0.2780 | 0.3840 | 0.4550 |

**Table 5. Rejection Frequencies: Homoskedastic $\chi^2_1$ Errors**

| $\beta_1$ | 0.0 | 0.25 | 0.50 | 0.75 | 1.00 |
|---|---|---|---|---|---|
| $\mathcal{LR}^{\text{2S-GMM}}$ | 0.0320 | 0.6940 | 0.8220 | 0.8930 | 0.9860 |
| $\mathcal{W}_{11}^{\text{2S-GMM}}$ | 0.0440 | 0.8700 | 0.9700 | 0.9820 | 0.9880 |
| $\mathcal{W}_{12}^{\text{2S-GMM}}$ | 0.0290 | 0.8740 | 0.9700 | 0.9810 | 0.9880 |
| $\mathcal{S}_{11}^{\text{2S-GMM}}$ | 0.0200 | 0.8110 | 0.9140 | 0.9590 | 0.9640 |
| $\mathcal{S}_{12}^{\text{2S-GMM}}$ | 0.0190 | 0.7280 | 0.8590 | 0.9060 | 0.9310 |
| $\mathcal{LR}^{\text{CUE}}$ | 0.0370 | 0.7740 | 0.8950 | 0.9460 | 0.9580 |
| $\mathcal{W}_{11}^{\text{CUE}}$ | 0.0510 | 0.8630 | 0.9740 | 0.9930 | 0.9900 |
| $\mathcal{W}_{12}^{\text{CUE}}$ | 0.0410 | 0.8750 | 0.9770 | 0.9930 | 0.9920 |
| $\mathcal{S}_{11}^{\text{CUE}}$ | 0.0200 | 0.7960 | 0.9140 | 0.9630 | 0.9600 |
| $\mathcal{S}_{12}^{\text{CUE}}$ | 0.0200 | 0.6940 | 0.8440 | 0.8980 | 0.9170 |
| $\mathcal{LR}^{\text{EL}}$ | 0.0490 | 0.7450 | 0.8730 | 0.9320 | 0.9400 |
| $\mathcal{W}_{11}^{\text{EL}}$ | 0.0440 | 0.8740 | 0.9710 | 0.9830 | 0.9920 |
| $\mathcal{W}_{12}^{\text{EL}}$ | 0.0330 | 0.8810 | 0.9700 | 0.9840 | 0.9920 |
| $\mathcal{W}_{23}^{\text{EL}}$ | 0.0240 | 0.8480 | 0.9590 | 0.9810 | 0.9880 |
| $\mathcal{W}_{24}^{\text{EL}}$ | 0.0250 | 0.8690 | 0.9660 | 0.9800 | 0.9910 |
| $\mathcal{S}_{11}^{\text{EL}}$ | 0.0170 | 0.8150 | 0.9180 | 0.9590 | 0.9620 |
| $\mathcal{S}_{12}^{\text{EL}}$ | 0.0200 | 0.7200 | 0.8550 | 0.9060 | 0.9130 |
| $\mathcal{S}_{23}^{\text{EL}}$ | 0.0260 | 0.7700 | 0.8810 | 0.9190 | 0.9290 |
| $\mathcal{S}_{24}^{\text{EL}}$ | 0.0260 | 0.7710 | 0.8870 | 0.9220 | 0.9280 |
| $\mathcal{LR}^{\text{ET}}$ | 0.0490 | 0.7590 | 0.8810 | 0.9390 | 0.9490 |
| $\mathcal{W}_{11}^{\text{ET}}$ | 0.0410 | 0.8200 | 0.9710 | 0.9890 | 0.9920 |
| $\mathcal{W}_{12}^{\text{ET}}$ | 0.0300 | 0.8150 | 0.9720 | 0.9910 | 0.9920 |
| $\mathcal{W}_{23}^{\text{ET}}$ | 0.0210 | 0.7610 | 0.9590 | 0.9830 | 0.9900 |
| $\mathcal{W}_{24}^{\text{ET}}$ | 0.0200 | 0.8640 | 0.9680 | 0.9850 | 0.9900 |
| $\mathcal{S}_{11}^{\text{ET}}$ | 0.0200 | 0.8240 | 0.9320 | 0.9720 | 0.9760 |
| $\mathcal{S}_{12}^{\text{ET}}$ | 0.0180 | 0.7330 | 0.8720 | 0.9140 | 0.9350 |
| $\mathcal{S}_{23}^{\text{ET}}$ | 0.0350 | 0.7700 | 0.8870 | 0.9330 | 0.9320 |
| $\mathcal{S}_{24}^{\text{ET}}$ | 0.0370 | 0.7660 | 0.8860 | 0.9300 | 0.9350 |

**Table 6. Rejection Frequencies: Heteroskedastic $N(0,1)$ Errors**

| $\beta_1$ | 0.0 | 0.25 | 0.50 | 0.75 | 1.00 |
|---|---|---|---|---|---|
| $\mathcal{LR}^{\text{2S-GMM}}$ | 0.0470 | 0.1110 | 0.2790 | 0.4270 | 0.5000 |
| $\mathcal{W}_{11}^{\text{2S-GMM}}$ | 0.1300 | 0.1940 | 0.4260 | 0.6270 | 0.7150 |
| $\mathcal{W}_{12}^{\text{2S-GMM}}$ | 0.1210 | 0.1850 | 0.4220 | 0.6210 | 0.6980 |
| $\mathcal{S}_{11}^{\text{2S-GMM}}$ | 0.0380 | 0.1890 | 0.3680 | 0.5200 | 0.6030 |
| $\mathcal{S}_{12}^{\text{2S-GMM}}$ | 0.0220 | 0.0920 | 0.2300 | 0.3740 | 0.4650 |
| $\mathcal{LR}^{\text{CUE}}$ | 0.0660 | 0.1520 | 0.3460 | 0.5100 | 0.5940 |
| $\mathcal{W}_{11}^{\text{CUE}}$ | 0.1280 | 0.2100 | 0.4330 | 0.6290 | 0.7290 |
| $\mathcal{W}_{12}^{\text{CUE}}$ | 0.1190 | 0.1940 | 0.4170 | 0.6210 | 0.7170 |
| $\mathcal{S}_{11}^{\text{CUE}}$ | 0.0360 | 0.1800 | 0.3720 | 0.5160 | 0.5880 |
| $\mathcal{S}_{12}^{\text{CUE}}$ | 0.0250 | 0.0930 | 0.2370 | 0.3800 | 0.4750 |
| $\mathcal{LR}^{\text{EL}}$ | 0.0800 | 0.1290 | 0.2880 | 0.4430 | 0.5230 |
| $\mathcal{W}_{11}^{\text{EL}}$ | 0.1310 | 0.2000 | 0.4160 | 0.6030 | 0.7040 |
| $\mathcal{W}_{12}^{\text{EL}}$ | 0.1160 | 0.1960 | 0.4080 | 0.5970 | 0.7020 |
| $\mathcal{W}_{23}^{\text{EL}}$ | 0.0940 | 0.1870 | 0.3970 | 0.5780 | 0.6650 |
| $\mathcal{W}_{24}^{\text{EL}}$ | 0.0840 | 0.1870 | 0.3850 | 0.5660 | 0.6610 |
| $\mathcal{S}_{11}^{\text{EL}}$ | 0.0300 | 0.1750 | 0.3600 | 0.5290 | 0.5950 |
| $\mathcal{S}_{12}^{\text{EL}}$ | 0.0180 | 0.0940 | 0.2210 | 0.3820 | 0.4610 |
| $\mathcal{S}_{23}^{\text{EL}}$ | 0.0320 | 0.1400 | 0.2770 | 0.4190 | 0.4720 |
| $\mathcal{S}_{24}^{\text{EL}}$ | 0.0340 | 0.1530 | 0.2810 | 0.4310 | 0.5100 |
| $\mathcal{LR}^{\text{ET}}$ | 0.0790 | 0.1410 | 0.3160 | 0.4730 | 0.5540 |
| $\mathcal{W}_{11}^{\text{ET}}$ | 0.1210 | 0.1790 | 0.3680 | 0.5660 | 0.6700 |
| $\mathcal{W}_{12}^{\text{ET}}$ | 0.1060 | 0.1780 | 0.3830 | 0.5890 | 0.6810 |
| $\mathcal{W}_{23}^{\text{ET}}$ | 0.0830 | 0.1650 | 0.3300 | 0.5340 | 0.6170 |
| $\mathcal{W}_{24}^{\text{ET}}$ | 0.0890 | 0.1880 | 0.3980 | 0.5900 | 0.6700 |
| $\mathcal{S}_{11}^{\text{ET}}$ | 0.0280 | 0.1910 | 0.3810 | 0.5460 | 0.6160 |
| $\mathcal{S}_{12}^{\text{ET}}$ | 0.0180 | 0.1120 | 0.2800 | 0.4450 | 0.5240 |
| $\mathcal{S}_{23}^{\text{ET}}$ | 0.0440 | 0.1340 | 0.2710 | 0.4160 | 0.5120 |
| $\mathcal{S}_{24}^{\text{ET}}$ | 0.0460 | 0.1270 | 0.2600 | 0.3980 | 0.4840 |

**Table 7. Rejection Frequencies: Heteroskedastic $t_3$ Errors**

| $\beta_1$ | 0.0 | 0.25 | 0.50 | 0.75 | 1.00 |
|---|---|---|---|---|---|
| $\mathcal{LR}^{\text{2S-GMM}}$ | 0.0410 | 0.1120 | 0.2160 | 0.3090 | 0.3690 |
| $\mathcal{W}_{11}^{\text{2S-GMM}}$ | 0.1220 | 0.1760 | 0.3530 | 0.4890 | 0.6040 |
| $\mathcal{W}_{12}^{\text{2S-GMM}}$ | 0.1170 | 0.1700 | 0.3550 | 0.4860 | 0.6000 |
| $\mathcal{S}_{11}^{\text{2S-GMM}}$ | 0.0320 | 0.1670 | 0.3120 | 0.4110 | 0.4880 |
| $\mathcal{S}_{12}^{\text{2S-GMM}}$ | 0.0230 | 0.0990 | 0.1900 | 0.2810 | 0.3580 |
| $\mathcal{LR}^{\text{CUE}}$ | 0.0620 | 0.1520 | 0.2810 | 0.3920 | 0.4520 |
| $\mathcal{W}_{11}^{\text{CUE}}$ | 0.1540 | 0.1440 | 0.3000 | 0.4630 | 0.5810 |
| $\mathcal{W}_{12}^{\text{CUE}}$ | 0.1410 | 0.1340 | 0.2950 | 0.4550 | 0.5800 |
| $\mathcal{S}_{11}^{\text{CUE}}$ | 0.0360 | 0.1490 | 0.2870 | 0.3850 | 0.4800 |
| $\mathcal{S}_{12}^{\text{CUE}}$ | 0.0280 | 0.0910 | 0.1820 | 0.2720 | 0.3440 |
| $\mathcal{LR}^{\text{EL}}$ | 0.0700 | 0.1330 | 0.2460 | 0.3370 | 0.4050 |
| $\mathcal{W}_{11}^{\text{EL}}$ | 0.1410 | 0.1680 | 0.3280 | 0.4880 | 0.5870 |
| $\mathcal{W}_{12}^{\text{EL}}$ | 0.1300 | 0.1780 | 0.3550 | 0.4940 | 0.5910 |
| $\mathcal{W}_{23}^{\text{EL}}$ | 0.1040 | 0.1550 | 0.3110 | 0.4530 | 0.5520 |
| $\mathcal{W}_{24}^{\text{EL}}$ | 0.0990 | 0.1520 | 0.3110 | 0.4590 | 0.5640 |
| $\mathcal{S}_{11}^{\text{EL}}$ | 0.0320 | 0.1650 | 0.3000 | 0.4040 | 0.4830 |
| $\mathcal{S}_{12}^{\text{EL}}$ | 0.0260 | 0.0960 | 0.2110 | 0.2940 | 0.3680 |
| $\mathcal{S}_{23}^{\text{EL}}$ | 0.0360 | 0.1200 | 0.2280 | 0.3190 | 0.3810 |
| $\mathcal{S}_{24}^{\text{EL}}$ | 0.0350 | 0.1250 | 0.2300 | 0.3310 | 0.3960 |
| $\mathcal{LR}^{\text{ET}}$ | 0.0690 | 0.1420 | 0.2630 | 0.3690 | 0.4280 |
| $\mathcal{W}_{11}^{\text{ET}}$ | 0.1320 | 0.1660 | 0.3330 | 0.4920 | 0.6100 |
| $\mathcal{W}_{12}^{\text{ET}}$ | 0.1120 | 0.1720 | 0.3480 | 0.5000 | 0.6210 |
| $\mathcal{W}_{23}^{\text{ET}}$ | 0.0890 | 0.1570 | 0.3050 | 0.4430 | 0.5630 |
| $\mathcal{W}_{24}^{\text{ET}}$ | 0.0970 | 0.1440 | 0.3130 | 0.4460 | 0.5680 |
| $\mathcal{S}_{11}^{\text{ET}}$ | 0.0370 | 0.1640 | 0.3110 | 0.4260 | 0.5030 |
| $\mathcal{S}_{12}^{\text{ET}}$ | 0.0240 | 0.1120 | 0.2220 | 0.3020 | 0.3800 |
| $\mathcal{S}_{23}^{\text{ET}}$ | 0.0440 | 0.1260 | 0.2280 | 0.3260 | 0.3920 |
| $\mathcal{S}_{24}^{\text{ET}}$ | 0.0520 | 0.1000 | 0.1950 | 0.2820 | 0.3470 |

**Table 8. Rejection Frequencies: Heteroskedastic $\chi^2_1$ Errors**

| $\beta_1$ | 0.0 | 0.25 | 0.50 | 0.75 | 1.00 |
|---|---|---|---|---|---|
| $\mathcal{LR}^{\text{2S-GMM}}$ | 0.0510 | 0.6420 | 0.8480 | 0.8740 | 0.8980 |
| $\mathcal{W}_{11}^{\text{2S-GMM}}$ | 0.0840 | 0.6690 | 0.9580 | 0.9660 | 0.9680 |
| $\mathcal{W}_{12}^{\text{2S-GMM}}$ | 0.0790 | 0.6470 | 0.9530 | 0.9620 | 0.9670 |
| $\mathcal{S}_{11}^{\text{2S-GMM}}$ | 0.0340 | 0.7510 | 0.9150 | 0.9290 | 0.9420 |
| $\mathcal{S}_{12}^{\text{2S-GMM}}$ | 0.0340 | 0.6140 | 0.8250 | 0.8300 | 0.8540 |
| $\mathcal{LR}^{\text{CUE}}$ | 0.0520 | 0.7510 | 0.9190 | 0.9440 | 0.9580 |
| $\mathcal{W}_{11}^{\text{CUE}}$ | 0.0860 | 0.6510 | 0.9440 | 0.9650 | 0.9770 |
| $\mathcal{W}_{12}^{\text{CUE}}$ | 0.0790 | 0.6600 | 0.9450 | 0.9630 | 0.9740 |
| $\mathcal{S}_{11}^{\text{CUE}}$ | 0.0330 | 0.7320 | 0.8970 | 0.9210 | 0.9350 |
| $\mathcal{S}_{12}^{\text{CUE}}$ | 0.0340 | 0.5900 | 0.8180 | 0.8450 | 0.8670 |
| $\mathcal{LR}^{\text{EL}}$ | 0.0690 | 0.6910 | 0.8910 | 0.9150 | 0.9340 |
| $\mathcal{W}_{11}^{\text{EL}}$ | 0.0680 | 0.7450 | 0.9550 | 0.9840 | 0.9790 |
| $\mathcal{W}_{12}^{\text{EL}}$ | 0.0620 | 0.7360 | 0.9520 | 0.9780 | 0.9760 |
| $\mathcal{W}_{23}^{\text{EL}}$ | 0.0490 | 0.6790 | 0.9010 | 0.9540 | 0.9660 |
| $\mathcal{W}_{24}^{\text{EL}}$ | 0.0430 | 0.6840 | 0.9070 | 0.9490 | 0.9640 |
| $\mathcal{S}_{11}^{\text{EL}}$ | 0.0350 | 0.7220 | 0.8900 | 0.9040 | 0.9190 |
| $\mathcal{S}_{12}^{\text{EL}}$ | 0.0300 | 0.6310 | 0.8210 | 0.8400 | 0.8600 |
| $\mathcal{S}_{23}^{\text{EL}}$ | 0.0480 | 0.6230 | 0.7930 | 0.8020 | 0.8360 |
| $\mathcal{S}_{24}^{\text{EL}}$ | 0.0470 | 0.6340 | 0.8100 | 0.8200 | 0.8420 |
| $\mathcal{LR}^{\text{ET}}$ | 0.0610 | 0.7170 | 0.9030 | 0.9260 | 0.9490 |
| $\mathcal{W}_{11}^{\text{ET}}$ | 0.0790 | 0.5440 | 0.9290 | 0.9670 | 0.9730 |
| $\mathcal{W}_{12}^{\text{ET}}$ | 0.0720 | 0.5650 | 0.9300 | 0.9680 | 0.9750 |
| $\mathcal{W}_{23}^{\text{ET}}$ | 0.0510 | 0.4400 | 0.8350 | 0.9160 | 0.9390 |
| $\mathcal{W}_{24}^{\text{ET}}$ | 0.0500 | 0.5800 | 0.8900 | 0.9430 | 0.9510 |
| $\mathcal{S}_{11}^{\text{ET}}$ | 0.0310 | 0.7300 | 0.9070 | 0.9270 | 0.9380 |
| $\mathcal{S}_{12}^{\text{ET}}$ | 0.0320 | 0.6190 | 0.8290 | 0.8610 | 0.8770 |
| $\mathcal{S}_{23}^{\text{ET}}$ | 0.0600 | 0.6430 | 0.8220 | 0.8460 | 0.8600 |
| $\mathcal{S}_{24}^{\text{ET}}$ | 0.0600 | 0.6060 | 0.7980 | 0.8190 | 0.8440 |