GEL-Based Inference with Unconditional Moment Inequality Restrictions

Nicky L. Grant
Department of Economics
University of Manchester

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

Preliminary. This draft: May 2017

Abstract

This paper studies the properties of generalised empirical likelihood (GEL) methods for the estimation of and inference on partially identified parameters in models specified by unconditional moment inequality constraints. The central result is, as in moment equality condition models, a large sample equivalence between the scaled optimised GEL objective function and that for generalised method of moments (GMM) with weight matrix equal to the inverse of the efficient GMM metric for moment equality restrictions. Consequently, the paper provides a generalisation of results in the extant literature for GMM for the non-diagonal GMM weight matrix setting. The paper demonstrates that GMM in such circumstances delivers a consistent estimator of the identified set and derives the corresponding rate of convergence. Based on these results the consistency of and rate of convergence for the GEL estimator of the identified set are obtained. A number of alternative equivalent GEL criteria are also considered and discussed. The paper proposes simple conservative uniformly consistent confidence regions for the identified set and the true parameter vector based on both GMM with a non-diagonal weight matrix and GEL. A simulation study examines the efficacy of the non-diagonal GMM and GEL procedures proposed in the paper compares them with the standard diagonal GMM method.

JEL Classification: C12, C13, C14, C19.
Keywords: Moment Inequalities, Generalised Empirical Likelihood, GMM, Set Identification.

1 Introduction

The primary concern of this paper is an examination of the properties of generalised empirical likelihood (GEL) methods for the estimation of and inference on partially identified parameters in models specified by unconditional moment inequality constraints. The central result is, as in moment equality condition models, a large sample equivalence between the scaled optimised GEL objective function and that for generalised method of moments (GMM) with weight matrix equal to the inverse of the moment variance, i.e., the efficient GMM metric for moment equality restrictions. Consequently, the paper provides a generalisation of results in the extant literature for GMM from the diagonal to the non-diagonal GMM weight matrix setting; see, *inter alia*, Chernozhukov et al. (2007), henceforth CHT. The paper demonstrates that GMM in such circumstances delivers a consistent estimator of the identified set and derives the corresponding rate of convergence. Based on these results the consistency of and rate of convergence for the GEL estimator of the identified set are obtained. A number of alternative equivalent GEL criteria are also considered and discussed. The paper proposes simple conservative uniformly consistent confidence regions for the identified set and the true parameter vector based on GMM with a non-diagonal weight matrix and GEL. A simulation study corroborates the main theoretical results of this paper and indicates that empirical likelihood and exponential tilting confidence region estimators have favourable coverage properties relative to GMM and especially continuous updating which has very poor coverage outside the identified set.

The econometric literature concerned with partially identified models has grown rapidly in recent years, especially that for models defined by moment inequality restrictions. The impetus for this research originally arose from the recognition that untenable and thus undesirable assumptions may often be imposed in econometric research to achieve point identification of model parameters thereby reducing the credibility of any resultant inference. The analysis of the properties of extremum-type parameter estimators in partially identified models specified by moment inequality restrictions has received particular attention. CHT provides general conditions for the consistency of estimators for the identified set and resampling methods to generate uniformly consistent confidence regions for either the identified set or the true parameter vector. To date much of this literature has concentrated on the GMM criterion and associated GMM estimators. CHT section 4, pp.1261-1267, develops confidence region estimators for the identified set and true vector of parameters based on GMM with a diagonal weight matrix whereas Rosen

(2008) does so for the latter based on the GMM criterion with the equality moment constraints efficient metric which avoids the necessity of CHT resampling techniques. An important recent contribution, Chen et al. (2016), develops confidence regions for the identified set based on inverting an optimal sample criterion where cut-off values are computed directly from MCMC simulations of a quasi-posterior distribution of the criterion. However, not unlike CHT for GMM, this method also requires a diagonal variance matrix assumption; see Assumption 3.2 and Theorem 3.1 of Chen et al. (2016). Menzel (2014) extends the CHT results for GMM to the case of many moment inequalities; cf. the many moment equalities GMM results of Han and Phillips (2008). Moment inequality selection methods and corresponding methods of inference based on GMM-type estimators are developed in Andrews and Guggenberger (2009), Andrews and Soares (2010) and Andrews and Barwick (2012). Extensions of GMM to conditional moment inequality models have also been considered; see, e.g., Andrews and Shi (2013, 2014), Armstrong (2014, 2015), Armstrong and Chan (2016) and Khan and Tamer (2009). Misspecified moment inequalities are studied in Ponomareva and Tamer (2011) and Bugni et al. (2012).

The criterion function approach of CHT and others, although of general applicability, can be computationally demanding. Another strand of research has focussed on econometric models with compact convex identified sets enabling the identified set to be characterised by its support function which thus provides a computationally tractable representation. See, e.g., Beresteanu and Mollinari (2008), Beresteanu et al. (2011) and Kaido and Santos (2014). Kaido (2016) presents a unification of the two approaches for compact convex identified sets, illustrating the applicability of the results in a number of examples and for models defined by a finite number of moment inequalities.

Despite the many substantial theoretical contributions to research on the estimation of set-identified parameters relatively little is known about the properties of GEL-type estimators. Exceptions are Moon and Schorfheide (2009), which adopts an empirical likelihood approach when parameters are point-identified by over-identifying moment equality restrictions and also subject to moment inequality restrictions, and Canay (2010), which obtains EL-based confidence regions for the true parameter vector when it is partially identified by a set of unconditional moment inequalities. More generally, the asymptotic properties of GEL methods of inference for the identified set and the true parameter vector, the topic of this paper, remain to be developed.

The paper is organised as follows. Section 2 briefly reviews the set-up describing models specified by unconditional moment inequality constraints. Section 3 details GMM and GEL criteria and associated constructs appropriate for estimation and inference in such models. The equivalence of definitions of the identified set based on population GMM and GEL criteria is also discussed and established here. Consistent estimators for the identified set based on GMM and GEL criteria are described in section 4 with, in particular, the asymptotic equivalence of various GEL criteria also shown. Conservative confidence region estimators for the identified set and the true parameter vector based on GMM with a non-diagonal weight matrix and GEL are proposed in section 5. Section 6 provides a simulation study for interval outcomes in a nonlinear regression model to examine the efficacy of GEL procedures proposed in the paper compared with the standard diagonal GMM method. Section 7 summarises and concludes. The appendices contain the technical assumptions of CHT, their verification for nondiagonal metric GMM and GEL together with the proofs of results stated in the text.

Throughout the text z_i, $(i = 1, ..., n)$, denotes a random sample of size n on the observation d_z-dimensional vector z. Positive (semi-) definite is abbreviated as p.(s.)d., f.c.r. full column rank and f.o.c. first order condition. The interior of a set A is denoted as $int(A)$. Superscripted vectors denote the requisite element, e.g., a^j is the jth element of vector a; $\|x\|_- = \|(x)_-\|$ with $(x)_- = \min\{x, 0\}$. 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. The symbols "$\Rightarrow$", "$\xrightarrow{p}$" and "$\xrightarrow{d}$" denote weak convergence, convergence in probability and convergence in distribution respectively and "with probability (approaching) 1" written as "w.p.(a.)1". The Hausdorff distance between sets A and B is defined as $d_H(A, B) = \max[\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)]$ where $d(b, A) = \inf_{a\in A} \|b - a\|$ and $d_H(A, B) = \infty$ if either A or B are empty.

2 Moment Inequality Restrictions

Let $m(z, \theta)$ denote a d_m-vector of known functions of the data observation vector z and the d_θ-vector $\theta \in \Theta$ of unknown parameters where $\Theta \subset \mathcal{R}^{d_\theta}$ is the corresponding parameter space. The moment indicator vector $m(z, \theta)$ will form the basis for inference in the following discussion and analysis. Also let $m(\theta) = \mathbb{E}_{P_0}[m(z, \theta)]$ and $\Omega(\theta) = \mathbb{E}_{P_0}[m(z, \theta)m(z, \theta)']$, $\theta \in \Theta$, where $\mathbb{E}_{P_0}[\cdot]$ denotes expectation taken with respect to the true population probability law (P_0) of z.

Assumption A.1. **(a)** *The parameter space* Θ *is a non-empty and compact subset of* $\mathcal{R}^{d_\theta}$*;* **(b)** $m(z, \theta)$ *is continuous at each* $\theta \in \Theta$ *w.p.1,* $\mathbb{E}_{P_0}[\sup_{\theta\in\Theta} \|m(z, \theta)\|^2] < C < \infty$ *for suitably large* $C > 0$*;* **(c)** $\Omega(\theta)$ *is finite and uniformly positive definite* $\theta \in \Theta$*;* **(d)** *the* *data z_i, $(i = 1, ..., n)$, are defined on a complete probability space $(\Omega, \mathcal{F}, P)$.*

Remark 2.1. Assumption A.1 repeats aspects of Condition M.2(a), p.1265, of CHT.

It is assumed that the true value θ_0 taken by θ satisfies the population unconditional moment inequality condition under P_0

$$\mathbb{E}_{P_0}[m(z, \theta)] \geq 0. \tag{2.1}$$

Remark 2.2. Moment inequalities such as (2.1) arise in many settings, e.g., interval outcomes in regression models of relevance for empirical models of auctions which forms the basis for the experimental design of the simulation study of section 6. See, *inter alia*, the CHT introduction which provides several more examples and the associated discussion in Romano and Shaikh (2008) for other common examples.

In many situations the common assumption that θ_0 uniquely satisfies the inequality restrictions (2.1) is untenable. A more general and less stringent requirement is that there exists a subset of Θ, here denoted by Θ_{P_0} and referred to as the *identified set,* for which these inequality constraints hold, i.e., the identified set Θ_{P_0} consists of all those elements $\theta \in \Theta$ that satisfy the moment inequality restrictions (2.1)

$$\Theta_{P_0} = \{\theta \in \Theta : \mathbb{E}_{P_0}[m(z, \theta)] \geq 0\}. \tag{2.2}$$

It is convenient for the following analysis to define a d_m-vector of complementary slackness parameters $t(\theta)$ by the identity

$$t(\theta) = \mathbb{E}_{P_0}[m(z, \theta)] \tag{2.3}$$

with the consequent equivalent re-expression of the moment inequality constraints (2.1) as the equality restrictions $t(\theta) - \mathbb{E}_{P_0}[m(z, \theta)] = 0$ together with the parametric inequality restrictions $t(\theta) \geq 0$. Thus, the identified set Θ_{P_0} may now be re-defined as

$$\Theta_{P_0} = \{\theta \in \Theta : t(\theta) - \mathbb{E}_{P_0}[m(z, \theta)] = 0, t(\theta) \geq 0\}. \tag{2.4}$$

In the following the identified set Θ_{P_0} is of inferential interest.

3 GMM and GEL

This section first discusses GMM for models specified by the moment inequality restrictions (2.1). A description of the application of GEL then follows; equivalent GEL variants and their properties are detailed in Appendix D. The section is concluded by an analysis and comparison of the corresponding GMM and GEL definitions of the identified set.

Let $m_i(\theta) = m(z_i, \theta)$, $(i = 1, ..., n)$, $\hat{m}_n(\theta) = \sum_{i=1}^n m_i(\theta)/n$ and $\hat{\Omega}_n(\theta) = \sum_{i=1}^n m_i(\theta)m_i(\theta)'/n$. Assumptions A.1(b) and (c) above ensure $\hat{m}_n(\theta) \xrightarrow{p} m(\theta)$ and $\hat{\Omega}_n(\theta) \xrightarrow{p} \Omega(\theta)$ uniformly $\theta \in \Theta$ by UWL.

3.1 GMM

Define the norm $\|x\|_W^2 = x'Wx$ where W is a p.s.d. matrix. A general formulation for GMM appropriate for the moment inequality constraints (2.1) is based on the objective function

$$\begin{aligned} \hat{Q}_n^W(\theta) &= \inf_{t \geq 0} (\hat{m}_n(\theta) - t)' W_n(\theta) (\hat{m}_n(\theta) - t) \\ &= \inf_{t \geq 0} \|\hat{m}_n(\theta) - t\|_{W_n(\theta)}^2 , \end{aligned} \tag{3.1}$$

where $W_n(\theta)$ is assumed to be uniformly p.s.d. $\theta \in \Theta$. The solution $\hat{t}_n(\theta)$ to (3.1) satisfies $\hat{t}_n^j(\theta) = 0$ if $\hat{m}_n^j(\theta) < 0$ and $\hat{m}_n^j(\theta)$ if $\hat{m}_n^j(\theta) \geq 0$, $(j = 1, ..., d_m)$. Cf. Rosen (2008); also see CHT and Romano and Shaikh (2008).

Assumption A.2-GMM. **(a)** *The GMM criterion function $\hat{Q}_n^W(\theta)$ is defined on a neighbourhood Θ' of Θ, and is measurable in $\theta \in \Theta'$;* **(b)** *there exists $W(\theta)$ such that* $\sup_{\theta \in \Theta} |W_n(\theta) - W(\theta)| = o_p(1)$ *where $W(\theta)$ is continuous with finite elements and uniformly positive definite $\theta \in \Theta$.*

Remark 3.1. Assumption A.2-GMM together with Assumption A.1 reproduces Conditions M.2(a) and M.2(e), p.1265, of CHT with an important exception; cf. Assumptions A4 and A5, p.110, of Rosen (2008). That is, CHT Assumption M.2(e), p.1265, which imposes diagonality on the asymptotic GMM weight matrix $W(\theta)$, is relaxed here. Consequently, the GMM criterion $\hat{Q}_n(\theta)$ in (3.1) may no longer be equivalently expressed asymptotically as the CHT sample criterion $\left\|\hat{m}_n(\theta)' W_n(\theta)^{1/2}\right\|_-^2$ unless $W(\theta)$ is diagonal. Assumption A.2-GMM(b) may be straightforwardly verified by application of UWL.

Remark 3.2. Of particular interest is the GMM objective function with the optimal GMM metric in the unconditional moment equality context, i.e., $W_n(\theta) = \hat{\Omega}_n(\theta)^{-1}$; *viz.*

$$\begin{aligned} \hat{Q}_n^{\Omega^{-1}}(\theta) &= \inf_{t \geq 0} (\hat{m}_n(\theta) - t)' \hat{\Omega}_n(\theta)^{-1} (\hat{m}_n(\theta) - t) \qquad (3.2) \\ &= \inf_{t \geq 0} \|\hat{m}_n(\theta) - t\|_{\hat{\Omega}_n(\theta)^{-1}}^2 . \end{aligned}$$

The population counterpart $Q^W(\theta)$ to the GMM criterion (3.1) is defined by

$$
\begin{aligned}
Q^W(\theta) &= \inf_{t\geq 0}(m(\theta)-t)'W(\theta)(m(\theta)-t) \\
&= \inf_{t\geq 0}\|m(\theta)-t\|^2_{W(\theta)}.
\end{aligned} \tag{3.3}
$$

3.2 GEL

It is well know that GEL is first order asymptotically equivalent to optimal GMM in the standard moment equality constraint setting. As is also widely appreciated, GEL includes as special cases empirical likelihood (EL) [Qin and Lawless (1994), Imbens (1997)], exponential tilting (ET) [Kitamura and Stutzer (1997), Imbens, Spady and Johnson (1998)], continuous updating estimation (CUE) [Hansen et al. (1996)] and estimators based on the the Cressie-Read power divergence family [Cressie and Read (1984)]. See *inter alia* Newey and Smith (2004) and Smith (1997, 2011). Canay (2010) develops an EL-based confidence region for the true parameter vector θ_0, but does not study the large sample properties of the EL estimator of the identified set.

To describe GEL let $\rho(v)$ be a function of a scalar v that is concave on its domain $\mathcal{V}$, an open interval containing zero. For expositional convenience but without loss of generality $\rho(0)$ is set equal to 0 below. The standard GEL criterion is then defined as

$$
\hat{P}^\rho_n(\theta,\lambda)=\sum_{i=1}^n \rho(\lambda' m_i(\theta))/n, \tag{3.4}
$$

in which each element of the auxiliary parameter vector $\lambda\in\mathcal{R}^{d_m}$ is associated with a corresponding element of the moment indicator vector $m_i(\theta)$, $(i=1,...,n)$; cf. Newey and Smith (2004) and Smith (1997, 2011).

Let $\hat{\Lambda}^+_n(\theta)=\hat{\Lambda}_n(\theta)\cap\{\lambda\geq 0\}$ where $\hat{\Lambda}_n(\theta)=\{\lambda:\lambda' m_i(\theta)\in\mathcal{V},(i=1,...,n)\}$ constrains the domain of $\rho(\cdot)$ to the concavity region $\mathcal{V}$ identically to the standard moment equality restrictions case; see Newey and Smith (2004). Optimization of $\hat{P}^\rho_n(\theta,\lambda)$ (3.4) with respect to λ is taken over $\hat{\Lambda}^+_n(\theta)$, where the non-negativity restriction $\lambda\geq 0$ reflects the moment inequality constraints (2.1). The profile GEL criterion function $\hat{P}^\rho_n(\theta)$ is then defined by

$$
\hat{P}^\rho_n(\theta)=\sup_{\lambda\in\hat{\Lambda}^+_n(\theta)}\hat{P}^\rho_n(\theta,\lambda). \tag{3.5}
$$

Let $\rho_1(\cdot)$ and $\rho_2(\cdot)$ denote the first and second derivatives of $\rho(\cdot)$ respectively. The next assumption provides the requisite conditions on the profile GEL criterion $\hat{P}^\rho_n(\theta)$

(3.5) and the function $\rho(\cdot)$.

Assumption A.2-GEL. **(a)** *$\hat{P}_n^\rho(\theta)$ is defined on a neighbourhood Θ' of Θ and is measurable in $\theta \in \Theta'$;* **(b)** *$\rho(\cdot)$ is strictly concave and twice continuously differentiable on an open interval $\mathcal{V}$ that includes 0 and $\rho_1(v) < 0$ for all $v \in \mathcal{V}$.*

Remark 3.3. Cf. Assumption A.2-GMM. Assumption A.2-GEL(b) is satisfied by the Cressie-Read (1984) family of divergence measures. In the following, without loss of generality, the first two derivatives of $\rho(\cdot)$ at zero are set to minus unity, i.e., $\rho_1(0) = \rho_2(0) = -1$.

For any $\theta \in \Theta$, define $\hat{\lambda}_n(\theta) = \arg\max_{\lambda \in \hat{\Lambda}_n^+(\theta)} \hat{P}_n^\rho(\theta, \lambda)$ as the solution to the f.o.c. with respect to λ for given θ, i.e.,

$$\sum_{i=1}^n \rho_1(\hat{\lambda}_n(\theta)' m_i(\theta)) m_i(\theta)/n \le 0,\ \hat{\lambda}_n(\theta) \ge 0. \tag{3.6}$$

In particular $\sum_{i=1}^n \rho_1(\hat{\lambda}_n(\theta)' m_i(\theta)) m_i^j(\theta)/n = 0$ and $\hat{\lambda}_n^j(\theta) > 0$ or $\sum_{i=1}^n \rho_1(\hat{\lambda}_n(\theta)' m_i(\theta)) m_i^j(\theta)/n < 0$ and $\hat{\lambda}_n^j(\theta) = 0$, $(j = 1, ..., d_m)$, i.e., $\hat{\lambda}_n(\theta)' \sum_{i=1}^n \rho_1(\hat{\lambda}_n(\theta)' m_i(\theta)) m_i(\theta)/n = 0$.

The GEL empirical or implied probabilities are then defined correspondingly by

$$\hat{\pi}_i^\rho(\theta, \lambda) = \frac{\rho_1(\lambda' m_i(\theta))}{\sum_{k=1}^n \rho_1(\lambda' m_k(\theta))}, (i = 1, ..., n); \tag{3.7}$$

cf. Back and Brown (1993), Newey and Smith (2004) and Brown and Newey (1992, 2002).

Remark 3.4. The GEL implied probabilities $\hat{\pi}_i^\rho(\theta) = \hat{\pi}_i^\rho(\theta, \hat{\lambda}_n(\theta))$, $(i = 1, ..., n)$, (3.7), are non-negative by Assumption A.2-GEL(b), sum to unity and satisfy the sample moment inequality condition $\sum_{i=1}^n \hat{\pi}_i^\rho(\theta) m_i(\theta) \ge 0$ (3.6) defining the f.o.c. for $\hat{\lambda}_n(\theta)$ for given θ. Cf. Assumption A.2-GEL(b).

Remark 3.5. The above optimisation problem may be cast alternatively in terms of the Lagrangean $\tilde{P}_n^\rho(\theta, \lambda, \tau) = \sum_{i=1}^n \rho(\lambda' m_i(\theta))/n + \tau' \lambda$ where τ is the d_m-vector of Lagrange multipliers associated with the inequality constraint $\lambda \ge 0$. The Lagrange multiplier estimator satisfies $\hat{\tau}_n(\theta) \ge 0$ with $\hat{\lambda}_n(\theta)' \hat{\tau}_n(\theta) = 0$ and, in particular, $\hat{\lambda}_n^j(\theta) = 0$ and $\hat{\tau}_n^j(\theta) > 0$ or $\hat{\lambda}_n^j(\theta) > 0$ and $\hat{\tau}_n^j(\theta) = 0$, $(j = 1, ..., d_m)$. The auxiliary parameter estimator $\hat{\lambda}_n(\theta)$ is the solution to the f.o.c. with respect to λ, i.e., $\sum_{i=1}^n \rho_1(\hat{\lambda}_n(\theta)' m_i(\theta)) m_i(\theta)/n +$

$\hat{\tau}_n(\theta) = 0$. Thus $\hat{\tau}_n(\theta)$ satisfies $\hat{\tau}_n(\theta) = -\sum_{i=1}^n \rho_1(\hat{\lambda}_n(\theta)' m_i(\theta)) m_i(\theta)/n$; cf. (3.6).

Remark 3.6. Appendix E gives alternative equivalent forms of GEL criteria; *viz.* $\tilde{P}_n^{\rho,a}(\theta,\lambda,\tau) = \sum_{i=1}^n \rho(\lambda'(m_i(\theta) - \tau))/n$ (E.1), $\tilde{P}_n^{\rho,b}(\theta,\lambda,\tau) = \sum_{i=1}^n [\rho(\lambda' m_i(\theta)) - \rho(\lambda'\tau)]/n$ (E.3) and $\tilde{P}_n^{\rho}(\theta,\lambda,\tau) = \sum_{i=1}^n \rho(\lambda' m_i(\theta))/n + \lambda'\tau$ (E.6), cf. Remark 3.5. Lemmas E.1-E.3 in Appendix E.1 provide detailed statements and, in particular, demonstrate that both the solutions to and the optimised values of the corresponding GEL saddle point problems (3.4), (E.6) and (E.1), (E.3) are identical, i.e., if $(\tilde{\theta}, \tilde{\lambda}, \tilde{\tau})$, where $\tilde{\tau} \in int(\mathcal{T})$, is a saddlepoint of $\tilde{P}_n^{\rho}(\theta,\lambda,\tau)$ or $\tilde{P}_n^{\rho,k}(\theta,\lambda,\tau)$, $(k = a, b)$, then $(\tilde{\theta}, \tilde{\lambda})$ is also a saddlepoint of $\hat{P}_n^{\rho}(\theta,\lambda)$ and, if $(\hat{\theta}, \hat{\lambda})$ is a saddlepoint of $\hat{P}_n^{\rho}(\theta,\lambda)$ and $\hat{\tau} \in int(\mathcal{T})$ for suitable definitions of the slackness parameter $\hat{\tau}$, then $(\hat{\theta}, \hat{\lambda}, \hat{\tau})$ is also saddlepoint of $\tilde{P}_n^{\rho}(\theta,\lambda,\tau)$ or $\tilde{P}_n^{\rho,k}(\theta,\lambda,\tau)$, $(k = a, b)$. Cf. Lemma A.1, p.150, of Moon and Schorfheide (2009).

3.3 Identified Set

The identified set Θ_{P_0} (2.2) is clearly identical to the GMM population counterpart

$$\Theta_{P_0}^W = \{\theta \in \Theta : \theta = \arg\min_{\theta \in \Theta} Q^W(\theta)\} \tag{3.8}$$

where $Q^W(\theta)$ is defined in (3.3).

Let $\hat{P}^{\rho}(\theta)$ denote the population counterpart to the profile GEL criterion $\hat{P}_n^{\rho}(\theta)$ (3.5), i.e., $\hat{P}^{\rho}(\theta) = \sup_{\lambda \geq 0} \mathbb{E}_{P_0}[\rho(\lambda' m(z,\theta)) - \rho(0)]$. The GEL population counterpart $\hat{\Theta}_{P_0}^{\rho}$ to the identified set Θ_{P_0} (2.2) is then defined as

$$\hat{\Theta}_{P_0}^{\rho} = \{\theta \in \Theta : \theta = \arg\min_{\theta \in \Theta} \hat{P}^{\rho}(\theta)\} \tag{3.9}$$

which similarly to Canay (2010) for EL may be shown to be identical to the identified set Θ_{P_0} (2.2).

Remark 3.7. Alternative but equivalent population counterparts $\tilde{\Theta}_{P_0}^{\rho}$ (E.8) and $\tilde{\Theta}_{P_0}^{\rho,k}$, $(k = a, b)$, (E.9) may be defined corresponding to the GEL criteria $\tilde{P}_n^{\rho}(\theta,\lambda,\tau)$ (E.6) and $\tilde{P}_n^{\rho,k}(\theta,\lambda,\tau)$, $(k = a, b)$, (E.1) and (E.3). See Appendix E.3 for a detailed description.

Lemma D.1 in Appendix D formally demonstrates the equivalence of $\hat{\Theta}_{P_0}^{\rho}$ (3.9) with the identified set Θ_{P_0} (2.2). Lemmas E.4 and E.5 in Appendix E do likewise for $\tilde{\Theta}_{P_0}^{\rho}$ (E.8) and $\tilde{\Theta}_{P_0}^{\rho,k}$ (E.9), $(k = a, b)$, with $\hat{\Theta}_{P_0}^{\rho}$ (3.9). Theorem 3.1 summarises these results.

Theorem 3.1. *Suppose that Assumptions A.1 and A.2-GEL are satisfied. Then the GEL population counterparts $\hat{\Theta}^{\rho}_{P_0}$ (3.9), $\tilde{\Theta}^{\rho}_{P_0}$ (E.8) and $\tilde{\Theta}^{\rho,k}_{P_0}$(E.9), ($k = a, b$), to the identified set Θ_{P_0} (2.2) are identical to Θ_{P_0}.*

4 Set Estimation

Let $v_n(\theta) = n^{1/2}(\hat{m}_n(\theta) - \mathbb{E}_{P_0}[m(z,\theta)])$, $\theta \in \Theta$, $\Omega(\theta_a, \theta_b) = \mathbb{E}_{P_0}[v_n(\theta_a)v_n(\theta_b)']$ and $\Omega(\theta) = \Omega(\theta, \theta)$ where $\theta_a, \theta_b \in \Theta$. The following assumptions correspond identically to CHT Conditions M.2(c), M.2(d) and M.2(f), pp.1265-1266, respectively.

Assumption A.3. *The process $v_n(\cdot)$ satisfies a P-Donsker property. In particular, $v_n(\cdot) \Rightarrow v(\cdot)$ where $v(\cdot)$ is a zero-mean Gaussian process on Θ' with covariance function $E_{P_0}[v(\theta_a)v(\theta_b)'] = \Omega(\theta_a, \theta_b)$.*

Assumption A.4. *There exist positive constants $C > 0$ and $\delta > 0$ such that $\|\mathbb{E}_{P_0}[m(z,\theta)]\|_- \geq C \cdot (d(\theta, \Theta_{P_0}) \wedge \delta)$ for all $\theta \in \Theta$ with continuous Jacobian $M(\theta) = \partial m(\theta)/\partial\theta'$ for each $\theta \in \Theta'$.*

Assumption A.5. *There exist positive constants $C > 0$, $M > 0$ and $\delta > 0$ such that, for all $\theta \in \Theta^{-\epsilon}_{P_0}$, $\min_{1 \leq j \leq d_m} \|\mathbb{E}_{P_0}[m^j(z,\theta)]\|_- \geq C \cdot (\epsilon \wedge \delta)$ and $d_H(\Theta^{-\varepsilon}_{P_0}, \Theta_{P_0}) \leq M\varepsilon$ for all $\varepsilon \in [0, \delta]$ where $\Theta^{-\varepsilon}_{P_0} = \{\theta \in \Theta_{P_0} : d(\theta, \Theta \backslash \Theta_{P_0}) \geq \varepsilon\}$.*

Proofs for the following results are provided in Appendices B and C respectively for GMM and GEL.

4.1 GMM

Let

$$\hat{\Theta}^W_n(c) = \{\theta \in \Theta : n\hat{Q}^W_n(\theta) \leq c\}. \tag{4.1}$$

Cf. CHT, eqs. (3.1) and (3.2), p.1253. The GMM estimator of the identified set Θ_{P_0} (2.2) is then defined as the set estimator $\hat{\Theta}^W_n(\hat{c}_W)$ (4.1) for some possibly data dependent level $\hat{c}_W$.

Appendix B establishes the validity for GMM of CHT Conditions C.1, p.1252, C.2, p.1253, and C.3, p.1255, under Assumptions A.1, A.2-GMM and A.3-A.5. These conditions are therefore sufficient for the statement of the following theorem on the consistency and rate of convergence of the GMM set estimator $\hat{\Theta}_n(\hat{c})$ for the identified set Θ_{P_0} given rate restrictions on $\hat{c}$ as provided in CHT Theorem 3.2, p.1255.

THEOREM 4.1. *Let $\hat{c}_W \geq q_n = \inf_{\theta\in\Theta} n\hat{Q}_n(\theta)$ w.p.a.1 and $\hat{c}_W = O_p(1)$. Then, under Assumptions A.1, A.2-GMM and A.3-A.5,* **(a)** *$\hat{\Theta}_n^W(\hat{c}_W)$ is a consistent estimator of the identified set Θ_{P_0}, i.e., $d_H(\hat{\Theta}_n^W(\hat{c}_W), \Theta_{P_0}) = o_p(1)$;* **(b)** *$d_H(\hat{\Theta}_n^W(\hat{c}_W), \Theta_{P_0}) = O_p(n^{-1/2})$.*

REMARK 4.1. Theorem 4.1 is established by verifying the conditions required for CHT Theorem 3.1, p.1254. Unlike CHT Condition M.2(e), p.1265, and the consequent Theorem 4.2, p.1266, Theorem 4.1 does not require the diagonality of the population GMM weight matrix $W(\theta)$, $\theta \in \Theta$, although similarly mild restrictions on the choice of the value $\hat{c}_W$ to those of CHT are imposed. CHT Theorem 4.2, p.1266, also obtains a limiting representation for the statistic $\sup_{\theta\in\Theta_{P_0}} n\hat{Q}_n^W(\theta)$ when the population GMM weight matrix $W(\theta)$, $\theta \in \Theta$, is diagonal. To the best of our knowledge there are as yet no results for the non-diagonal case. Section 5.1 proposes conservative bounds appropriate for GMM criteria with a non-diagonal population weight matrix $W(\theta)$, $\theta \in \Theta$, and, likewise, GEL criteria.

REMARK 4.2. Alternatively, cf. CHT Theorem 3.1, p.1254, a similar result holds if $\hat{c}_W \geq \sup_{\theta\in\Theta_{P_0}} n\hat{Q}_n^W(\theta)$ w.p.a.1 and $\hat{c}_W/n = o_p(1)$ with Theorem 4.1(b) restated as $d_H(\hat{\Theta}_n^W(\hat{c}_W), \Theta_{P_0}) = O_p((1 \vee \hat{c}_W)/n)^{1/2}$; cf. Proposition 2, p.110, of Rosen (2008) which sets $\hat{c}_W \to \infty$ and $\hat{c}_W/n = o(1)$. Since, in general, $\sup_{\theta\in\Theta_{P_0}} n\hat{Q}_n^W(\theta)$ is unknown, CHT, p.1254, suggests the choice $\hat{c}_W = o(\log(n))$ which yields a rate of convergence of $(\log(n)/n)^{1/2}$.

4.2 GEL

Let

$$\hat{\Theta}_n^\rho(\hat{c}_\rho) = \{\theta \in \Theta : n\hat{P}_n^\rho(\theta) \leq \hat{c}_\rho\} \tag{4.2}$$

where the profile GEL criterion $\hat{P}_n^\rho(\theta)$ is defined in (3.5). The GEL estimator of Θ_{P_0} (4.2) based on (3.4) is the solution to a saddle point problem and is described by the set estimator $\hat{\Theta}_n^\rho(\hat{c}_\rho)$ for some possibly data dependent $\hat{c}_\rho$.

REMARK 4.3. Write $\tilde{P}_n^\rho(\theta) = \inf_{\tau\in\mathbb{T}} \sup_{\lambda\in\hat{\Lambda}_n(\theta)} \tilde{P}_n^\rho(\theta,\lambda,\tau)$ and $\tilde{P}_n^{\rho,k}(\theta) = \inf_{\tau\in\mathbb{T}} \sup_{\lambda\in\hat{\Lambda}_n^k(\theta,\tau)} \tilde{P}_n^{\rho,k}(\theta,\lambda,\tau)$ $(k = a, b)$, see Ramrk 3.6, and define

$$\tilde{\Theta}_n^\rho(\hat{c}_\rho) = \{\theta \in \Theta : n\tilde{P}_n^\rho(\theta) \leq \hat{c}_\rho\} \tag{4.3}$$

and

$$\tilde{\Theta}_n^{\rho,k}(\hat{c}_\rho) = \{\theta \in \Theta : n\tilde{P}_n^{\rho,k}(\theta) \leq \hat{c}_\rho\}, (k = a, b). \tag{4.4}$$

Consequently, the set estimators $\hat{\Theta}_n^\rho(\hat{c}_\rho)$ (4.2), $\tilde{\Theta}_n^\rho(\hat{c}_\rho)$ (4.3) and $\tilde{\Theta}_n^{\rho,k}(\hat{c}_\rho)$, $(k = a, b)$, (4.4) based on the respective GEL criteria (3.4), (E.6), (E.1) and (E.3) evaluated using the same critical value $\hat{c}_\rho$ are identical given their equivalence to $\tilde{P}_n^\rho(\theta, \lambda, \tau)$ established in Appendix E.1.

Appendix C establishes the corresponding validity of Conditions C.1, p.1252, C.2, p.1253, and C.3, p.1255, of CHT under Assumptions A.1, A.2-GEL and A.3-A.5. Hence, a similar result to Theorem 4.1 for GMM may be stated on the consistency and rate of convergence of the GEL set estimators $\hat{\Theta}_n^\rho(\hat{c}_\rho)$ (4.2), $\tilde{\Theta}_n^\rho(\hat{c}_\rho)$ (4.3) and $\tilde{\Theta}_n^{\rho,k}(\hat{c}_\rho)$, $(k = a, b)$, (4.4) for the identified set Θ_{P_0} with some possibly data-dependent $\hat{c}_\rho$.

Theorem 4.2. *Let* $\hat{c}_\rho \geq q_n^\rho = \inf_{\theta\in\Theta} n\hat{P}_n^\rho(\theta)$, $\inf_{\theta\in\Theta} n\tilde{P}_n^{\rho,k}(\theta)$, $(k = a, b)$, *or* $\inf_{\theta\in\Theta} n\tilde{P}_n^\rho(\theta)$ *w.p.a.*1 *and* $\hat{c}_\rho = O_p(1)$. *Then, under Assumptions A.1, A.2–GEL and A.3-A.5,* **(a)** $\hat{\Theta}_n^\rho(\hat{c}_\rho)$, $\tilde{\Theta}_n^\rho(\hat{c}_\rho)$ *and* $\tilde{\Theta}_n^{\rho,k}(\hat{c}_\rho)$, $(k = a, b)$, *are consistent estimators of the identified set* Θ_{P_0}, *i.e.,* $d_H(\hat{\Theta}_n^\rho(\hat{c}_\rho), \Theta_{P_0})$, $d_H(\tilde{\Theta}_n^\rho(\hat{c}_\rho), \Theta_{P_0})$, $d_H(\tilde{\Theta}_n^{\rho,k}(\hat{c}_\rho), \Theta_{P_0})$, $(k = a, b)$, *are* $o_p(1)$; **(b)** $d_H(\hat{\Theta}_n^\rho(\hat{c}_\rho), \Theta_{P_0})$, $d_H(\tilde{\Theta}_n^\rho(\hat{c}_\rho), \Theta_{P_0})$, $d_H(\tilde{\Theta}_n^{\rho,k}(\hat{c}_\rho), \Theta_{P_0})$, $(k = a, b)$, *are* $O_p(n^{-1/2})$.

Remark 4.4. Theorem 4.2 is proved for the alternative GEL criterion $\tilde{P}_n^\rho(\theta, \lambda, \tau)$ (E.6) but also applies to $\hat{P}_n^\rho(\theta, \lambda)$ (3.4) and $\tilde{P}_n^{\rho,k}(\theta, \lambda, \tau)$, $(k = a, b)$, (E.1), (E.3), given their equivalence to $\tilde{P}_n^\rho(\theta, \lambda, \tau)$ established in Appendix E.1. Proofs analogous to those of Newey and Smith (2004) are developed to show these results. In particular, the GEL criterion function is shown to be first-order equivalent to the optimal GMM criterion (3.2) in the unconditional moment equality context and then Theorem 4.1 with population GMM weight matrix $\Omega(\theta)^{-1}$, $\theta \in \Theta$, is invoked.

Remark 4.5. Similarly to Remark 4.2 above, if $\hat{c}_\rho \geq \sup_{\theta\in\Theta_{P_0}} n\hat{P}_n^\rho(\theta)$, $\geq \sup_{\theta\in\Theta_{P_0}} n\tilde{P}_n^{\rho,k}(\theta)$,$(k = a, b)$, or $\geq \sup_{\theta\in\Theta_{P_0}} n\tilde{P}_n^\rho(\theta)$ w.p.a.1 and $\hat{c}_\rho/n = o_p(1)$, then $d_H(\hat{\Theta}_n^\rho(\hat{c}_\rho), \Theta_{P_0})$, $d_H(\tilde{\Theta}_n^{\rho,k}(\hat{c}_\rho), \Theta_{P_0})$, $(k = a, b)$, and $d_H(\tilde{\Theta}_n^\rho(\hat{c}_\rho), \Theta_{P_0})$ are $O_p((1 \vee \hat{c}_\rho)/n)^{1/2}$.

5 Confidence Region Estimation

Confidence regions for the identified set Θ_{P_0} and the true parameter value θ_0 are of particular interest. Section 5.1 constructs a conservative confidence region for the identified set Θ_{P_0}. Section 5.2 develops conservative GEL-based confidence regions for the true parameter value θ_0 similar to those of Rosen (2008, Section 4, pp.111-113).

5.1 Conservative Confidence Regions for Θ_{P_0}

A critical matter is a suitable choice for the possibly data dependent $\hat{c}_W$ or $\hat{c}_\rho$ respectively satisfying the hypotheses of Theorems 4.1 and 4.2 thereby ensuring that the GMM estimator $\hat{\Theta}_n^W(\hat{c}_W) = \{\theta \in \Theta : n\hat{Q}_n^W(\theta) \leq \hat{c}_W\}$, cf. (4.1), or GEL estimators $\hat{\Theta}_n^\rho(\hat{c}_\rho)$ (4.2), $\tilde{\Theta}_n^\rho(\hat{c}_\rho)$ (4.3) and $\tilde{\Theta}_n^{\rho,k}(\hat{c}_\rho)$ (4.4), $(k = a, b)$, of the identified set Θ_{P_0} possesses a confidence region property; see CHT section 3.3, pp.1256-1257. CHT section 4.2, pp.1265-1267, addresses this issue for moment inequalities when the GMM asymptotic weighting matrix $W(\theta)$ is diagonal; see CHT Condition M.2(e), p.1265.

Suppose b moment inequalities bind, i.e., $m^j(\theta) = 0$, $(j = 1, ..., b)$, and the remainder do not, i.e., $m^j(\theta) > 0$, $(j = b+1, ..., d_m)$, and $c = d_m - b$; note that b and thus c depend on θ. In principle, the critical value $\hat{c}_W$ describing the GMM confidence region estimator $\hat{\Theta}_n^W(\hat{c}_W) = \{\theta \in \Theta : n\hat{Q}_n^W(\theta) \leq \hat{c}_W\}$ would be obtained from consideration of the distribution of the limit quantity $\mathcal{C}^W = \sup_{\theta \in \Theta_{P_0}} \mathcal{C}^W(\theta)$ for the optimised GMM criterion $\sup_{\theta \in \Theta_{P_0}} n\hat{Q}_n^W(\theta)$, where $\mathcal{C}^W(\theta) = (v(\theta) - s(\theta))' W(\theta)(v(\theta) - s(\theta))$, $s(\theta) = \arg\min_{s_b \in \mathcal{R}_+^b, s_c \in \mathcal{R}^c} (v(\theta) - s)' W(\theta)(v(\theta) - s)$, $s = (s_b', s_c')'$ with s_b those b elements of s corresponding to the b binding moment inequalities and s_c the remainder. See Lemmas A.2 and A.3 in Appendix A together with the Proof of CHT Condition C.1(d) in Appendix B. Let $\hat{c}_W(1-\alpha)$ denote a consistent estimator of the $1-\alpha$ quantile $c_W(1-\alpha)$ of the limit quantity $\mathcal{C}^W$. Then $\hat{\Theta}_n(\hat{c}_W(1-\alpha))$ (4.2) defines asymptotically an $(1-\alpha)$ level confidence region for Θ_{P_0} as $\lim_{n\to\infty} \mathcal{P}\{\Theta_{P_0} \subseteq \hat{\Theta}_n(\hat{c}_W(1-\alpha))\} = 1-\alpha$ and $\hat{\Theta}_n(\hat{c}_W(1-\alpha))$ is a consistent estimator of Θ_{P_0} in Hausdorff distance at rate $n^{-1/2}$; see Theorem 4.1 and CHT, p.1266. Similar results may be stated for the GEL estimators $\hat{\Theta}_n^\rho(\hat{c}_\rho(1-\alpha))$ (4.2), $\hat{\Theta}_n^\rho(\hat{c}_\rho(1-\alpha))$ (4.3) and $\hat{\Theta}_n^{\rho,k}(\hat{c}_\rho(1-\alpha))$ (4.4), $(k = a, b)$, given the limiting relationship of the GEL criteria to the GMM-CUE criterion $n\hat{Q}_n^W(\theta)$ when $W_n(\theta) = \hat{\Omega}_n(\theta)^{-1}$; see Theorem 4.2 and Appendix C.

Remark 5.1. To the best of our knowledge, no formal results yet exist establishing the asymptotic validity of sub-sampling methods for approximating the distribution of the limit GMM quantity $\mathcal{C}^W$ with a non-diagonal GMM weight matrix $W(\theta)$, in particular, $\Omega(\theta)^{-1}$, required for simulating the GEL confidence region estimator critical value $\hat{c}_\rho$. Cf. CHT section 3.4, pp.1257-1258.

To deal with the difficulty outlined in Remark 5.1, a simple valid but conservative confidence region estimator for the identified set Θ_{P_0} is now described. The difficulty is easily circumvented by replacing the optimal GMM slackness parameter estimator $\hat{t}_n(\theta)$ by $[\hat{m}_n(\theta)]_-$, i.e., the estimator that solves a GMM criterion with diagonal weight-matrix as metric, thereby bounding the GMM criterion $n\hat{Q}_n^W(\theta)$ (3.1) above; cf. CHT Condition M.2(e), p.1265. Let

$$\underline{\hat{Q}}_n^W(\theta) = [\hat{m}_n(\theta)]_-' W_n(\theta)[\hat{m}_n(\theta)]_-. \tag{5.1}$$

Then, by definition,

$$\hat{Q}_n^W(\theta) \leq \underline{\hat{Q}}_n^W(\theta)$$

for all n and $\theta \in \Theta$.

Remark 5.2. The population counterpart $\underline{Q}^W(\theta)$ to the bounding GMM criterion $\underline{\hat{Q}}_n^W(\theta)$ (5.1) is defined by $\underline{Q}^W(\theta) = [m(\theta)]_-' W(\theta)[m(\theta)]_- = \left\|[m(\theta)]_-' W(\theta)^{1/2}\right\|^2$; cf. $Q^W(\theta)$ (3.3).

The Proofs of CHT Conditions C.4, p.1256, and C.5, p.1257, in Appendix B establish the limiting behaviour of the scaled bounding GMM criterion $n\underline{\hat{Q}}_n^W(\theta)$ (5.1); cf. CHT Proof of Theorem 4.2 Steps 4 and 5, pp.1279-1280. The Proof of CHT Condition C.4 in Appendix B, in particular, see (B.2) and (B.3) of Appendix B, establishes that the limit $\underline{\mathcal{C}}^W$ of $\underline{\mathcal{C}}_n^W = \sup_{\theta \in \Theta_{P_0}} n\underline{\hat{Q}}_n^W(\theta)$ is described by

$$\underline{\mathcal{C}}^W = \sup_{\theta \in \Theta_{P_0}} \|[v(\theta) + \xi(\theta)]_-\|^2_{W(\theta)}$$

where $\xi^j(\theta) = 0$ if $m^j(\theta) = 0$, $(j = 1, ..., b)$, and $\xi^j(\theta) = \infty$ if $m^j(\theta) > 0$, $(j = b+1, ..., d_m)$, for $\theta \in \Theta_{P_0}$.

Correspondingly the $1-\alpha$ quantile $\underline{c}_W(1-\alpha)$ of the limit $\underline{\mathcal{C}}^W$ of the scaled bounding GMM criterion $n\underline{\hat{Q}}_n^W(\theta)$ (5.1) satisfies

$$\mathcal{P}\{\underline{\mathcal{C}}^W \leq \underline{c}_W(1-\alpha)\} = 1-\alpha,$$

i.e., $\lim_{n\to\infty} \mathcal{P}\{\sup_{\theta\in\Theta_{P_0}} n\underline{Q}_n^W(\theta) \leq \underline{c}_W(1-\alpha)\} = 1-\alpha$. It is then immediate that $\lim_{n\to\infty} \mathcal{P}\{\sup_{\theta\in\Theta_{P_0}} n\hat{Q}_n^W(\theta) \leq \underline{c}_W(1-\alpha)\} \geq \lim_{n\to\infty} \mathcal{P}\{\sup_{\theta\in\Theta_{P_0}} n\underline{\hat{Q}}_n^W(\theta) \leq \underline{c}_W(1-\alpha)\}$. Hence the asymptotic level of the confidence region $\hat{\Theta}_n^W(\underline{c}_W(1-\alpha))$ (4.1) is bounded below by $1-\alpha$, i.e.,

$$\lim_{n\to\infty} \mathcal{P}\{\Theta_{P_0} \subseteq \hat{\Theta}_n^W(\underline{c}_W(1-\alpha))\} \geq 1-\alpha. \tag{5.2}$$

REMARK 5.3. To implement the confidence region estimator $\hat{\Theta}_n^W(\underline{c}_W(1-\alpha))$ (4.1) requires a consistent estimate of the quantile $\underline{c}_W(1-\alpha)$ of the limit $\underline{\mathcal{C}}^W$ of $\underline{\mathcal{C}}_n^W = \sup_{\theta\in\Theta_{P_0}} n\hat{\underline{Q}}_n^W(\theta)$. A simulation procedure similar to that outlined in CHT Remarks 4.2, pp.1263-1264, and 4.5, p.1267, suffices. In particular, let z_i^*, $(i=1,...,n)$, denote n `i.i.d.` draws from the standard normal $N(0,1)$ distribution. Thus, the process $v_n^*(\theta) = n^{-1/2}\sum_{i=1}^n m_i(\theta)z_i^*$ is zero-mean Gaussian with covariance function $\sum_{i=1}^n m_i(\theta_a)m_i(\theta_b)'/n$. Define $\hat{\xi}_n^j(\theta) = 0$ if $\hat{m}_n^j(\theta) \leq c_j((\log n)/n)^{1/2}$ and ∞ if $\hat{m}_n^j(\theta) > c_j((\log n)/n)^{1/2}$ for some positive constants $c_j > 0$, $(j=1,...,d_m)$. Also let $\hat{\Theta}_n$ denote a consistent estimator of Θ_{P_0}; see section 4. Quantiles of the limit $\underline{\mathcal{C}}^W$ can then be estimated by simulation from the distribution of $\hat{\underline{C}}_n^{W*} = \sup_{\theta\in\hat{\Theta}_n} \hat{\underline{Q}}_n^{W*}(\theta)$ where $\hat{\underline{Q}}_n^{W*}(\theta) = \|[v_n^*(\theta) + \hat{\xi}_n(\theta)]_-\|_{W_n(\theta)}$.

5.2 Confidence Regions for θ_0

This section is concerned with GMM and GEL estimation of confidence regions for the true parameter value θ_0. Of central interest here is the optimal GMM criterion in the unconditional moment equality context, i.e., $\hat{Q}_n^{\Omega^{-1}}(\theta)$ (3.2) when the GMM metric $W_n(\theta) = \hat{\Omega}_n(\theta)^{-1}$. CHT section 5, pp.1267-1270, analyses the issue with an asymptotically diagonal GMM weight matrix whereas Rosen (2008) deals with the optimal GMM criterion.To ease the notational burden the optimal GMM metric Ω^{-1} is omitted in the following discussion.

Let $b(\theta)$ denote the number of binding moments for $\theta\in\Theta_{P_0}$. Define $c(\theta) = d_m - b(\theta)$, $\theta\in\Theta_{P_0}$. Without loss of generality also let $m^j(\theta) = 0$, $(j=1,...,b(\theta))$, and $m^j(\theta) > 0$, $(j=b(\theta)+1,...,d_m)$, $\theta\in\Theta_{P_0}$.

By Lemma A.3 in Appendix A

$$\begin{aligned} n\hat{Q}_n(\theta) &= \inf_{s_b\in\mathcal{R}_+^b, s_c\in\mathcal{R}^c} (v(\theta)-s)'\Omega(\theta)^{-1}(v(\theta)-s) + o_p(1) \\ &= (v(\theta)-s(\theta))'\Omega(\theta)^{-1}(v(\theta)-s(\theta)) + o_p(1) \end{aligned}$$

uniformly $\theta\in\Theta_{P_0}$. Therefore, cf. Rosen (2008, Proposition 3, p.110), uniformly $\theta\in\Theta_{P_0}$,

$$\lim_{n\to\infty}\mathcal{P}\{n\hat{Q}_n(\theta) > c\} = \sum_{j=1}^{b(\theta)} w(b(\theta), b(\theta)-j), \Omega(\theta))\mathcal{P}\{\chi_j^2 > c\}, \tag{5.3}$$

a weighted chi-bar square distribution, where χ_j^2, $(j=1,...,b(\theta))$, denote independent chi-square random variates with j degrees of freedom respectively. The weights $w(b(\theta), b(\theta)-j), \Omega(\theta))$, $(j=1,...,b(\theta))$, in (5.3) are defined in Kudo (1963) and Wolak (1987) and correspond to the probability that exactly j of the $b(\theta)$ binding inequality constraints bind, i.e., $\mathcal{P}\{s(\theta)$ has j zero components$\}$, $(j=1,...,b(\theta))$; e.g., ${}^{b(\theta)}C_j/2^{b(\theta)}$ if $\Omega(\theta)$ is diagonal. See the discussion in Rosen (2008, p.111).

Clearly the GMM statistic (3.2) $n\hat{Q}_n(\theta)$ (3.2) is asymptotically non-pivotal. As noted in Rosen (2008), if both $b(\theta)$ and $\Omega(\theta)$ were known, the limiting distribution (5.3) could easily be simulated with a valid confidence region for the true value θ_0 obtained by inversion of the non-rejection region $\{n\hat{Q}_n(\theta)\leq c\}$ with c determined to deliver the desired confidence level from (5.3). The limiting distribution (5.3), however, is discontinuous in $b(\theta)$ rendering an estimator for this limiting distribution based on simulation after substitution of consistent estimators $\hat{b}_n(\theta)$ and $\hat{\Omega}_n(\theta)$ for $b_n(\theta)$ and $\Omega(\theta)$ respectively inconsistent. Consequently, Rosen (2008, p.111) suggests using a least favourable asymptotic distribution approach based on an estimated upper bound for $b(\theta)$. In particular, define $\hat{b}_n(\theta)=\sum_{j=1}^{d_m}1[\hat{m}_n^j(\theta)<C((\log n)/n)^{1/2}]$ for some constant $C>0$. Then, since $\lim_{n\rightarrow\infty}\mathcal{P}\{\hat{b}_n(\theta)=b(\theta)\}=1$, uniformly $\theta\in\Theta_{P_0}$, see CHT Remark 4.2, p.1267, Rosen (2008) proposes the upper bound estimator $\hat{b}_n^{\sup}=\sup_{\theta\in\hat{\Theta}_n(\hat{c})}\hat{b}_n(\theta)$ where $\hat{\Theta}_n(\hat{c})$ is the consistent GMM identified set estimator (4.1) with level $\hat{c}$ satisfying Theorem 4.1. Let $b^{\sup}=\sup_{\theta\in\Theta_{P_0}}b(\theta)$. Then

$$\sup_{\theta\in\Theta_{P_0}}\lim_{n\rightarrow\infty}\mathcal{P}\{n\hat{Q}_n(\theta)>c\}\leq\frac{1}{2}\mathcal{P}\{\chi^2_{b^{\sup}}>c\}+\frac{1}{2}\mathcal{P}\{\chi^2_{b^{\sup}-1}>c\};$$

see Rosen (2008, Corollary 1, p.113). Therefore, setting c such that

$$\alpha=\frac{1}{2}\mathcal{P}\{\chi^2_{b^{\sup}}>c\}+\frac{1}{2}\mathcal{P}\{\chi^2_{b^{\sup}-1}>c\},$$

a conservative $1-\alpha$ level confidence region for θ_0 is given by

$$\begin{aligned}\inf_{\theta\in\Theta_{P_0}}\lim_{n\rightarrow\infty}\mathcal{P}\{n\hat{Q}_n(\theta)&\leq c\}=1-\sup_{\theta\in\Theta_{P_0}}\lim_{n\rightarrow\infty}\mathcal{P}\{n\hat{Q}_n(\theta)>c\}\\&\geq 1-\alpha.\end{aligned}$$

See the associated discussion in Rosen (2008, section 4, pp.111-113).

Remark 5.4. The various scaled optimised GEL criteria $2n\hat{P}_n^\rho(\theta)$ (3.4), $2n\tilde{P}_n^\rho(\theta,\tilde{\lambda}_n(\theta),\tilde{\tau}_n(\theta))$ (E.6) and $2n\tilde{P}_n^{\rho,k}(\theta,\tilde{\lambda}_n(\theta),\tilde{\tau}_n(\theta))$, $(k=a,b)$, (E.1), (E.3) are asymptotically equivalent to the GMM criterion $n\hat{Q}_n(\theta)$ (3.2), uniformly $\theta\in\Theta_{P_0}$; see Lemma C.4 in Appendix C. Therefore, valid conservative GEL confidence regions for θ_0 asymptotically equivalent to that defined in (5.4) are given by substitution of these GEL criteria for $n\hat{Q}_n(\theta)$ in (5.4) based on the respective consistent GEL identified set estimators $\hat{\Theta}_n^\rho(\hat{c}_\rho)$ (4.2), $\hat{\Theta}_n^\rho(\hat{c}_\rho)$ (4.3) or $\hat{\Theta}_n^{\rho,k}(\hat{c}_\rho)$, $(k=a,b)$, (4.4), in place of the GMM identified set estimator $\hat{\Theta}_n(\hat{c})$ (4.1).

6 Simulation Evidence

This section reports the results from a simulation study to assess the performance of some of the confidence region estimators for the identified set Θ_{P_0} based on the GMM and GEL statistics developed in Section 5.1 for the identified set in a nonlinear interval conditional mean regression model.

6.1 Experimental Design

The nonlinear conditional mean regression for the latent scalar variable y given the scalar covariate x is described by

$$y = x^{\theta_0} + u$$

where $u|x \sim N(0,1)$, x is uniformly distributed on the unit interval $[0,1]$ and $\theta_0 = 1$ is the true value of the scalar parameter θ.

The regressand y is only partially observed according to the interval observation rule

$$y_1 \leq y \leq y_2$$

with $y_1 = y - \omega_1 x^2$ and $y_2 = y + \omega_2 x$ observed where ω_1, $\omega_2 \geq 0$. Hence

$$\mathbb{E}_{P_0}[y_1|x] \leq x^{\theta_0} \leq \mathbb{E}_{P_0}[y_2|x] \text{ a.s. } x.$$

and

$$\mathbb{E}_{P_0}[y_1 x] \leq \mathbb{E}_{P_0}[x^{\theta_0+1}] \leq \mathbb{E}_{P_0}[y_2 x]$$

Defining the moment indicator vector $m(z,\theta) = \left(-(y_1 - x^\theta)x, (y_2 - x^\theta)x\right)'$,

$$\begin{aligned} \mathbb{E}_{P_0}[m(z,\theta)] &= \begin{pmatrix} -(E[x] - \omega_1 E[x^3] - E[x^{\theta+1}] \\ E[x] + \omega_2 E[x^2] - E[x^{\theta+1}] \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{2} + \frac{\omega_1}{4} + \frac{1}{\theta+2} \\ \frac{1}{2} + \frac{\omega_2}{3} - \frac{1}{\theta+2} \end{pmatrix}. \end{aligned}$$

Therefore, with the moment inequaility contraint $\mathbb{E}_{P_0}[m(z,\theta)] \geq 0$, the identified set Θ_{P_0} is given by the interval

$$\Theta_{P_0} = \left[-\frac{4\omega_2}{3+2\omega_2}, \frac{\omega_1}{1-\omega_1/2}\right].$$

To obtain the moment matrix $\Omega(\theta_0) = \mathbb{E}_{P_0}[m(z,\theta_0)m(z,\theta_0)']$, note that $m^1(z,\theta_0) = -ux + \omega_1 x^3$ and $m^2(z,\theta_0) = ux + \omega_2 x^2$. Hence

$$\Omega(\theta_0) = \begin{pmatrix} 1/3 + \omega_1^2/7 & \omega_1\omega_2/6 - 1/3 \\ \omega_1\omega_2/6 - 1/3 & 1/3 + \omega_2^2/5 \end{pmatrix}$$

which is diagonal when $\omega_1\omega_2 = 2$. The experiments consider the diagonal $\Omega(\theta_0)$ case throughout with values for ω_1 and ω_2

$$\omega_1 = \frac{2}{3}, \omega_2 = 3.$$

Hence, the resultant identified set

$$\Theta_{P_0} = [-\frac{4}{3}, 1]$$

and moment matrix

$$\Omega(\theta_0) = \begin{pmatrix} 8/3 & 0 \\ 0 & 32/15 \end{pmatrix}.$$

REMARK 6.1. GMM with a diagonal metric is thus studied under the most favourable conditions as compared to GEL. When $\theta = \theta_0$ the GEL criterion function is asymptotically equivalent to that GMM but, importantly, is not generally for $\theta \in \Theta_{P_0}/\{\theta_0\}$. In this example the moment matrix $\Omega(\theta)$ cannot be diagonal everywhere on Θ_0 even when Θ_{P_0} is a singleton, i.e., $\omega_1 = \omega_2 = 0$, as is likely to be the case in practice for moment inequality models when the moment matrix cannot be diagonal everywhere on Θ_0.

Experimental data are generated as random samples of size $n = 50$, 100, 500 and 1000 from the joint distribution of (y, x). Each simulation experiment comprises 500 replications.

6.2 Criteria

The criteria $n\hat{Q}^j_n(\theta)$, (j =EL, ET, CUE, GMM), are considered where $\hat{Q}^j_n(\theta) = \hat{P}^\rho_n(\theta)$ (3.5),, (j =EL, ET, CUE), with $\rho(v) = \log(1-v)$ [EL] empirical likelihood, $\rho(v) = \exp(-v) - 1$ [ET] exponential tilting and $\rho(v) = -1/2[(1+v)^2 - 1]$ [CUE] continuous updating estimator respectively with $\hat{Q}^{\text{GMM}}_n(\theta) = \hat{Q}^W_n(\theta)$ [GMM] the GMM objective function with the metric $W_n(\theta)$ diagonal with diagonal elements those of `the efficient` moment equality metric $\hat{\Omega}_n(\theta)^{-1}$ thus mimicing the approach in CHT section 4.2, pp.1265-67.

Each criterion is evaluated across the grid $\Theta_n = \{-9, ..., 0, ..., 10u\}$.[1]

[1]To reduce computer processing time the grid spacing was increased for values of θ in Θ_n further away from the bounds defining Θ_{P_0}, though always maintaining $d_H(\Theta_{P_0}, \Theta_n) = O(1/n)$.

6.3 Level

The definitions of the level c_n^j used to evaluate the coverage probability $\mathcal{P}\{n\hat{Q}_n^j(\theta) \leq c_n^j\}$ are

$$c_n^j \in \{\tilde{c}_n^j, \log\log n/2, \log n/2, \sqrt{n}/2\},$$

where $\tilde{c}_n^j = \inf_{\Theta_n} n\hat{Q}_n^j(\theta) + 0.001$, ($j$ =EL, ET, CUE, GMM), with the grid Θ_n given in section 6.2.

Remark 6.2. Hence, the moment function satisfies the degeneracy condition CHT Condition C.3, p.1255, see Appendix B, rendering each set estimator a consistent estimator of the identified set Θ_{P_0}; see section 4.

6.4 Results

Figures 1-4 about here

Figures 1-4 indicate that the coverage probability $\mathcal{P}\{n\hat{Q}_n^j(\theta) \leq c_n^j\}$ converges to 1 for $\theta \in \Theta_{P_0}$ for $\tilde{c}_n^j > \inf_{\theta\in\Theta_n} n\hat{Q}_n^j(\theta)$ verifying the consistency results for GMM and GEL of Theorems 4.1 and 4.2 and with $c_n^j/n \to 0$; cf. Remarks 4.2 and 4.5.

With faster rates of growth for the level c_n^j the coverage probability $\mathcal{P}\{n\hat{Q}_n^j(\theta) \leq c_n^j\}$ is closer to 1 for more $\theta \in \Theta_{P_0}$. However, there is also a tendence for an increase in the coverage probability $\mathcal{P}\{n\hat{Q}_n^j(\theta) \leq c_n^j\}$ for $\theta \in \Theta/\Theta_{P_0}$; this is especially marked in the smaller samples, see Figures 1 and 2 with $n = 50$ and 100 respectively, for EL and ET as compared with CUE and GMM when $\theta > 1$, the upper bound of the identified set Θ_{P_0}. The reverse appears to be the case for $\theta < -4/3$, the lower bound of Θ_{P_0}. CUE exhibits very hign coverage probabilities for $\theta < -4/3$; to a lesser extent a similar occurence is observed for GMM whereas the coverage probabilities for both EL and ET are close to zero for $\theta < -4/3$. The differences in coverage probabilities when $\theta > 1$ for EL and ET as compared to CUE and GMM are much less pronounced than those for $\theta < -4/3$. For the larger sample sizes $n = 500$ and 1000 the coverage probabilities are very similar for EL, ET, CUE, GMM for both $\theta \in \Theta_{P_0}$ and $\theta > 1$ but CUE and GMM continue to display very poor properties, especially CUE; see Figures 3 and 4.

To study the properties of the conservative inferential procedures described in section 5.1 based on the GEL confidence region estimator $\hat{\Theta}_n^\rho(\hat{c}_\rho(1-\alpha))$ (4.2), the quantiles of the bounding statistic $\sup_{\theta\in\Theta_{P_0}} \underline{\hat{Q}}_n^{\Omega^{-1}}(\theta)$ (5.1) are required where $\underline{\hat{Q}}_n^{\Omega^{-1}}(\theta) = [\hat{m}_n(\theta)]_-'\hat{\Omega}_n(\theta)^{-1}[\hat{m}_n(\theta)]_-$. As described in Remark 5.3 of section 5.1 suitable estimates are provided by simulation of the quantiles from $\underline{\hat{C}}_n^{\Omega^{-1}*} = \sup_{\theta\in\hat{\Theta}_n} \underline{\hat{Q}}_n^{\Omega^{-1}*}(\theta)$ where $\underline{\hat{Q}}_n^{\Omega^{-1}*}(\theta) = \|[v_n^*(\theta) + \hat{\xi}_n(\theta)]_-\|_{\hat{\Omega}_n(\theta)^{-1}}$. Note that if interest is in ascertaining whether a hypothesised identified set $\Theta_{P_0}^0$ is credible by examining whether $\Theta_{P_0}^0 \subseteq \hat{\Theta}_n^{\rho}(\underline{c}_{\Omega^{-1}}^*(1-\alpha))$, where $\underline{c}_{\Omega^{-1}}^*(1-\alpha)$ is the $1-\alpha$ quantile of $\underline{\hat{C}}_n^{\Omega^{-1}*}$, it is only necessary to take the supremum of $\underline{\hat{Q}}_n^{\Omega^{-1}*}(\theta)$ over $\Theta_{P_0}^0$ rather than an estimator $\hat{\Theta}_n$ which greatly simplifies inference and also does not add any further sampling uncertainty to the estimate of the distribution of the limit variate $\underline{\mathcal{C}}^{\Omega^{-1}}$; see below Remark 5.2.

Figure 5 about here

Figure 5 displays the estimated quantiles of $\sup_{\theta\in\Theta_{P_0}} n\hat{Q}_n(\theta)^j$, ($j$ =ET, EL, CUE), together with those of the bounding variate $\sup_{\theta\in\Theta_{P_0}} \underline{\hat{Q}}_n^{\Omega^{-1}*}(\theta)$ (5.1). As expected these are bounded below by those of $\sup_{\theta\in\Theta_{P_0}} \underline{\hat{Q}}_n^{\Omega^{-1}*}(\theta)$ with the bound rather conservative for the range $1-\alpha \in (0.9, 1.0)$ typically used in practice for inference. The ET, EL and CUE quantiles are almost identical at all sample sizes which corroborates empirically their first order equivalence on Θ_{P_0} detailed in Theorem 4.2. The corresponding quantile estimates for $\sup_{\theta\in\Theta_0} \hat{Q}_n^{\mathrm{GMM}}(\theta)$ are also plotted for comparison although the quantile lower bound does not apply to GMM.

7 Concluding Remarks

This paper examines the properties of GEL methods for the estimation of the identified set in models specified by unconditional moment inequality constraints. The paper extends the results for GMM estimation in CHT section 4, pp.1261-1267, to permit a nondiagonal weight matrix in the GMM criterion, in particular, the inverse of the moment variance matrix, the optimal GMM metric appropriate for moment equality conditions. Unlike the moment equality context, this extension of GMM to GEL is relatively nontrivial. Analogously to moment equality condition models, an asymptotic equivalence exists between various scaled optimised GEL criteria and that for GMM with optimal moment equality weight matrix. Consequently, similarly to CHT, conditions are provided for consistent GEL estimation of the identified set at the parametric rate $n^{1/2}$. When the moment matrix is nondiagonal on the identified set the limit of the scaled optimised GEL statistic differs from that for GMM with diagonal weight matrix which the case studied in CHT section 4, pp.1261-1267. To the best of our knowledge there are, as yet, no results for the asymptotic validity of a bootstrap or sub-sampling approximation to the limiting distribution of these statistics; cf. the application of CHT section 3.4, pp.1257-1258, to GMM with a diagonal weight matrix given in CHT section 4, pp.1261-1267. A conservative confidence region estimator for the identified set is therefore developed. The GMM criterion with non-diagonal weight matrix may be bounded above by a statistic the limit of which can be approximated using a resampling method similar to that described in CHT Remarks 4.2, pp.1263-1264, and 4.5, p.1267, for GMM with a diagonal weight matrix. Conservative GMM and GEL confidence region estimators for the true parameter, cf. Rosen (2008), are also described. A simulation study for a nonlinear interval nonlinear conditional mean regression model corroborates the main theoretical results of the paper with favourable small sample properties for conservative EL and ET confidence region estimators for the identified set.

Appendix

The argument θ is suppressed for expositional simplicity throughout the Appendices where there is no possibility of confusion.

Throughout the Appendix, C will denote a generic positive constant that may be different in different uses with CS, M and T the Cauchy-Schwarz, Markov and triangle inequalities respectively. In addition UWL is a uniform weak law of large numbers; CMT the continuous mapping theorem; w.p.a.1 "with probability approaching one" and $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ the minimum and maximum eigenvalues respectively of $\cdot$.

The following convention is employed. $\mathbb{E}_{P_0}[m^j(z,\theta)] < 0$, $(j = 1, ..., a)$, $\mathbb{E}_{P_0}[m^j(z,\theta)] = 0$, $(j = a+1, ..., a+b)$, and $\mathbb{E}_{P_0}[m^j(z,\theta)] > 0$, $(j = a+b+1, ..., d_m)$. Defining $c = d_m - a - b$, a, b and thus c depend on θ. Vectors are correspondingly partitioned, e.g., $s = (s_a', s_b', s_c')'$ such that s_a corresponds to $\mathbb{E}_{P_0}[m^j(z,\theta)] < 0$, $(j = 1, ..., a)$, i.e., those a elements of s for which (2.1) is false, s_b to $\mathbb{E}_{P_0}[m^j(z,\theta)] = 0$, $(j = a+1, ..., a+b)$, i.e., those b elements of s corresponding to the b binding moment inequalities and s_c to $\mathbb{E}_{P_0}[m^j(z,\theta)] > 0$, $(j = a+b+1, ..., d_m)$, i.e., the remainder.

Let $\Lambda_n = \{\lambda \in \mathcal{R}^{d_m} : \|\lambda\| \leq Cn^{-1/2}\}$. Also let $\Theta_{P_0}^{-\epsilon} = \{\theta \in \Theta : d(\theta, \Theta \backslash \Theta_{P_0}) \geq \epsilon\}$ where $\epsilon > 0$. A closed ball of radius $\delta > 0$ is denoted by $\mathcal{B}_\delta = \{\theta \in \mathcal{R}^{d_\theta} : \|\theta\| \leq \delta\}$. Recall Θ' is a neighbourhood of Θ in $\mathcal{R}^{d_\theta}$.

Recall the GMM sample criterion (3.1) $\hat{Q}_n^W(\theta) = \inf_{t \geq 0} \|\hat{m}_n(\theta) - t\|^2_{W_n(\theta)}$. Also define $\bar{Q}_n^W(\theta) = \inf_{t \geq 0} \|\hat{m}_n(\theta) - t\|^2_{W(\theta)}$ and $nQ_n^W(\theta) = \inf_{s_b \in \mathcal{R}_+^b, s_c \in \mathcal{R}^c} \|v_n(\theta) - s\|^2_{W_n(\theta)}$. Recall the corresponding GMM population criterion $Q^W(\theta) = \inf_{t \geq 0} \|m(\theta) - t\|^2_{W(\theta)}$, $\theta \in \Theta$, where $m(\theta) = \mathbb{E}_{P_0}[m(z,\theta)]$.

Define

$$\hat{t}_n(\theta) = \arg\min_{t\geq 0} \|\hat{m}_n(\theta) - t\|^2_{W_n(\theta)},$$

i.e., $\hat{Q}_n^W(\theta) = \left\|\hat{m}_n(\theta) - \hat{t}_n(\theta)\right\|^2_{W_n(\theta)}$.

Appendix A: Preliminary Lemmas

To simplify the Proofs for CHT Conditions C.1, p.1252, and C.2, p.1253, Lemmas A.1 and A.2 show that the weighting matrix $W_n(\theta)$ in the GMM criterion $\hat{Q}_n^W(\theta)$ (3.1) may be replaced by $W(\theta)$ w.p.a.1 uniformly $\theta \in \Theta$, i.e., $n\hat{Q}_n^W(\theta) = nQ_n^W(\theta) + o_p(1)$ uniformly $\theta \in \Theta$.

Lemma A.1. *Let Assumptions A.1, A.2-GMM and A.3 be satisfied. Then*

$$\sup_{\theta\in\Theta_{P_0}} \left|n\hat{Q}_n^W(\theta) - nQ_n^W(\theta)\right| = o_p(1).$$

Proof. The proof closely follows that for Proposition 3, pp.115-116, in Rosen (2008). Consider

$$\begin{aligned} n\hat{Q}_n^W(\theta) &= \inf_{t\geq 0} n \|\hat{m}_n(\theta) - t\|^2_{W_n(\theta)} \\ &= \inf_{s\geq -n^{1/2}m(\theta)} \|v_n(\theta) - s\|^2_{W_n(\theta)} \end{aligned}$$

where $s = n^{1/2}(t - m(\theta))$. Write $\hat{s}_n(\theta) = \arg\min_{s\geq -n^{1/2}m(\theta)} \|v_n(\theta) - s\|^2_{W_n(\theta)}$.

Suppose $\theta \in \Theta_{P_0}$. Thus, s_a is empty, i.e., $a = 0$; also $c = d_m - b$ and $m(\theta) \geq 0$, $\theta \in \Theta_{P_0}$. Let $m_c(\theta) = (m^{b+1}(\theta), ..., m^{d_m}(\theta))'$. In this case

$$\begin{aligned} n\hat{Q}_n^W(\theta) &= \inf_{s_b\geq 0, s_c\geq -n^{1/2}m_c(\theta)} \|v_n(\theta) - s\|^2_{W_n(\theta)} \\ &= \|v_n(\theta) - \hat{s}_n(\theta)\|^2_{W_n(\theta)} \end{aligned}$$

with solution $\hat{s}_n(\theta) = (\hat{s}_{bn}(\theta)', \hat{s}_{cn}(\theta)')' = \arg\min_{s_b\geq 0, s_c\geq -n^{1/2}m_c(\theta)} \|v_n(\theta) - s\|^2_{W_n(\theta)}$. Now

$$\begin{aligned} nQ_n^W(\theta) &= \inf_{s_b\in\mathcal{R}^b_+, s_c\in\mathcal{R}^c} \|v_n(\theta) - s\|^2_{W_n(\theta)} \\ &= \|v_n(\theta) - s_n(\theta)\|^2_{W_n(\theta)} \end{aligned}$$

with solution $s_n(\theta) = (s_{bn}(\theta)', s_{cn}(\theta)')' = \arg\min_{s_b\in\mathcal{R}^b_+, s_c\in\mathcal{R}^c} \|v_n(\theta) - s\|^2_{W_n(\theta)}$. Note that $s_{cn}(\theta) = v_{cn}(\theta)$, $\theta \in \Theta_{P_0}$, and thus, from Lemma 1, p.115, of Rosen (2008), w.p.a.1

$nQ_n^W(\theta) = v_{bn}(\theta)'(W_n^{bb}(\theta))^{-1}v_{bn}(\theta)$ where $W_n^{bb}(\theta)$ denotes the top left hand $b \times b$ submatrix of $W_n(\theta)^{-1}$.

To show that $\sup_{\theta\in\Theta_{P_0}} \left|n\hat{Q}_n^W(\theta) - nQ_n^W(\theta)\right| = o_p(1)$, i.e., $\sup_{\theta\in\Theta_{P_0}} \|\hat{s}_n(\theta) - s_n(\theta)\| = o_p(1)$, it is only necessary to demonstrate

$$\sup_{\theta\in\Theta_{P_0}} \|\hat{s}_{cn}(\theta) - s_{cn}(\theta)\| = o_p(1)$$

or $\sup_{\theta\in\Theta_{P_0}} \|\hat{s}_{cn}(\theta) - v_{cn}(\theta)\| = o_p(1)$. Now, since v_n is P-Donsker, $\sup_{\theta\in\Theta} \|v_n(\theta)\| = O_p(1)$ by Assumption A.3, i.e., for any ε, $\delta > 0$, there exists $N(\varepsilon,\delta)$ such that, for all $n > N(\varepsilon,\delta)$, $\mathcal{P}\{\sup_{\theta\in\Theta_{P_0}} \|v_n(\theta)\| < \varepsilon\} > 1-\delta$. Choose $\varepsilon = \max_j \sup_{\theta\in\Theta_{P_0}} m^j(\theta)$ such that for all $n > N(\varepsilon,\delta)$

$$\mathcal{P}\{\sup_{\theta\in\Theta_{P_0}} \|v_n(\theta)\| < \max_j \sup_{\theta\in\Theta_{P_0}} m^j(\theta)\} > 1-\delta.$$

In particular, for all $n > N(\varepsilon,\delta)$, with probability at least $1-\delta$, $\sup_{\theta\in\Theta_{P_0}} |v_n^j(\theta)| < \max_j \sup_{\theta\in\Theta} m^j(\theta)$ and, thus, $\hat{s}_n^j(\theta) = v_n^j(\theta)$, $(j = b+1,...,d_m)$, uniformly $\theta \in \Theta_{P_0}$, i.e.,

$$\mathcal{P}\{\sup_{\theta\in\Theta_{P_0}} \|\hat{s}_{cn}(\theta) - v_{cn}(\theta)\| = 0\} > 1-\delta.$$

Therefore,

$$n\hat{Q}_n^W(\theta) = nQ_n^W(\theta) + o_p(1)$$

uniformly $\theta \in \Theta_{P_0}$.■

Lemma A.2. *Let Assumptions A.1, A.2-GMM and A.3 be satisfied. Then*

$$\inf_{\theta\in\Theta/\Theta_{P_0}} n\hat{Q}_n^W(\theta) \xrightarrow{p} \infty.$$

Proof. Let $\theta \in \Theta/\Theta_{P_0}$. In this case, since s_a is no longer empty, define $m_a(\theta) = (m^1(\theta),...,m^a(\theta))'$. Hence,

$$\begin{aligned} n\hat{Q}_n^W(\theta) &= \inf_{s\geq -n^{1/2}m(\theta)} \|v_n(\theta) - s\|^2_{W_n(\theta)} \\ &= \inf_{s_a\geq -n^{1/2}m_a(\theta), s_b\geq 0, s_c\geq -n^{1/2}m_c(\theta)} \|v_n(\theta) - s\|^2_{W_n(\theta)} \\ &\geq \inf_{s_a\geq -n^{1/2}m_a(\theta), s_b\in\mathcal{R}^b_+, s_c\in\mathcal{R}^c} \|v_n(\theta) - s\|^2_{W_n(\theta)} \\ &\geq \inf_{s_a\geq -n^{1/2}m_a(\theta)} \|v_{an}(\theta) - s_a\|^2_{(W_n^{aa}(\theta))^{-1}} \end{aligned}$$

w.p.a.1 where $W_n^{aa}(\theta)$ denotes the $a \times a$ top left hand sub matrix of $W_n(\theta)^{-1}$ corresponding to $m_a(\theta)$; see Lemma 1, p.115, of Rosen (2008). Now $\sup_{\theta\in\Theta} \|v_n(\theta)\| = O_p(1)$ by Assumption A.3. Thus, since $-n^{1/2}m_a(\theta) \to \infty$ if $\theta \in \Theta/\Theta_{P_0}$ and $\sup_{\theta\in\Theta} \|W_n(\theta) - W(\theta)\| = o_p(1)$ with $W(\theta)$ uniformly p.d. from Assumption A.2-GMM(b), the statistic $n\hat{Q}_n^W(\theta)$ diverges, i.e., $n\hat{Q}_n^W(\theta) \xrightarrow{p} \infty$, uniformly $\theta \in \Theta/\Theta_{P_0}$.■

LEMMA A.3. *Let Assumptions A.1, A.2-GMM and A.3 be satisfied. Then*

$$n\hat{Q}_n^W(\theta) = \inf_{s_b\in\mathcal{R}_+^b, s_c\in\mathcal{R}^c} \|v(\theta) - s\|_{W(\theta)}^2 + o_p(1)$$

uniformly $\theta \in \Theta_{P_0}$.

PROOF. Now $\sup_{\theta\in\Theta} \|W_n(\theta) - W(\theta)\| = o_p(1)$ by Assumption A.2-GMM(b). Thus, $nQ_n^W(\theta) = v_n^b(\theta)'(W^{bb}(\theta))^{-1}v_n^b(\theta) + o_p(1) \left\|v_n^b(\theta)\right\|_-^2$ from Lemma 1, p.115, of Rosen (2008), as $s_n^c(\theta) = v_n^c(\theta)$, $\theta \in \Theta_{P_0}$, where $W^{bb}(\theta)$ denotes the top left hand $b \times b$ sub-matrix of $W(\theta)^{-1}$. Then, from Lemma A.1,

$$n\hat{Q}_n^W(\theta) = \inf_{s_b\in\mathcal{R}_+^b, s_c\in\mathcal{R}^c} \|v_n(\theta) - s\|_{W(\theta)}^2 + o_p(1)$$

uniformly $\theta \in \Theta_{P_0}$, noting $\sup_{\theta\in\Theta} \left\|v_n^b(\theta)\right\| = O_p(1)$. Now $\sup_{\theta\in\Theta} \|v_n(\theta) - v(\theta)\| = o_p(1)$ by v_n P-Donsker from Assumption A.3 yielding

$$n\hat{Q}_n^W(\theta) = \inf_{s_b\in\mathcal{R}_+^b, s_c\in\mathcal{R}^c} \|v(\theta) - s\|_{W(\theta)}^2 + o_p(1)$$

uniformly $\theta \in \Theta_{P_0}$.■

Define

$$s(\theta) = \arg \min_{s_b\in\mathcal{R}_+^b, s_c\in\mathcal{R}^{q-b}} \|v(\theta) - s\|_{W(\theta)}^2 .$$

LEMMA A4. *Suppose that Assumptions A.1, A.2-GMM and A.3 hold. Then*

$$\sup_{\theta\in\Theta_{P_0}} \|s(\theta)\| = O_p(1).$$

PROOF. The dependence on θ is ignored for ease of exposition.

Now,

$$\inf_{s_b\in\mathcal{R}_+^b, s_c\in\mathcal{R}^c} \|v - s\|_W^2 = \inf_{s_b\in\mathcal{R}_+^b} \|(v - s)_b\|_{W_{bb}}^2$$

where $(\cdot)_b$ denotes the first b elements of $(\cdot)$; see Lemma 1, p.115, of Rosen (2008).

Therefore, from the first order conditions, either (a) $[-W_{bb}(v-s)_b]^j = 0$ and $s^j > 0$ or (b) $[-W_{bb}(v-s)_b]^j > 0$ and $s^j = 0$, $(j = 1, ..., b)$. Define $\mathcal{J} = \{j : [-W_{bb}(v-s)_b]^j = 0$ and $s^j > 0$, $(j = 1, ..., b)\}$. Now, from (a) and (b), $\sum_{k\in\mathcal{J}} W_{bbjk}(v-s)^k = -\sum_{k\notin\mathcal{J}} W_{bbjk}v^k = O_p(1)$, $j \in \mathcal{J}$, uniformly $\theta \in \Theta_{P_0}$, since $\sup_{\theta\in\Theta} \|v(\theta)\| = O_p(1)$ and $\sup_{\theta\in\Theta} \|W(\theta)\| = O(1)$ by Assumptions A.2-GMM(b) and A.3. Hence, $s^j = O_p(1)$, $j \in \mathcal{J}$, uniformly $\theta \in \Theta_{P_0}$ and $s^j = 0$, $j \in \mathcal{J}^c$. Hence the result follows because $(v-s)_c = -W_{cc}W_{cb}(v-s)_b$; see eq. (22), p.115, of Rosen (2008).■

Appendix B: Proofs for GMM

Appendix B establishes the validity of CHT Conditions C.1-C.3 for the GMM criterion $n\hat{Q}_n^W(\theta)$ (3.1) under Assumptions A.1, A.2-GMM and A.3-A.5. CHT Conditions C.4 and C.5 are established for the bounding GMM statistic $n\underline{\hat{Q}}_n^W(\theta)$ (5.1). The relevant CHT constants and sequences are defined as $\gamma = 2$, $a_n = n$ and $b_n = n^{1/2}$. See CHT Theorem 4.2, p.1266.

CHT Condition C.1. Consistency: (**a**) *The parameter space* Θ *is a nonempty compact subset of* $\mathcal{R}^{d_\theta}$. (**b**) *There is a lower semi-continuous population criterion function* $Q : \Theta \to \mathcal{R}_+$ *such that* $\inf_\Theta Q = 0$. *Let* $\Theta_{P_0} = \arg\inf_\Theta Q$ *be the set of its minimisers, called the identified set.* (**c**) *There is a sample criterion function* $\hat{Q}_n(\theta) = \hat{Q}_n(\theta, \{z_i\}_{i=1}^n)$ *that takes values in* $\mathcal{R}_+$ *and is jointly measurable in the parameter* $\theta \in \Theta$ *and the data* z_i, $(i = 1, ..., n)$ *defined on a complete probability space* $(\Omega, \mathcal{F}, P)$. **(d)** *The sample criterion function is uniformly no smaller than the population function in large samples, that is,* $\sup_\Theta(Q - \hat{Q}_n)_+ = O_p(n^{-1/2})$. (**e**) *The sample criterion converges to the limit criterion function over the identified set* Θ_{P_0} *at the rate* $1/n$, *that is,* $\sup_{\Theta_{P_0}} \hat{Q}_n = O_p(n^{-1})$.

Proof. **(a)** Holds by Assumption A.1(a). **(b)** Recall the population GMM criterion function $Q^W(\theta) = \inf_{t\geq 0} \|m(\theta) - t\|^2_{W(\theta)} \geq 0$; see (3.3). Now, $t^j(\theta) = m^j(\theta)$ if $m^j(\theta) > 0$ and 0 if $m^j(\theta) = 0$, $(j = 1, ..., d_m)$, $\theta \in \Theta_{P_0}$. Hence, $Q^W(\theta)$ takes a zero value on Θ_{P_0}, i.e., $\inf_{\theta\in\Theta} Q^W(\theta) = 0$. **(c)** Holds by Assumptions A.1(b) and A.1(d). **(d)** Lemmas A.1 and A.3 establish that $n\hat{Q}_n^W(\theta) = \inf_{s_b\in\mathcal{R}_+^b, s_c\in\mathcal{R}^{d_m-b}} \|v(\theta) - s\|^2_{W(\theta)} + o_p(1)$ uniformly $\theta \in \Theta_{P_0}$ and Lemma A.2 that $n\hat{Q}_n^W(\theta) \xrightarrow{p} \infty$ uniformly $\theta \in \Theta/\Theta_{P_0}$. **(e)** By (b) $Q^W(\theta) = 0$ uniformly $\theta \in \Theta_{P_0}$. Thus $\sup_{\theta\in\Theta_{P_0}} \left|\hat{Q}_n^W(\theta) - Q^W(\theta)\right| = \sup_{\theta\in\Theta_{P_0}} \left|\hat{Q}_n^W(\theta)\right| = O_p(n^{-1})$ again using Lemmas A.1 and A.3.■

CHT Condition C.2. Existence of a Polynomial Minorant: *There exist positive constants* (δ, κ) *such that for an* $\varepsilon \in (0,1)$ *there are* $(\kappa_\varepsilon, n_\varepsilon)$ *such that for all* $n \geq n_\varepsilon$, $\hat{Q}_n(\theta) \geq \kappa \cdot [d(\theta, \Theta_{P_0}) \wedge \delta]^2$ *uniformly on* $\{\theta \in \Theta : d(\theta, \Theta_{P_0}) \geq (\kappa_\varepsilon/n)^{1/2}\}$ *with probability at least* $1 - \varepsilon$.

Proof. Write $W(\theta) = X(\theta)\Lambda(\theta)X(\theta)'$, $\theta \in \Theta$, where the matrix of eigenvectors $X(\theta)$ is orthonormal, i.e., $X(\theta)^{-1} = X(\theta)'$, and eigenvalue matrix $\Lambda(\theta)$ diagonal, $\theta \in \Theta$. Hence, since $X(\theta)X(\theta)' = I_{d_m}$, as $\sup_{\theta\in\Theta} \|W_n(\theta) - W(\theta)\| = o_p(1)$ from Assumption A.2-GMM(b), w.p.a.1 uniformly $\theta \in \Theta$,

$$\begin{aligned}
n\hat{Q}_n^W(\theta) &\geq \inf_{\theta\in\Theta} \lambda_{\min}(W_n(\theta)) \cdot n \min_{t\geq 0} \|\hat{m}_n(\theta) - t\|^2 \\
&= \inf_{\theta\in\Theta} \lambda_{\min}(W(\theta)) \cdot \left\|n^{1/2}\hat{m}_n(\theta)\right\|_-^2 \\
&= \inf_{\theta\in\Theta} \lambda_{\min}(W(\theta)) \cdot \left\|v_n(\theta) + n^{1/2}m(\theta)\right\|_-^2 \\
&= \inf_{\theta\in\Theta} \lambda_{\min}(W(\theta)) \cdot \left\|n^{1/2}m(\theta)\right\|_-^2 \\
&\quad \times \left\|v_n(\theta) + n^{1/2}m(\theta)\right\|_-^2 / \left\|n^{1/2}m(\theta)\right\|_-^2
\end{aligned}$$

where the inequality follows from Assumption A.2-GMM(b) since $\inf_{\theta\in\Theta} \lambda_{\min}(W(\theta)) > 0$ as $W(\theta)$ is uniformly p.d. $\theta \in \Theta$. Now, by Assumption A.4, $\left\|n^{1/2}m(\theta)\right\|_-^2 \geq C \cdot n \cdot (d(\theta, \Theta_{P_0}) \wedge \delta)^2$ for some $C > 0$ and $\delta > 0$. Therefore, as in CHT Proof of Theorem 4.2 Step 1, p.1278, for any $\varepsilon > 0$, with probability at least $1 - \varepsilon$,

$$n\hat{Q}_n^W(\theta) \geq \frac{1}{2} \inf_{\theta\in\Theta} \lambda_{\min}(W(\theta)) \cdot C \cdot n \cdot (d(\theta, \Theta_{P_0}) \wedge \delta)^2$$

uniformly $\{\theta \in \Theta : d(\theta, \Theta_{P_0}) \geq (\kappa_\varepsilon/n)^{1/2}\}$, $n > n_\varepsilon$, for some $(\kappa_\varepsilon, n_\varepsilon)$, from $\sup_{\theta\in\Theta} \|v_n(\theta)\| = O_p(1)$ by the P-Donsker property of Assumption A.3 and $\|y + x\|_- / \|x\|_- \to 1$ as $\|x\|_- \to \infty$ for any $y \in \mathcal{R}^{d_m}$.■

CHT Condition C.3. Degeneracy: *There is a sequence of subsets* Θ_n *of* Θ, *which could be data dependent, such that* $\hat{Q}_n$ *vanishes on these subsets, that is,* $\hat{Q}_n(\theta) - \inf_{\theta\in\Theta} \hat{Q}_n(\theta) = 0$ *for each* $\theta \in \Theta_n$, *for each* n, *and these sets can approximate the identified set arbitrarily well in the Hausdorff distance, that is,* $d_H(\Theta_n, \Theta_{P_0}) \leq \epsilon_n$ *for some sequence* $\epsilon_n = O_p(n^{-1/2})$.

PROOF. Similarly to the Proof of CHT Condition C.2 above, w.p.a.1 uniformly $\theta \in \Theta_{P_0}$,

$$\begin{aligned} n\hat{Q}_n^W(\theta) &\leq \sup_{\theta\in\Theta} \lambda_{\max}(W(\theta)) \cdot n \min_{t\geq 0} \|\hat{m}_n(\theta) - t\|^2 \\ &= \sup_{\theta\in\Theta} \lambda_{\max}(W(\theta)) \cdot \left\|v_n(\theta) + n^{1/2}m(\theta)\right\|_-^2 \\ &\leq \sup_{\theta\in\Theta} \lambda_{\max}(W(\theta)) \cdot \sum_{j=1}^{d_m} [v_n^j(\theta) + n^{1/2}m^j(\theta)]_-^2 \\ &\leq \sup_{\theta\in\Theta} \lambda_{\max}(W(\theta)) \cdot d_m \cdot [O_p(1) + n^{1/2} \cdot C \cdot (d(\theta, \Theta\backslash\Theta_{P_0}) \wedge \delta)]_-^2 \end{aligned}$$

where the first inequality follows from $W(\theta)$ uniformly p.d. $\theta \in \Theta$ and bounded by Assumption A.2-GMM(b), the second by T and the third inequality by Assumption A.5. The conclusion follows as in CHT Proof of Theorem 4.2 Step 2, p.1278, since, with $\epsilon_n = O_p(n^{-1/2})$, $\hat{Q}_n^W(\theta) = 0$ for $\theta \in \Theta_{P_0}^{-\epsilon_n}$.■

The Proofs of CHT Conditions C.4, p.1256, and C.5, p.1257, given below concern the bounding statistic $n\underline{\hat{Q}}_n^W(\theta)$ (5.1) for the GMM criterion $n\hat{Q}_n^W(\theta)$ (3.1). These results establish the validity of the asymptotically conservative inference procedure for Θ_{P_0} described in section 5.1.

Define $\underline{\mathcal{C}}_n^W = \sup_{\theta\in\Theta_{P_0}} \underline{\hat{Q}}_n^W(\theta)$ and $\underline{\mathcal{C}}^W = \sup_{\theta\in\Theta_{P_0}} \|[v(\theta) + \xi(\theta)]_-\|_{W(\theta)}^2$ where $\xi^j(\theta) = 0$ if $m^j(\theta) = 0$, $(j = 1, ..., b)$, and $\xi^j(\theta) = \infty$ if $m^j(\theta) > 0$, $(j = b+1, ..., d_m)$, $\theta \in \Theta_{P_0}$.

CHT CONDITION C.4. Convergence of $\underline{\mathcal{C}}_n^W$: *$\mathcal{P}[\underline{\mathcal{C}}_n^W \leq \underline{c}_W] \to \mathcal{P}[\underline{\mathcal{C}}^W \leq \underline{c}_W]$ for each $\underline{c}_W \in [0, \infty)$, where the distribution function of $\underline{\mathcal{C}}^W$ is non-degenerate and continuous on $[0, \infty)$.*

PROOF. Define $\theta_n(\lambda) = \theta + n^{-1/2}\lambda$ and $l_n^W(\theta, \lambda) = n\underline{\hat{Q}}_n^W(\theta_n(\lambda))$. Then, for $(\theta, \lambda) \in \Theta \times \mathcal{B}_\delta$,

$$\begin{aligned} l_n^W(\theta, \lambda) &= [n^{1/2}\hat{m}_n(\theta_n(\lambda))]_-' W_n(\theta_n(\lambda))[n^{1/2}\hat{m}_n(\theta_n(\lambda))]_- \\ &= [v_n(\theta_n(\lambda)) + n^{1/2}m(\theta_n(\lambda))]_-' W_n(\theta_n(\lambda))[v_n(\theta_n(\lambda)) + n^{1/2}m(\theta_n(\lambda))]_- \\ &= \left\|[v_n(\theta_n(\lambda)) + n^{1/2}m(\theta_n(\lambda))]_-\right\|_{W_n(\theta_n(\lambda))}^2 \end{aligned}$$

First, by the P-Donsker property of $v_n(\theta)$ of Assumption A.3, $v_n(\theta) \Rightarrow v(\theta)$ and $v(\theta)$ stochastically equicontinuous. Hence, $v_n(\theta_n(\lambda)) \Rightarrow v(\theta)$ uniformly $(\theta, \lambda) \in \Theta \times \mathcal{B}_\delta$. Secondly, from Assumption A.2-GMM(b), $\sup_{\theta\in\Theta} |W_n(\theta) - W(\theta)| = o_p(1)$ and $W(\theta)$

continuous, thus $W_n(\theta_n(\lambda)) \xrightarrow{p} W(\theta)$ uniformly $(\theta,\lambda) \in \Theta \times \mathcal{B}_\delta$. Therefore, uniformly $(\theta,\lambda) \in \Theta \times \mathcal{B}_\delta$,

$$l_n^W(\theta,\lambda) = \left\|[v(\theta) + n^{1/2}m(\theta_n(\lambda))]_-\right\|^2_{W(\theta)} + o_p(1). \tag{B.1}$$

Next define $l_\infty^W(\theta,\lambda) = [v(\theta)+M(\theta)\lambda+\xi(\theta)]'_- W(\theta)[v(\theta)+M(\theta)\lambda+\xi(\theta)]_- = \|[v(\theta)+M(\theta)\lambda+\xi(\theta)]_-\|^2_W$ By Assumption A.4 $n^{1/2}m(\theta_n(\lambda)) = M(\theta)\lambda + \xi(\theta) + o(1)$ uniformly $(\theta,\lambda) \in \Theta_{P_0} \times \mathcal{B}_\delta$. Therefore, from (B.1),

$$l_n^W(\theta,\lambda) - l_\infty^W(\theta,\lambda) = o_p(1) \tag{B.2}$$

uniformly $L^\infty(\Theta_{P_0} \times \mathcal{B}_\delta)$.

Now, by definition, $\underline{\mathcal{C}}_n^W = \sup_{\theta\in\Theta_{P_0}} l_n^W(\theta,0)$ and $\underline{\mathcal{C}}^W = \sup_{\theta\in\Theta_{P_0}} l_\infty^W(\theta,0)$. Therefore, by (B.2),

$$\underline{\mathcal{C}}_n^W \xrightarrow{d} \underline{\mathcal{C}}^W. \tag{B.3}$$

■

CHT Condition C.5. Approximability of $\underline{\mathcal{C}}_n^W$: *Let Θ_n be any sequence of subsets of Θ such that $d_H(\Theta_n,\Theta_{P_0}) = o_p(n^{-1/2})$ and define $\underline{\mathcal{C}}_n^{W\prime} = \sup_{\theta\in\Theta_n} n\hat{\underline{Q}}_n^W(\theta)$. Then for any $\underline{c}_W \geq 0$, we have that $\mathcal{P}[\underline{\mathcal{C}}_n^{W\prime} \leq \underline{c}_W] = \mathcal{P}[\underline{\mathcal{C}}_n^W \leq \underline{c}_W] + o(1)$.*

Proof. By arguments similar to those in CHT Proof of Theorem 4.2 Step 4, pp.1279-80,

$$\begin{aligned}
\underline{\mathcal{C}}_n^{W\prime} &= \sup_{\theta\in\Theta_n} n\hat{\underline{Q}}_n^W(\theta) \\
&= \sup_{\theta\in\Theta_n} \left\|[v(\theta) + n^{1/2}m(\theta) + o_p(1)]_-\right\|^2_{W(\theta)} \\
&= \sup_{\theta\in\Theta_{P_0}} \left\|[v(\theta) + n^{1/2}m(\theta) + o_p(1)]_-\right\|^2_{W(\theta)}
\end{aligned}$$

using the approximation device in the Proof of CHT Condition C.4 above, cf. CHT Proof of Theorem 4.2 Step 2, p.1278, the stochastic equicontinuity of $\theta \to (v(\theta), W(\theta))$ and $\left\|n^{1/2}(m(\theta) - m(\theta'))\right\| = o(1)$ uniformly on $\{\theta,\theta' \in \Theta : \|\theta - \theta'\| \leq o_p(n^{-1/2})\}$. The conclusion then follows as in CHT Proof of Theorem 4.2 Step 3, p.1279.■

Appendix C: Proofs for GEL

Let $\Lambda_n = \{\lambda \in \mathcal{R}^{d_m} : \|\lambda\| \leq Cn^{-1/2}\}$. In the following $\hat{Q}_n(\theta)$ and $Q(\theta)$ refer to sample and population GMM criteria that respectively employ the efficient metrics $\hat{\Omega}_n(\theta)^{-1}$ and $\Omega(\theta)^{-1}$ appropriate for unconditional moment equaility restrictions.

LEMMA C.1. *If Assumptions A.1 and A.2-GEL hold then* **(a)** $\max_{1\leq i\leq n}\sup_{\theta\in\Theta,\lambda\in\Lambda_n}|\lambda' m_i(\theta)| \xrightarrow{p} 0$; **(b)** *w.p.a.1,* $\Lambda_n \subseteq \hat{\Lambda}_n(\theta)$ *for all* $\theta \in \Theta$.

PROOF. Follows directly from Newey and Smith (2004, Lemma A1, p.239) and the extension Parente and Smith (2011, Lemma A.1, p.101).■

Statements and proofs are given for the alternative GEL criterion $\tilde{P}_n^\rho(\theta)$ (E.6) defined in Appendix D; those for the GEL criterion $\hat{P}_n^\rho(\theta)$ (3.4) and alternative GEL criteria $\tilde{P}_n^{\rho,k}(\theta)$, $(k = a, b)$, (E.1) and (E.3), follow similarly.

Recall $\hat{\Omega}_n(\theta) = \sum_{i=1}^n m_i(\theta)m_i(\theta)'/n$. The next Lemma and its proof mirror Newey and Smith (2004,Lemma A2, p.239) for the moment equality case.

LEMMA C.2. *Let* $\theta \in \Theta_{P_0}$. *Let the arbitrary sequence* $\tau_n(\theta)$ *obey* $\|\hat{m}_n(\theta) - \tau_n(\theta)\| = O_p(n^{-1/2})$ *uniformly* $\theta \in \Theta_{P_0}$. *If Assumptions A.1 and A.2-GEL are satisfied, then* $\tilde{\lambda}_n(\theta) = \arg\max_{\lambda\in\hat{\Lambda}_n(\theta)} \tilde{P}_n^\rho(\theta, \lambda, \tau_n(\theta))$ *exists w.p.a.1,* $\tilde{\lambda}_n(\theta) = O_p(n^{-1/2})$ *and* $\sup_{\lambda\in\hat{\Lambda}_n(\theta)} \tilde{P}_n^\rho(\theta, \lambda, \tau_n(\theta)) \leq O_p(n^{-1})$.

PROOF. By Assumption A.2-GEL and UWL $\hat{\Omega}_n(\theta) \xrightarrow{p} \Omega(\theta)$ uniformly $\theta \in \Theta$. Then, by $\Omega(\theta)$ p.d. uniformly $\theta \in \Theta$ from Assumption A.1(c), the smallest eigenvalue of $\hat{\Omega}_n(\theta)$ is bounded away from zero w.p.a.1. By Lemma C.1 and twice continuous differentiability of $\rho(\cdot)$ in a neighborhood of zero from Assumption A.2-GEL(b), $\tilde{P}_n^\rho(\theta, \lambda, \tau_n(\theta))$ is twice continuously differentiable on Λ_n w.p.a.1 uniformly $\theta \in \Theta$. Write $\lambda_n = \lambda_n(\theta)$. Then, $\lambda_n = \arg\max_{\lambda\in\Lambda_n} \tilde{P}_n^\rho(\theta, \lambda, \tau_n(\theta))$ exists w.p.a.1. Furthermore, for any $\dot{\lambda}$ on the line segment joining λ_n and 0, by Lemma C.1 and $\rho_2(0) = -1$, $\max_{1\leq i\leq n} \rho_2(\dot{\lambda}' m_i(\theta)) < -1/2$ w.p.a.1. Hence, by a Taylor expansion around $\lambda = 0$ with Lagrange remainder, there is

$\dot{\lambda}$ on the line joining λ_n and 0 such that

$$\begin{aligned} 0 &= \tilde{P}_n^\rho(\theta, 0, \tau_n(\theta)) \\ &\le \tilde{P}_n^\rho(\theta, \lambda_n, \tau_n(\theta)) = -(\hat{m}_n(\theta) - \tau_n(\theta))'\lambda_n + \frac{1}{2}\lambda_n'[\sum_{i=1}^n \rho_2(\dot{\lambda}' m_i(\theta)) m_i(\theta) m_i(\theta)'/n]\lambda_n \\ &\le -(\hat{m}_n(\theta) - \tau_n(\theta))'\lambda_n - \frac{1}{4}\lambda_n'\hat{\Omega}_n(\theta)\lambda_n \le \|\lambda_n\|\|\hat{m}_n(\theta) - \tau(\theta)\| - C\|\lambda_n\|^2 \end{aligned}$$

uniformly $\theta \in \Theta_{P_0}$. Adding $C\|\lambda_n\|^2$ to both sides and dividing by $\|\lambda_n\|$ yields $C\|\lambda_n\| \le \|\hat{m}_n(\theta) - \tau_n(\theta)\|$ w.p.a.1. By hypothesis, $\hat{m}_n(\theta) - \tau_n(\theta) = O_p(n^{-1/2})$, $\theta \in \Theta_{P_0}$, and, thus, $\|\lambda_n\| = O_p(n^{-1/2})$. Therefore, w.p.a.1 $\lambda_n \in int(\Lambda_n)$ and hence $\partial\tilde{P}_n^\rho(\theta, \lambda_n, \tau_n(\theta))/\partial\lambda = 0$, the first order conditions for an interior maximum. By Lemma C.1, w.p.a.1 $\lambda_n \in \hat{\Lambda}_n(\theta)$, so by the concavity of $\tilde{P}_n^\rho(\theta, \lambda, \tau_n(\theta))$ and convexity of $\hat{\Lambda}_n(\theta)$ it follows that $\tilde{P}_n^\rho(\theta, \lambda_n, \tau_n(\theta)) = \sup_{\lambda \in \hat{\Lambda}_n^\rho(\theta)} \tilde{P}_n^\rho(\theta, \lambda, \tau_n(\theta))$, giving the first and second conclusions with $\lambda_n = \tilde{\lambda}_n$. Then, by the last inequality of the above equation, $\|\hat{m}_n(\theta) - \tau_n(\theta)\| = O_p(n^{-1/2})$, and $\|\lambda_n\| = O_p(n^{-1/2})$, we obtain $\tilde{P}_n^\rho(\theta, \tilde{\lambda}_n, \tau_n(\theta)) \le \|\tilde{\lambda}_n\|\|\hat{m}_n(\theta) - \tau_n(\theta)\| - C\|\tilde{\lambda}_n\|^2 = O_p(n^{-1})$ uniformly $\theta \in \Theta_{P_0}$.■

Lemma C.3. *Let $\theta \in \Theta_{P_0}$. If Assumptions A.1 and A.2-GEL are satisfied, then $\tilde{\lambda}_n(\theta) = \arg\max_{\lambda \in \hat{\Lambda}_n(\theta)} \tilde{P}_n^\rho(\theta, \lambda, \tilde{\tau}_n(\theta))$ exists w.p.a.1, $\tilde{\lambda}_n(\theta) = O_p(n^{-1/2})$, $\sup_{\theta \in \Theta_{P_0}} \|\hat{m}_n(\theta) - \tilde{\tau}_n(\theta)\| \le O_p(n^{-1/2})$ and $\sup_{\lambda \in \hat{\Lambda}_n(\theta)} \tilde{P}_n^\rho(\theta, \lambda, \tilde{\tau}_n(\theta)) \le O_p(n^{-1})$.*

Proof. From the Proofs of Lemmas E.1 and E.3 below the population auxiliary paramater $\lambda(\theta) = 0$, $\theta \in \Theta_{P_0}$. Thus, the population slackness parameter $\tau(\theta) = \mathbb{E}_{P_0}[m(z, \theta)] \ge 0$. In particular, $\tau^j(\theta) > [=]0$ if and only if $m^j(\theta) > [=]0$, $(j = 1, ..., d_m)$. Let $\tilde{\lambda}_n$ satisfy $\tilde{P}_n^\rho(\theta, \tilde{\lambda}_n, \tau(\theta)) = \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \tilde{P}_n^\rho(\theta, \lambda, \tau(\theta))$. Then, $\tilde{P}_n^\rho(\theta, \tilde{\lambda}_n(\theta), \tilde{\tau}_n(\theta)) \le \tilde{P}_n^\rho(\theta, \tilde{\lambda}_n(\theta), \tau(\theta)) \le \tilde{P}_n^\rho(\theta, \tilde{\lambda}_n, \tau(\theta))$ uniformly $\theta \in \Theta_{P_0}$. Therefore, from the Proof of Lemma C.2, $\tilde{\lambda}_n(\theta) = O_p(n^{-1/2})$ and $\tilde{P}_n^\rho(\theta, \tilde{\lambda}_n(\theta), \tilde{\tau}_n(\theta)) \le O_p(n^{-1})$, i.e.,

$$\inf_{\tau \in \mathcal{T}} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \tilde{P}_n^\rho(\theta, \lambda, \tau) \le O_p(n^{-1})$$

uniformly $\theta \in \Theta_{P_0}$.

Let $\bar{\lambda}_n = -n^{-1/2}(\hat{m}_n(\theta) - \tilde{\tau}_n(\theta))/\|\hat{m}_n(\theta) - \tilde{\tau}_n(\theta)\|$ and, thus, $\bar{\lambda}_n \in \Lambda_n$, $\theta \in \Theta_{P_0}$. By Lemma C.1, $\max_{1 \le i \le n} |\bar{\lambda}_n' m_i(\theta)| \xrightarrow{p} 0$ and $\bar{\lambda}_n \in \hat{\Lambda}_n(\theta)$ w.p.a.1. Thus, for any $\dot{\lambda}$ on the line joining $\bar{\lambda}_n$ and 0, w.p.a.1 $\rho_2(\dot{\lambda}' m_i(\theta)) \ge -C$, $(i = 1, ..., n)$. Also, by UWL and Assumption A.2, the largest eigenvalue of $\sum_{i=1}^n m_i(\theta) m_i(\theta)'/n$ is bounded above w.p.a.1.

An expansion then gives

$$\begin{aligned}
\tilde{P}_n^\rho(\theta,\bar\lambda_n,\tilde\tau_n(\theta)) &= -(\hat m_n(\theta)-\tilde\tau_n(\theta))'\bar\lambda_n + \frac{1}{2}\bar\lambda_n'[\sum_{i=1}^n \rho_2(\dot\lambda' m_i(\theta))m_i(\theta)m_i(\theta)'/n]\bar\lambda_n \\
&\geq n^{-1/2}\|\hat m_n(\theta)-\tilde\tau_n(\theta)\| - C\frac{1}{2}\bar\lambda_n'\hat\Omega_n(\theta)\bar\lambda_n \geq n^{-1/2}\|\hat m_n(\theta)-\tilde\tau_n(\theta)\| - Cn^{-1}
\end{aligned}$$

w.p.a.1. uniformly $\theta\in\Theta_{P_0}$. Hence,

$$n^{-1/2}\|\hat m_n(\theta)-\tilde\tau_n(\theta)\| - Cn^{-1} \leq \tilde P_n^\rho(\theta,\bar\lambda,\tilde\tau_n(\theta)) \leq \tilde P_n^\rho(\theta,\tilde\lambda_n(\theta),\tilde\tau_n(\theta)) \leq O_p(n^{-1}). \tag{C.1}$$

Solving eq. (C.1) for $\|\hat m_n(\theta)-\tilde\tau_n(\theta)\|$ then gives

$$\|\hat m_n(\theta)-\tilde\tau_n(\theta)\| \leq O_p(n^{-1/2}). \tag{C.2}$$

uniformly $\theta\in\Theta_{P_0}$.■

Recall $\tilde P_n^\rho(\theta) = \inf_{\tau\in\mathcal{T}}\sup_{\lambda\in\hat\Lambda_n(\theta)}\tilde P_n^\rho(\theta,\lambda,\tau)$.

Lemma C.4. *Under Assumptions A.1 and A.2-GEL*

$$\begin{aligned}
2n\tilde P_n^\rho(\theta) &= n\hat Q_n(\theta)+o_p(1) \\
&= \inf_{s_b\in\mathcal{R}_+^b,\, s_c\in\mathcal{R}^{d_m-b}}\|v(\theta)-s\|^2_{\Omega(\theta)^{-1}} + o_p(1),
\end{aligned}$$

uniformly $\theta\in\Theta_{P_0}$.

Proof. Cf. Canay (2010, Proof of Theorem 3.1, pp.418-419). Let the arbitrary sequence τ_n obey $\|\hat m_n(\theta)-\tau_n\| = O_p(n^{-1/2})$; cf. Lemmas C.1 and C.2 above. Define $\tilde\lambda_n(\tau_n) = \arg\max_{\lambda\in\hat\Lambda_n(\theta)}\tilde P_n^\rho(\theta,\lambda,\tau_n)$. Therefore, cf. the Proof of Lemma C.2 above, w.p.a.1, $\tilde\lambda_n(\tau_n)\in int(\hat\Lambda_n(\theta))$ and $\tilde\lambda_n(\tau_n)$ satisfies the first order conditions for an interior maximum $\partial\tilde P_n^\rho(\theta,\lambda,\tau_n)/\partial\lambda = 0$, i.e., $\tilde\lambda_n(\tau_n) = -\hat\Omega_n^{-1}(\hat m_n(\theta)-\tau_n)+o_p(n^{-1/2})$ uniformly $\theta\in\Theta_{P_0}$. Hence, defining $\dot\lambda(\tau)$ on the line joining $\tilde\lambda_n(\tau)$ and 0,

$$\begin{aligned}
2n\tilde P_n^\rho(\theta,\tilde\lambda_n(\tilde\tau_n),\tilde\tau_n) &= 2\inf_{\tau\in\mathcal{T}}\sum_{i=1}^n\rho(\tilde\lambda_n(\tau)'m_i(\theta))+\tilde\lambda_n(\tau)'\tau \\
&= 2\inf_{\tau\in\mathcal{T}} -n(\hat m_n(\theta)-\tau)'\tilde\lambda_n(\tau)+\frac{1}{2}n\tilde\lambda_n(\tau)'[\sum_{i=1}^n\rho_2(\dot\lambda(\tau)'m_i(\theta))m_i(\theta)m_i(\theta)'/n]\tilde\lambda_n(\tau) \\
&= 2\inf_{\tau\in\mathcal{T}} -n(\hat m_n(\theta)-\tau)'\tilde\lambda_n(\tau)-\frac{1}{2}n\tilde\lambda_n(\tau)'\hat\Omega_n(\theta)\tilde\lambda_n(\tau)+o_p(1) \\
&= \inf_{\tau\in\mathcal{T}} n\|\hat m_n(\theta)-\tau\|^2_{\hat\Omega_n^{-1}}+o_p(1) \\
&= n\hat Q_n(\theta)+o_p(1),
\end{aligned}$$

uniformly $\theta \in \Theta_{P_0}$, using Lemmas C.1 and C.3.

It then follows by Lemma A.3 in Appendix A that, uniformly $\theta \in \Theta_{P_0}$,

$$\begin{aligned} 2n\tilde{P}_n^{\rho}(\theta, \tilde{\lambda}_n, \tilde{\tau}_n) &= \inf_{t \geq 0} n \left\| \hat{m}_n(\theta) - t \right\|^2_{\Omega(\theta)^{-1}} + o_p(1) \\ &= \inf_{s_b \in \mathcal{R}^b_+, s_c \in \mathcal{R}^{d_m - b}} \left\| v(\theta) - s \right\|^2_{\Omega(\theta)^{-1}} + o_p(1). \end{aligned}$$

■

Lemma C.5. *Under Assumptions A.1-A.2-GEL, $2n\tilde{P}_n^{\rho}(\theta) \xrightarrow{p} \infty$ uniformly $\theta \in \Theta/\Theta_{P_0}$.*

Proof. The structure of the following argument closely resembles that of Smith (2007, Proof of Theorem 4.1, pp.112-114); cf. KTA, Proof of Theorem 3.1, pp.1686-1688, for EL.

Let $c > 0$ such that $(-c, c) \in \mathcal{V}$. Define $C_n = \{z \in \mathbb{R}^{d_z} : \sup_{\theta \in \Theta} \|m(z, \theta)\| \leq cn^{1/2}\}$ and $m_{ni}(\theta) = I_i m_i(\theta)$, where $I_i = I\{z_i \in C_n\}$. Let $\bar{\lambda}(\theta, \tau) = -(m(\theta) - \tau)/(1 + \|m(\theta) - \tau\|)$; note that $n^{-1/2}\bar{\lambda}(\theta, \tau) \in \Lambda_n$.

Then,

$$\begin{aligned} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \tilde{P}_n^{\rho}(\theta, \lambda, \tau) &\geq \tilde{Q}_n^{\rho}(\theta, \tau) \\ &= \sum_{i=1}^n \rho(n^{-1/2}\bar{\lambda}(\theta, \tau)' m_{ni}(\theta))/n + n^{-1/2}\bar{\lambda}(\theta, \tau)'\tau. \end{aligned} \tag{C.3}$$

Now

$$\rho(n^{-1/2}\bar{\lambda}(\theta, \tau)' m_{ni}(\theta)) + n^{-1/2}\bar{\lambda}(\theta, \tau)'\tau = -n^{-1/2}\bar{\lambda}(\theta, \tau)'(m_i(\theta) - \tau) + r_{ni}(t),$$

for some $t \in (0, 1)$ and remainder

$$\begin{aligned} r_{ni}(t) &= n^{-1/2}\bar{\lambda}(\theta, \tau)' m_i(\theta)(1 - I_i) \\ &\quad + n^{-1/2}\bar{\lambda}(\theta, \tau)' m_{ni}(\theta)[\rho_1(tn^{-1/2}\bar{\lambda}(\theta, \tau)' m_{ni}(\theta)) - \rho_1(0)]. \end{aligned} \tag{C.4}$$

From Lemma C.4 $\sup_{\theta \in \Theta, n^{-1/2}\lambda \in \Lambda_n, 1 \leq i \leq n} \left|\rho_1(n^{-1/2}\lambda' m_i(\theta)) - \rho_1(0)\right| \xrightarrow{p} 0$. Also $\max_{1 \leq i \leq n}(1 - I_i) = o_p(1)$. Hence, from eq. (C.4),

$$\begin{aligned} n^{1/2} \sum_{i=1}^n r_{ni}(t)/n &= o_p(1)\bar{\lambda}(\theta, \tau)'\hat{m}_n(\theta) + o_p(1)\bar{\lambda}(\theta, \tau)'\hat{m}_n(\theta) \\ &\quad - o_p(1)\bar{\lambda}(\theta, \tau)' \sum_{i=1}^n m_i(\theta)(1 - I_i)/n \\ &= o_p(1)\bar{\lambda}(\theta, \tau)'\hat{m}_n(\theta) \end{aligned}$$

uniformly $\theta \in \Theta$ and $\tau \in \mathcal{T}$. Thus,

$$\begin{aligned} n^{1/2} \sup_{\theta\in\Theta,\tau\in\mathcal{T}} \left| \sum_{i=1}^n r_{ni}(t)/n \right| &\leq o_p(1) \sup_{\theta\in\Theta} \|\hat{m}_n(\theta)\| \\ &= o_p(1)O_p(1) = o_p(1) \end{aligned}$$

as $\sup_{\theta\in\Theta} \|\hat{m}_n(\theta)\| \leq \sup_{\theta\in\Theta} \|m(\theta)\| + o_p(1)$ by T and UWL. Therefore, substituting eq. (C.3), $n^{1/2}\tilde{Q}_n^\rho(\theta,\tau) = -\bar{\lambda}(\theta,\tau)'(\hat{m}_n(\theta)-\tau) + o_p(1)$ uniformly $\theta \in \Theta$ and $\tau \in \mathcal{T}$. By UWL

$$n^{1/2} \sup_{\theta\in\Theta,\tau\in\mathcal{T}} \left| \tilde{Q}_n(\theta,\tau) - \tilde{Q}(\theta,\tau) \right| = o_p(1), \tag{C.5}$$

where

$$\begin{aligned} n^{1/2}\tilde{Q}(\theta,\tau) &= -\bar{\lambda}(\theta,\tau)'(m(\theta)-\tau) \\ &= \frac{\|m(\theta)-\tau\|^2}{1+\|m(\theta)-\tau\|}. \end{aligned}$$

Thus, from eqs. (C.3) and (C.5), cf. KTA, eqs. (A.6) and (A.7), p.1687,

$$n^{1/2} \inf_{\tau\in\mathcal{T}} \sup_{\lambda\in\hat{\Lambda}_n(\theta)} \tilde{P}_n^\rho(\theta,\lambda,\tau) \geq n^{1/2} \inf_{\tau\in\mathcal{T}} \tilde{Q}(\theta,\tau) + o_p(1) \tag{C.6}$$

uniformly $\theta \in \Theta$.

The function $\|m(\theta)-\tau\|^2/(1+\|m(\theta)-\tau\|)$ is continuous in θ and τ. By definition of the identified set Θ_{P_0}, $\inf_{\tau\in\mathcal{T}} \|m(\theta)-\tau\|^2/(1+\|m(\theta)-\tau\|)$ takes the value zero for all $\theta \in \Theta_{P_0}$ and is strictly positive for all $\theta \in \Theta/\Theta_{P_0}$, i.e.,

$$\inf_{\tau\in\mathcal{T}} \tilde{Q}(\theta,\tau) = 0 \Longleftrightarrow \theta \in \Theta_{P_0}$$

and

$$\inf_{\tau\in\mathcal{T}} \tilde{Q}(\theta,\tau) > 0 \Longleftrightarrow \theta \notin \Theta_{P_0}.$$

Therefore, from eq. (C.6), uniformly $\theta \in \Theta/\Theta_{P_0}$, $n^{1/2}\inf_{\tau\in\mathcal{T}}\sup_{\lambda\in\hat{\Lambda}_n(\theta)} \tilde{P}_n^\rho(\theta,\lambda,\tau) \xrightarrow{p} \infty$.

Similarly to the Proof of Condition C.1(d) for GMM, $2n\tilde{P}_n^\rho(\theta) = 2n\tilde{P}_n^\rho(\theta,\tilde{\lambda}_n(\theta),\tilde{\tau}_n(\theta)) \xrightarrow{p} \infty$ uniformly $\theta \in \Theta/\Theta_{P_0}$.■

Recall from section 3.4 the population criterion defined by $\tilde{P}^\rho(\theta) = \inf_{\tau\in\mathcal{T}}\sup_{\lambda\in\mathcal{R}^{d_m}} \tilde{P}^\rho(\theta,\lambda,\tau)$ with $\tilde{P}^\rho(\theta,\lambda,\tau) = \mathbb{E}_{P_0}[\rho(\lambda'm(z,\theta))-\rho_0]+\lambda'\tau$ correponding to the alternative GEL criterion $\tilde{P}_n^\rho(\theta)$ (E.6). Proofs are presented for $\tilde{P}_n^\rho(\theta)$ (E.6); those for $\hat{P}_n^\rho(\theta)$ (3.4) and $\tilde{P}_n^{\rho,k}(\theta)$, $(k=a,b)$, (E.1) and (E.3), follow similarly. In the following discussion $\tilde{P}^\rho(\theta)$ substitutes for $\hat{Q}_n(\theta)$ in the statements of CHT Conditions C.1-C.3 in Appendix B.

Proof of CHT Condition C.1. **(a)** Holds by Assumption A.1. **(b)** Follows for $\tilde{P}^\rho(\theta)$ from Lemmas E.1 and E.3 below as $\tilde{P}^\rho(\theta) = \hat{P}^\rho(\theta) = 0$ for all $\theta \in \Theta_{P_0}$, i.e., $\inf_{\theta\in\Theta} \tilde{P}^\rho(\theta) = 0$. **(c)** Holds by Assumption A.2-GEL(a). **(d)** Lemma C.4 establishes that $2n\tilde{P}_n^\rho(\theta, \tilde{\lambda}_n(\theta), \tilde{\tau}_n(\theta)) = n\hat{Q}_n(\theta) + o_p(1)$ uniformly $\theta \in \Theta_{P_0}$ and Lemma C.5 that $2n\tilde{P}_n^\rho(\theta, \tilde{\lambda}_n(\theta), \tilde{\tau}_n(\theta)) \xrightarrow{p} \infty$ uniformly $\theta \in \Theta/\Theta_{P_0}$. **(e)** By (b) $\tilde{P}^\rho(\theta) = 0$ uniformly $\theta \in \Theta_{P_0}$. Thus $\sup_{\theta\in\Theta_{P_0}} \left|\tilde{P}_n^\rho(\theta) - \tilde{P}^\rho(\theta)\right| = \sup_{\theta\in\Theta_{P_0}} \left|\tilde{P}_n^\rho(\theta)\right| = O_p(n^{-1})$ using Lemma C.3.■

Proof of CHT Condition C.2. Similarly to the Proof of Condition C.2 for GMM in Appendix B, from Lemmas C.4 and C.5, w.p.a.1 uniformly $\theta \in \Theta$,

$$\begin{aligned} 2n\tilde{P}_n^\rho(\theta) &= n\hat{Q}_n(\theta) \\ &\geq \inf_{\theta\in\Theta} \left\|n^{1/2}m(\theta)\right\|_-^2 \\ &\quad \times \left\|v_n(\theta) + n^{1/2}m(\theta)\right\|_-^2 / \lambda_{\max}(\Omega(\theta)) \left\|n^{1/2}m(\theta)\right\|_-^2 . \end{aligned}$$

Therefore, for any $\varepsilon > 0$, with probability at least $1 - \varepsilon$,

$$2n\tilde{P}_n^\rho(\theta) \geq \frac{1}{2} \inf_{\theta\in\Theta} C \cdot n \cdot (d(\theta, \Theta_{P_0}) \wedge \delta)^2 / \lambda_{\max}(\Omega(\theta))$$

uniformly $\{\theta \in \Theta : d(\theta, \Theta_{P_0}) \geq (\kappa_\varepsilon/n)^{1/2}\}$, $n > n_\varepsilon$, for some $(\kappa_\varepsilon, n_\varepsilon)$, from $\sup_{\theta\in\Theta} \|v_n(\theta)\| = O_p(1)$ by the P-Donsker property of Assumption A.3, Assumption A.4 and $\|y + x\|_- / \|x\|_- \to 1$ as $\|x\|_- \to \infty$ for any $y \in \mathcal{R}^{d_m}$.■

Proof of CHT Condition C.3. Similarly to the Proof of Condition C.3 for GMM in Appendix B, w.p.a.1 uniformly $\theta \in \Theta_{P_0}$,

$$\begin{aligned} 2n\tilde{P}_n^\rho(\theta) &= n\hat{Q}_n(\theta) \\ &\leq \sup_{\theta\in\Theta} d_m \cdot [O_p(1) + n^{1/2} \cdot C \cdot (d(\theta, \Theta \backslash \Theta_{P_0}) \wedge \delta)]_-^2 / \lambda_{\min}(\Omega(\theta)) \end{aligned}$$

where the inequality follows from $\Omega(\theta)$ uniformly p.d. $\theta \in \Theta$ and bounded by Assumption A.2-GMM(b) and Assumption A.5. The conclusion follows as in the Proof of Condition C.3 for GMM, since with $\epsilon_n = O_p(n^{-1/2})$, $\tilde{P}_n^\rho(\theta) = 0$ on $\Theta_{P_0}^{-\epsilon_n}$.■

Appendix D: Identified Set

Recall the partition of the index set $\{1, ..., d_m\}$ according to $m^j(\theta) < 0$, $(j = 1, ..., a)$, $m^j(\theta) = 0$, $(j = a+1, ..., a+b)$ and $m^j(\theta) > 0$, $(j = a+b+1, ..., d_m)$. Let $c = d_m - a - b$. Note again that a, b and thus c depend on θ. Also recall the notation $m(\theta) = \mathbb{E}_{P_0}[m(z, \theta)]$.

Recall $\hat{\Theta}^{\rho}_{P_0} = \{\theta \in \Theta : \theta = \arg\min_{\theta \in \Theta} \hat{P}^{\rho}(\theta)\}$ (3.9).

Lemma D.1. *Suppose that Assumptions A.1 and A.2-GEL are satisfied. Then* $\hat{\Theta}^{\rho}_{P_0} = \Theta_{P_0}$.

Proof. Now

$$\begin{aligned} \hat{P}^{\rho}(\theta) &= \sup_{\lambda \geq 0} \mathbb{E}_{P_0}[\rho(\lambda(\theta)' m(z, \theta))] \\ &= \mathbb{E}_{P_0}[\rho(\lambda(\theta)' m(z, \theta))] \geq 0 \end{aligned}$$

since $\rho(\lambda' m(z, \theta)) = 0$ at $\lambda = 0$.

Fix $\theta \in \Theta_{P_0}$; thus $a = 0$. Consider $j \in \{b+1, ..., d_m\}$, i.e., $m^j(\theta) > 0$. Suppose that the associated auxiliary parameter $\lambda^j(\theta) > 0$. Now

$$\begin{aligned} \mathbb{E}_{P_0}[\rho(\lambda(\theta)' m(z, \theta))] &\leq \rho(\lambda(\theta)' m(\theta)) \\ &< 0; \end{aligned}$$

a contradiction. The first inequality holds by Jensen's inequality from $\rho(\cdot) < 0$ and the strict concavity of $\rho(\cdot)$ on $\mathcal{V}$ by Assumption A.2-GEL(b). The second inequality follows from $\lambda(\theta)' m(\theta) > 0$ since $m^j(\theta) = 0$, $j \in \{1, ..., b\}$, and $\lambda^j(\theta) \geq 0$ with at least one $\lambda^j(\theta) > 0$, $j \in \{b+1, ..., d_m\}$, from above. Hence, the associated auxiliary parameter $\lambda_j(\theta) = 0$, $j \in \{b+1, ..., d_m\}$, and $\mathbb{E}_{P_0}[\rho(\lambda(\theta)' m(z, \theta))]$ is maximised at $\rho(0)$ by setting $\lambda^j(\theta) = 0$, $j \in \{1, ..., d_m\}$. Therefore, $\hat{P}^{\rho}(\theta) = 0$ if $\theta \in \Theta_{P_0}$, i.e., $\Theta_{P_0} \subseteq \hat{\Theta}^{\rho}_{P_0}$.

To conclude, suppose $a \neq 0$, i.e., $\theta \in \Theta / \Theta_{P_0}$, and so there exists $j \in \{1, ..., a\}$ such that $m^j(\theta) < 0$. Now, as above, $\mathbb{E}_{P_0}[\rho(\lambda' m(z, \theta))] = 0$ and $\partial \mathbb{E}_{P_0}[\rho(\lambda' m(z, \theta)) - \rho(0)] / \partial \lambda^j = \mathbb{E}_{P_0}[\rho_1(\lambda' m(z, \theta)) m^j(z, \theta)] > 0$ at $\lambda = 0$. Define λ such that $\lambda^j = \epsilon$ for some small $\epsilon > 0$ and $\lambda^k = 0$ for $k \neq j$. Then, by continuity, $\hat{P}^{\rho}(\theta) \geq \mathbb{E}_{P_0}[\rho(\lambda^j m^j(z, \theta))] > 0$. Cf. Canay (2010, Proof of Lemma B.3, p.423). Hence, $\theta \in \Theta / \hat{\Theta}^{\rho}_{P_0}$, i.e., $\hat{\Theta}^{\rho}_{P_0} \subseteq \Theta_{P_0}$.■

Appendix E: Alternative GEL Criteria

A number of alternative but equivalent GEL criteria may also be defined.

Mirroring the GMM criterion (3.1), the introduction of the d_m-vector of complementary slackness parameters $\tau \geq 0$, cf. (2.3), directly into (3.4) defines the alternative GEL criterion

$$\tilde{P}_n^{\rho,a}(\theta, \lambda, \tau) = \sum_{i=1}^{n} \rho(\lambda'(m_i(\theta) - \tau))/n. \tag{E.1}$$

The GEL criterion $\tilde{P}_n^{\rho,a}(\theta, \lambda, \tau)$ (E.1) is then optimised over $\lambda \in \tilde{\Lambda}_n^a(\theta, \tau)$, where $\tilde{\Lambda}_n^a(\theta, \tau) = \{\lambda : \lambda'(m_i(\theta) - \tau) \in \mathcal{V}, i = 1, ..., n\}$, and $\tau \in \mathcal{T}$ for given $\theta \in \Theta$ with the slackness parameter space $\mathcal{T} = \{\tau \in \mathcal{R}^{d_m} : \tau \geq 0, \|\tau\| \leq C\}$ and $C > 0$ defined by the boundedness condition in Assumption A.1(b). The slackness parameter estimator $\tilde{\tau}_n^a(\theta)$ solves the corresponding f.o.c. with respect to τ, i.e., $\partial \tilde{P}_n^{\rho,a}(\theta, \tilde{\lambda}_n^a(\theta), \tilde{\tau}_n^a(\theta))/\partial\tau \geq 0$, $\tau \geq 0$. Now $\tilde{\lambda}_n^a(\theta) \geq 0$ since $\partial \tilde{P}_n^{\rho,a}(\theta, \tilde{\lambda}_n^a(\theta), \tilde{\tau}_n^a(\theta))/\partial\tau = -\sum_{i=1}^n \rho_1(\tilde{\lambda}_n^a(\theta)'(m_i(\theta) - \tilde{\tau}_n^a(\theta)))\tilde{\lambda}_n^a(\theta)/n$ and $\sum_{i=1}^n \rho_1(\tilde{\lambda}_n^a(\theta)'(m_i(\theta) - \tilde{\tau}_n^a(\theta))) < 0$ from Assumption A.2-GEL(b). In particular, either $\tilde{\lambda}_n^{a,j}(\theta) = 0$ and $\tilde{\tau}_n^{a,j}(\theta) > 0$ or $\tilde{\lambda}_n^{a,j}(\theta) > 0$ and $\tilde{\tau}_n^{a,j}(\theta) = 0$, $(j = 1, ..., d_m)$, and, thus, $\tilde{\lambda}_n^a(\theta)'\tilde{\tau}_n^a(\theta) = 0$. Hence, the auxiliary parameter constraint space $\tilde{\Lambda}_n^a(\theta, \tau)$ simplifies to $\hat{\Lambda}_n(\theta)$. The auxiliary parameter estimator $\tilde{\lambda}_n^a(\theta)$ solves the corresponding f.o.c. with respect to λ, i.e., $\sum_{i=1}^n \rho_1(\tilde{\lambda}_n^a(\theta)'(m_i(\theta) - \tilde{\tau}_n^a(\theta)))(m_i(\theta) - \tilde{\tau}_n^a(\theta))/n = 0$. Consequently, the slackness parameter estimator $\tilde{\tau}_n^a(\theta)$ satisfies

$$\begin{aligned}
\tilde{\tau}_n^a(\theta) &= \frac{\sum_{i=1}^n \rho_1(\tilde{\lambda}_n^a(\theta)'(m_i(\theta) - \tilde{\tau}_n^a(\theta)))m_i(\theta)}{\sum_{k=1}^n \rho_1(\tilde{\lambda}_n^a(\theta)'(m_k(\theta) - \tilde{\tau}_n^a(\theta)))} \\
&= \sum_{i=1}^n \hat{\pi}_i^\rho(\theta, \tilde{\lambda}_n^a(\theta))m_i(\theta),
\end{aligned}$$

since $\tilde{\lambda}_n^a(\theta)'\tilde{\tau}_n^a(\theta) = 0$; cf. (3.6). Therefore, $\tilde{\lambda}_n^a(\theta) = \hat{\lambda}_n(\theta)$ and, thus, $\tilde{P}_n^{\rho,a}(\theta, \tilde{\lambda}_n(\theta), \tilde{\tau}_n(\theta)) = \hat{P}_n^\rho(\theta)$.

Remark E.1. Note that $\lim_{n\to\infty} \mathcal{P}\{\tilde{\tau}_n^a(\theta) \in \mathcal{T}\} = 1$ since $\sup_{\theta\in\Theta} \|\hat{m}_n(\theta) - m(\theta)\| = o_p(1)$ by UWL from Assumption A.1(b). Thus, the upper bound C is not binding in $\mathcal{T}$ w.p.a.1.

Remark E.2. The GEL implied probabilities defined from (E.2),

$$\tilde{\pi}_i^{\rho,a}(\theta, \lambda, \tau) = \frac{\rho_1(\lambda'(m_i(\theta) - \tau))}{\sum_{k=1}^n \rho_1(\lambda'(m_k(\theta) - \tau))}, (i = 1, ..., n), \tag{E.2}$$

are non-negative and sum to unity. Moreover, since $\tilde{\lambda}_n^a(\theta)'\tilde{\tau}_n^a(\theta) = 0$, $\tilde{\pi}_i^{\rho,a}(\theta, \tilde{\lambda}_n^a(\theta), \tilde{\tau}_n^a(\theta)) = \hat{\pi}_i^\rho(\theta, \tilde{\lambda}_n^a(\theta))$, $(i = 1, ..., n)$.

The GEL criterion (E.1) may be re-centred by separating out the slackness parameter $\tau \geq 0$ to form

$$\tilde{P}_n^{\rho,b}(\theta, \lambda, \tau) = \sum_{i=1}^{n} [\rho(\lambda' m_i(\theta)) - \rho(\lambda' \tau)]/n, \tag{E.3}$$

which is then optimised over $\lambda \in \tilde{\Lambda}_n^b(\theta, \tau)$, where $\tilde{\Lambda}_n^b(\theta, \tau) = \{\lambda : \lambda' m_i(\theta) \in \mathcal{V}, i = 1, ..., n, \lambda' \tau \in \mathcal{V}\}$ and $\tau \in \mathcal{T}$ for given $\theta \in \Theta$. As above $\tilde{\lambda}_n^a(\theta) \geq 0$ since $\partial \tilde{P}_n^{\rho,b}(\theta, \tilde{\lambda}_n^b(\theta), \tilde{\tau}_n^b(\theta))/\partial \tau = -\rho_1(\tilde{\lambda}_n(\theta)' \tilde{\tau}_n(\theta)) \tilde{\lambda}_n(\theta) \geq 0$ noting $\rho_1(\tilde{\lambda}_n(\theta)' \tilde{\tau}_n(\theta)) < 0$ from Assumption A.2-GEL(b). Similarly, either $\tilde{\lambda}_n^{b,j}(\theta) = 0$ and $\tilde{\tau}_n^{b,j}(\theta) > 0$ or $\tilde{\lambda}_n^{b,j}(\theta) > 0$ and $\tilde{\tau}_n^{b,j}(\theta) = 0$, $(j = 1, ..., d_m)$, and, thus, $\tilde{\lambda}_n^b(\theta)' \tilde{\tau}_n^b(\theta) = 0$. Examining the f.o.c. with respect to λ, i.e., $\sum_{i=1}^n \rho_1(\tilde{\lambda}_n^b(\theta)' m_i(\theta)) m_i(\theta)/n - \rho_1(\tilde{\lambda}_n^b(\theta)' \tilde{\tau}_n^b(\theta)) \tilde{\tau}_n^b(\theta)) = 0$, the slackness parameter estimator $\tilde{\tau}_n^b(\theta)$ satisfies

$$\tilde{\tau}_n^b(\theta) = \frac{\sum_{i=1}^n \rho_1(\tilde{\lambda}_n^b(\theta)' m_i(\theta)) m_i(\theta)/n}{\rho_1(\tilde{\lambda}_n^b(\theta)' \tilde{\tau}_n^b(\theta))}. \tag{E.4}$$

Hence, the auxiliary parameter constraint space $\tilde{\Lambda}_n^b(\theta, \tau)$ is not fully binding and reduces to $\hat{\Lambda}_n(\theta)$ as previously. Consequently, $\tilde{P}_n^{\rho,b}(\theta, \tilde{\lambda}_n^b(\theta), \tilde{\tau}_n^b(\theta)) = \hat{P}_n^{\rho}(\theta)$.

Remark E.3. Noting $\rho_1(0) = -1$, since $\tilde{\lambda}_n^b(\theta)' \tilde{\tau}_n^b(\theta)$, the slackness parameter estimator (E.4) $\tilde{\tau}_n^b(\theta) = -\sum_{i=1}^n \rho_1(\tilde{\lambda}_n^b(\theta)' m_i(\theta)) m_i(\theta)/n$; cf. (3.6). The GEL implied probabilities implicitly defined from (E.4) as $\rho_1(\lambda' m_i(\theta))/n\rho_1(\theta' \tau)$, $(i = 1, ..., n)$, although non-negative by Assumption A.2-GEL(b), do not sum to unity. Even if evaluated at $\tilde{\lambda}_n^b(\theta)$ and $\tilde{\tau}_n^b(\theta)$, the GEL implied probabilities $-\rho_1(\tilde{\lambda}_n^b(\theta)' m_i(\theta))/n$, $(i = 1, ..., n)$, do not sum to unity. Exploiting (3.7) guarantees non-negativity and unit summability, i.e.,

$$\tilde{\pi}_i^{\rho,b}(\theta, \lambda, \tau) = \frac{\rho_1(\lambda' m_i(\theta))}{\sum_{k=1}^n \rho_1(\lambda' m_k(\theta))}, (i = 1, ..., n). \tag{E.5}$$

Moreover, $\tilde{\pi}_i^{\rho,b}(\theta, \tilde{\lambda}_n^b(\theta), \tilde{\tau}_n^b(\theta)) = \hat{\pi}_i^{\rho}(\theta, \tilde{\lambda}_n^b(\theta))$, $(i = 1, ..., n)$.

Consider the Lagrangean

$$\tilde{P}_n^{\rho}(\theta, \lambda, \tau) = \sum_{i=1}^{n} \rho(\lambda' m_i(\theta))/n + \lambda' \tau \tag{E.6}$$

in which the slackness parameter vector τ now denotes a d_m-vector of Lagrange multipliers associated with the inequality constraint $\lambda \geq 0$; cf. $G^*(\theta, v, \lambda)$ defined in Moon and Schorfheide (2009, eq. (16), p.140). Here the GEL criterion $\tilde{P}_n^{\rho}(\theta, \lambda, \tau)$ (E.6) is optimised over $\lambda \in \hat{\Lambda}_n(\theta)$ and $\tau \in \mathcal{T}$ for given $\theta \in \Theta$. The Lagrange multiplier parameter estimator $\tilde{\tau}_n(\theta)$ satisfies $\partial \tilde{P}_n^\rho(\theta, \tilde{\lambda}_n(\theta), \tilde{\tau}_n(\theta))/\partial \tau \geq 0$, $\tau \geq 0$. Thus, the auxiliary parameter estimator $\tilde{\lambda}_n(\theta) \geq 0$ as $\partial \tilde{P}_n^\rho(\theta, \tilde{\lambda}_n(\theta), \tilde{\tau}_n(\theta))/\partial \tau = \tilde{\lambda}_n(\theta)$. Moreover, $\tilde{\lambda}_n(\theta)'\tilde{\tau}_n(\theta) = 0$ with, in particular, $\tilde{\lambda}_n^j(\theta) = 0$ and $\tilde{\tau}_n^j(\theta) > 0$ or $\tilde{\lambda}_n^j(\theta) > 0$ and $\tilde{\tau}_n^j(\theta) = 0$, $(j = 1, ..., d_m)$. From the f.o.c. with respect to λ, i.e., $\sum_{i=1}^n \rho_1(\tilde{\lambda}_n(\theta)' m_i(\theta)) m_i(\theta)/n + \tilde{\tau}_n(\theta) = 0$, the Lagrange multiplier estimator $\hat{\tau}_n(\theta) \geq 0$ satisfies

$$\tilde{\tau}_n(\theta) = -\sum_{i=1}^n \rho_1(\tilde{\lambda}_n(\theta)' m_i(\theta)) m_i(\theta)/n, \tag{E.7}$$

cf. (3.6). Substituting $\tilde{\lambda}_n(\theta)$ and $\tilde{\tau}_n(\theta)$, $\tilde{P}_n^\rho(\theta, \tilde{\lambda}_n(\theta), \tilde{\tau}_n(\theta)) = \hat{P}_n^\rho(\theta, \tilde{\lambda}_n(\theta))$. Therefore, from the strict concavity of $\rho(\cdot)$ on $\mathcal{V}$ by Assumption A.2-GEL(b), $\tilde{\lambda}_n(\theta) = \hat{\lambda}_n(\theta)$ and, likewise, $\tilde{P}_n^\rho(\theta, \tilde{\lambda}_n(\theta), \tilde{\tau}_n(\theta)) = \hat{P}_n^\rho(\theta)$.

Remark E.4. The GEL implied probabilities defined from (E.7) as $-\rho_1(\tilde{\lambda}_n(\theta)' m_i(\theta))/n$, $(i = 1, ..., n)$, are non-negative by Assumption A.2-GEL(b) but do not sum to unity. The redefinition $\tilde{\pi}_i^\rho(\theta, \tilde{\lambda}_n(\theta), \tilde{\tau}_n(\theta)) = \rho_1(\tilde{\lambda}_n(\theta)' m_i(\theta))/\sum_{k=1}^n \rho_1(\tilde{\lambda}_n(\theta)' m_k(\theta))$ guarantees non-negativity and unit summability; cf. Remark D.3. Moreover, $\tilde{\pi}_i^\rho(\theta, \tilde{\lambda}_n(\theta), \tilde{\tau}_n(\theta)) = \hat{\pi}_i^\rho(\theta, \hat{\lambda}_n(\theta))$, $(i = 1, ..., n)$.

E.1 GEL Estimator Equivalence

Lemma E.1. *The solutions to the saddle point problems (3.4) and (E.1) are identical, i.e.,* **(a)** *if $(\tilde{\lambda}(\theta), \tilde{\tau}(\theta))$, where $\tilde{\tau}(\theta) \in int(\mathcal{T})$, is a saddlepoint of $\tilde{P}_n^{\rho,a}(\theta, \lambda, \tau)$ then $\tilde{\lambda}(\theta)$ is also a maximiser of $\hat{P}_n^\rho(\theta, \lambda)$;* **(b)** *if $\hat{\lambda}(\theta)$ is a maximiser of $\hat{P}_n^\rho(\theta, \lambda)$ and $\hat{\tau}(\theta) \in int(\mathcal{T})$, where $\hat{\tau}^j(\theta) = \sum_{i=1}^n \hat{\pi}_i^\rho(\theta, \hat{\lambda}(\theta)) m_i^j(\theta)$ if $\hat{\lambda}^j(\theta) = 0$ and 0 if $\hat{\lambda}^j(\theta) > 0$, $(j = 1, ..., d_m)$, then $(\hat{\lambda}(\theta), \hat{\tau}(\theta))$ is a saddlepoint of $\tilde{P}_n^{\rho,c}(\theta, \lambda, \tau)$.*

Proof. To prove (a), first note that $\tilde{\lambda} \geq 0$, since the solution $\tilde{\tau}$ satisfies $\partial \tilde{P}_n^{\rho,a}(\theta, \tilde{\lambda}, \tilde{\tau})/\partial \tau \geq 0$ and $\partial \tilde{P}_n^{\rho,a}(\theta, \lambda, \tau)/\partial \tau = -\sum_{i=1}^n \rho_1(\lambda'(m_i(\theta) - \tau))\lambda/n$ with $\sum_{i=1}^n \rho_1(\lambda'(m_i(\theta) - \tau))/n < 0$ by Assumption A.2-GEL(b). In particular, $\tilde{\lambda}^j = 0$ and $\tilde{\tau}^j > 0$ or $\tilde{\lambda}^j > 0$ and $\tilde{\tau}^j = 0$, $(j = 1, ..., d_m)$, and $\tilde{\lambda}'\tilde{\tau} = 0$. The solution $\tilde{\lambda}$ satisfies $\partial \tilde{P}_n^{\rho,a}(\theta, \tilde{\lambda}, \tilde{\tau})/\partial \lambda = 0$, i.e., $\sum_{i=1}^n \rho_1(\tilde{\lambda}'(m_i(\theta) - \tilde{\tau}))(m_i(\theta) - \tilde{\tau})/n = 0$, and, thus, $\tilde{\tau}^j = \sum_{i=1}^n \rho_1(\tilde{\lambda}' m_i(\theta)) m_i^j(\theta)/\sum_{k=1}^n \rho_1(\tilde{\lambda}' m_k(\theta)) > 0$ if $\tilde{\lambda}^j = 0$ or 0 if $\tilde{\lambda}^j > 0$, $(j = 1, ..., d_m)$. Now $\tilde{P}_n^{\rho,a}(\theta, \lambda, \tilde{\tau}) = \hat{P}_n^\rho(\theta, \lambda) - \sum_{i=1}^n \rho_1(\lambda'(m_i(\theta) - \tau_*))\lambda'\tilde{\tau}/n \geq \hat{P}_n^\rho(\theta, \lambda)$ for $\tau_* \in (0, \tilde{\tau})$ since $\lambda \geq 0$ and $\sum_{i=1}^n \rho_1(\lambda'(m_i(\theta) - \tau_*)) < 0$ by Assumption A.2-GEL(b). Therefore, from the saddlepoint property with respect to λ,

$\hat{P}_n^\rho(\theta,\tilde{\lambda}) = \tilde{P}_n^{\rho,a}(\theta,\tilde{\lambda},\tilde{\tau}) \geq \tilde{P}_n^{\rho,a}(\theta,\lambda,\tilde{\tau}) \geq \hat{P}_n^\rho(\theta,\lambda)$.

For (b), $\hat{\lambda}'\hat{\tau} = 0$ from the definition of $\hat{\tau}$ with $\hat{\tau}^j = 0$ and $\hat{\lambda}^j > 0$ or $\hat{\tau}^j > 0$ and $\hat{\lambda}^j = 0$ $(j = 1, ..., d_m)$, from the first order condition $\partial\hat{P}_n^\rho(\theta,\lambda)/\partial\lambda \leq 0$, $\lambda \geq 0$, cf. (3.6). For the saddle point property with respect to $\tau \geq 0$, $\tilde{P}_n^{\rho,a}(\theta,\hat{\lambda},\hat{\tau}) = \hat{P}_n^\rho(\theta,\hat{\lambda}) \leq \tilde{P}_n^{\rho,a}(\theta,\hat{\lambda},\tau)$ since $\hat{\lambda} \geq 0$ and $\tilde{P}_n^{\rho,a}(\theta,\hat{\lambda},\tau) = \hat{P}_n^\rho(\theta,\hat{\lambda}) - \sum_{i=1}^n \rho_1(\hat{\lambda}'(m_i(\theta)-\tau_*))\hat{\lambda}'\tau/n$ for $\tau_* \in (0,\tau)$ with $\sum_{i=1}^n \rho_1(\hat{\lambda}'(m_i(\theta)-\tau_*))/n < 0$ from Assumption A.2-GEL(b). For λ, $\tilde{P}_n^{\rho,a}(\theta,\hat{\lambda},\hat{\tau}) = \hat{P}_n^\rho(\theta,\hat{\lambda}) \geq \tilde{P}_n^{\rho,a}(\theta,\lambda,\hat{\tau})$ since, noting $\partial\tilde{P}_n^{\rho,a}(\theta,\hat{\lambda},\hat{\tau})/\partial\lambda = 0$, $\hat{P}_n^{\rho,a}(\theta,\lambda,\hat{\tau}) = \tilde{P}_n^{\rho,a}(\theta,\hat{\lambda},\hat{\tau}) + \sum_{i=1}^n \rho_2(\lambda_*'(m_i(\theta)-\hat{\tau}))[(m_i(\theta)-\hat{\tau})'(\lambda-\hat{\lambda})]^2/2n \leq \tilde{P}_n^{\rho,a}(\theta,\hat{\lambda},\hat{\tau})$ for $\lambda_* \in (\lambda,\hat{\lambda})$ and $\rho_2(\cdot) < 0$ by the concavity of $\rho(\cdot)$ from Assumption A.2-GEL(b).■

Lemma E.2. *The solutions to the saddle point problems (3.4) and (E.3) are identical, i.e.,* **(a)** *if $(\tilde{\lambda}(\theta),\tilde{\tau}(\theta))$, where $\tilde{\tau}(\theta) \in int(\mathcal{T})$, is a saddlepoint of $\tilde{P}_n^{\rho,b}(\theta,\lambda,\tau)$ then $\tilde{\lambda}(\theta)$ is also a maximiser of $\hat{P}_n^\rho(\theta,\lambda)$;* **(b)** *if $\hat{\lambda}(\theta)$ is a maximiser of $\hat{P}_n^\rho(\theta,\lambda)$ and $\hat{\tau}(\theta) \in int(\mathcal{T})$, where $\hat{\tau}^j(\theta) = -\sum_{i=1}^n \rho_1(\hat{\lambda}(\theta)'m(\theta))m_i^j(\theta)/n$ if $\hat{\lambda}^j(\theta) = 0$ and 0 if $\hat{\lambda}^j(\theta) > 0$, $(j = 1, ..., d_m)$, then $(\hat{\lambda}(\theta),\hat{\tau}(\theta))$ is a saddlepoint of $\tilde{P}_n^{\rho,b}(\theta,\lambda,\tau)$.*

Proof. The proof follows on similar lines to that for Lemma E.1.

For (a), $\tilde{\lambda} \geq 0$ since $\tilde{\tau}$ satisfies $\partial\tilde{P}_n^{\rho,b}(\theta,\tilde{\lambda},\tilde{\tau})/\partial\tau \geq 0$ and $\partial\tilde{P}_n^{\rho,b}(\theta,\lambda,\tau)/\partial\tau = -\rho_1(\lambda'\tau)\lambda$ with $\rho_1(\lambda'\tau) < 0$ by Assumption A.2-GEL(b). Likewise, $\tilde{\lambda}'\tilde{\tau} = 0$ with $\tilde{\lambda}^j = 0$ and $\tilde{\tau}^j > 0$ or $\tilde{\lambda}^j > 0$ and $\tilde{\tau}^j = 0$, $(j = 1, ..., d_m)$. In this case $\tilde{\lambda}$ satisfies $\partial\tilde{P}_n^{\rho,b}(\theta,\tilde{\lambda},\tilde{\tau})/\partial\lambda = 0$, i.e., $\sum_{i=1}^n \rho_1(\tilde{\lambda}'m_i(\theta))m_i(\theta)/n - \rho_1(\tilde{\lambda}'\tilde{\tau})\tilde{\tau} = 0$, and, thus, $\tilde{\tau}^j = -\sum_{i=1}^n \rho_1(\tilde{\lambda}'m_i(\theta))m_i^j(\theta)/n$ if $\tilde{\lambda}^j = 0$ or 0 if $\tilde{\lambda}^j > 0$, $(j = 1, ..., d_m)$, from the normalisation $\rho_1(0) = -1$ of Remark 3.3. Now $\tilde{P}_n^{\rho,b}(\theta,\lambda,\tilde{\tau}) = \hat{P}_n^\rho(\theta,\lambda) - \rho_1((\lambda'\tilde{\tau})_*)\lambda'\tilde{\tau} \geq \hat{P}_n^\rho(\theta,\lambda)$ for $(\lambda'\tilde{\tau})_* \in (0,(\lambda'\tilde{\tau}))$ since $\lambda \geq 0$ and $\rho_1((\lambda'\tilde{\tau})_*) < 0$ by Assumption A.2-GEL(b). Therefore, $\hat{P}_n^\rho(\theta,\tilde{\lambda}) = \tilde{P}_n^{\rho,b}(\theta,\tilde{\lambda},\tilde{\tau}) \geq \tilde{P}_n^{\rho,b}(\theta,\lambda,\tilde{\tau}) \geq \hat{P}_n^\rho(\theta,\lambda)$.

For the proof of (b), as in the Proof of Lemma E.1(b), $\hat{\lambda}'\hat{\tau} = 0$ with $\tilde{\tau}^j = 0$ and $\hat{\lambda}^j > 0$ or $\tilde{\tau}^j > 0$ and $\hat{\lambda}^j = 0$, $(j = 1, ..., d_m)$. For the saddle point property with respect to $\tau \geq 0$, $\tilde{P}_n^{\rho,b}(\theta,\hat{\lambda},\hat{\tau}) = \hat{P}_n^\rho(\theta,\hat{\lambda}) \leq \tilde{P}_n^{\rho,b}(\theta,\hat{\lambda},\tau)$ since $\hat{\lambda} \geq 0$ and $\tilde{P}_n^{\rho,b}(\theta,\hat{\lambda},\tau) = \hat{P}_n^\rho(\theta,\hat{\lambda}) - \rho_1(\hat{\lambda}'\tau_*)\hat{\lambda}'\tau$ for $\tau_* \in (0,\tau)$ with $\rho_1(\hat{\lambda}'\tau_*) < 0$. For λ, $\tilde{P}_n^{\rho,b}(\theta,\hat{\lambda},\hat{\tau}) = \hat{P}_n^\rho(\theta,\hat{\lambda}) \geq \tilde{P}_n^{\rho,b}(\theta,\lambda,\hat{\tau})$ since, noting $\partial\tilde{P}_n^{\rho,b}(\theta,\hat{\lambda},\hat{\tau})/\partial\lambda = 0$ and $\rho_2(\cdot) < 0$, $\hat{P}_n^{\rho,b}(\theta,\lambda,\hat{\tau}) = \tilde{P}_n^{\rho,b}(\theta,\hat{\lambda},\hat{\tau}) + \sum_{i=1}^n \rho_2(\lambda_*'m_i(\theta))[m_i(\theta)' \times(\lambda-\hat{\lambda})]^2/2n + \sum_{i=1}^n \rho_2(\lambda_*'\hat{\tau})[\hat{\tau}'(\lambda-\hat{\lambda})]^2/2n \leq \tilde{P}_n^{\rho,a}(\theta,\hat{\lambda},\hat{\tau})$ for $\lambda_* \in (\lambda,\hat{\lambda})$.■

Lemma E.3. *The solutions to the saddle point problems (3.4) and (E.6) are identical, i.e.,* **(a)** *if $(\tilde{\lambda}(\theta),\tilde{\tau}(\theta))$, where $\tilde{\tau}(\theta) \in int(\mathcal{T})$, is a saddlepoint of $\tilde{P}_n^\rho(\theta,\lambda,\tau)$ then $\tilde{\lambda}(\theta)$ is also a maximiser of $\hat{P}_n^\rho(\theta,\lambda)$;* **(b)** *if $\hat{\lambda}(\theta)$ is a maximiser of $\hat{P}_n^\rho(\theta,\lambda)$ and $\hat{\tau}(\theta) \in int(\mathcal{T})$, where $\hat{\tau}^j(\theta) = -\sum_{i=1}^n \rho_1(\hat{\lambda}(\theta)'m_i(\theta))m_i^j(\theta)/n$ if $\hat{\lambda}^j(\theta) = 0$ and 0 if $\hat{\lambda}^j(\theta) > 0$,*

$(j = 1, ..., d_m)$, *then* $(\hat{\lambda}(\theta), \hat{\tau}(\theta))$ *is a saddlepoint of* $\tilde{P}_n^\rho(\theta, \lambda, \tau)$.

Proof. The proof again follows along similar lines to the Proof of Lemma E.1. Cf. Moon and Schorfheide (2009, Lemma A.1, p.150).

For (a), $\tilde{\lambda} \geq 0$ since $\tilde{\tau}$ satisfies $\partial \tilde{P}_n^\rho(\theta, \tilde{\lambda}, \tilde{\tau})/\partial\tau \geq 0$ and $\partial \tilde{P}_n^\rho(\theta, \lambda, \tau)/\partial\tau = \lambda$. Likewise, $\tilde{\lambda}'\tilde{\tau} = 0$ with $\tilde{\lambda}^j = 0$ and $\tilde{\tau}^j > 0$ or $\tilde{\lambda}^j > 0$ and $\tilde{\tau}^j = 0$, $(j = 1, ..., d_m)$. In this case $\tilde{\lambda}$ satisfies $\partial \tilde{P}_n^{\rho,b}(\theta, \tilde{\lambda}, \tilde{\tau})/\partial\lambda = 0$, i.e., $\sum_{i=1}^n \rho_1(\tilde{\lambda}' m_i(\theta)) m_i(\theta)/n + \tilde{\tau} = 0$, and, thus, $\tilde{\tau}^j = -\sum_{i=1}^n \rho_1(\tilde{\lambda}' m_i(\theta)) m_i^j(\theta)/n$ if $\tilde{\lambda}^j = 0$ or 0 if $\tilde{\lambda}^j > 0$, $(j = 1, ..., d_m)$. Now $\tilde{P}_n^\rho(\theta, \lambda, \tilde{\tau}) = \hat{P}_n^\rho(\theta, \lambda) + \lambda'\tilde{\tau} \geq \hat{P}_n^\rho(\theta, \lambda)$ since $\lambda \geq 0$. Therefore, $\hat{P}_n^\rho(\theta, \tilde{\lambda}) = \tilde{P}_n^\rho(\theta, \tilde{\lambda}, \tilde{\tau}) \geq \tilde{P}_n^\rho(\theta, \lambda, \tilde{\tau}) \geq \hat{P}_n^\rho(\theta, \lambda)$.

For (b), $\hat{\lambda}'\hat{\tau} = 0$ with $\tilde{\tau}^j = 0$ and $\hat{\lambda}^j > 0$ or $\tilde{\tau}^j > 0$ and $\hat{\lambda}^j = 0$, $(j = 1, ..., d_m)$, from the first order condition $\partial \hat{P}_n^\rho(\theta, \lambda)/\partial\lambda \leq 0$, $\lambda \geq 0$. For the saddle point property with respect to $\tau \geq 0$, $\tilde{P}_n^\rho(\theta, \hat{\lambda}, \hat{\tau}) = \hat{P}_n^\rho(\theta, \hat{\lambda}) \leq \tilde{P}_n^\rho(\theta, \hat{\lambda}, \tau)$ since $\hat{\lambda} \geq 0$. For λ, $\tilde{P}_n^\rho(\theta, \hat{\lambda}, \hat{\tau}) = \hat{P}_n^\rho(\theta, \hat{\lambda}) \geq \tilde{P}_n^\rho(\theta, \lambda, \hat{\tau})$ since, noting $\partial \tilde{P}_n^\rho(\theta, \hat{\lambda}, \hat{\tau})/\partial\lambda = 0$ and $\rho_2(\cdot) < 0$, $\tilde{P}_n^\rho(\theta, \lambda, \hat{\tau}) = \tilde{P}_n^\rho(\theta, \hat{\lambda}, \hat{\tau}) + \sum_{i=1}^n \rho_2(\lambda_*' m_i(\theta))[m_i(\theta)'(\lambda - \hat{\lambda})]^2/2n \leq \tilde{P}_n^{\rho,a}(\theta, \hat{\lambda}, \hat{\tau})$ for $\lambda_* \in (\lambda, \hat{\lambda})$.■

E.2 Identified Set

Alternative but equivalent population versions of the GEL identified set $\hat{\Theta}_{P_0}^\rho$ (3.9), cf. Canay (2010) for EL, may be defined corresponding to the alternative GEL criteria $\tilde{P}_n^\rho(\theta, \lambda, \tau)$ (E.6) and $\tilde{P}_n^{\rho,k}(\theta, \lambda, \tau)$, $(k = a, b)$, (E.1), (E.3), described in Appendix E.1. The respective population criteria are defined by $\tilde{P}^\rho(\theta) = \inf_{\tau\in\mathcal{T}} \sup_{\lambda\in\mathcal{R}^{d_m}} \tilde{P}^\rho(\theta, \lambda, \tau)$ with $\tilde{P}^\rho(\theta, \lambda, \tau) = \mathbb{E}_{P_0}[\rho(\lambda' m(z, \theta))] + \lambda'\tau$ and $\tilde{P}^{\rho,k}(\theta) = \inf_{\tau\in\mathcal{T}} \sup_{\lambda\in\mathcal{R}^{d_m}} \tilde{P}^{\rho,k}(\theta, \lambda, \tau)$, $(k = a, b)$, with $\tilde{P}^{\rho,a}(\theta, \lambda, \tau) = \mathbb{E}_{P_0}[\rho(\lambda'(m(z, \theta) - \tau))]$ and $\tilde{P}^{\rho,b}(\theta, \lambda, \tau) = \mathbb{E}_{P_0}[\rho(\lambda' m(z, \theta)) - \rho(\lambda'\tau)]$. The respective GEL population counterparts to identified set Θ_{P_0} are

$$\tilde{\Theta}_{P_0}^\rho = \{\theta \in \Theta : \theta = \arg\min_{\theta\in\Theta} \tilde{P}^\rho(\theta)\}, \tag{E.8}$$

and

$$\tilde{\Theta}_{P_0}^{\rho,k} = \{\theta \in \Theta : \theta = \arg\min_{\theta\in\Theta} \tilde{P}^{\rho,k}(\theta)\}, (k = a, b). \tag{E.9}$$

Recall the partition of the index set $\{1, ..., d_m\}$ according to $m^j(\theta) < 0$, $(j = 1, ..., a)$, $m^j(\theta) = 0$, $(j = a+1, ..., a+b)$ and $m^j(\theta) > 0$, $(j = a+b+1, ..., d_m)$. Let $c = d_m - a - b$. Note again that a, b and thus c depend on θ. Also recall the notation $m(\theta) = \mathbb{E}_{P_0}[m(z, \theta)]$.

Recall $\hat{\Theta}_{P_0}^\rho = \{\theta \in \Theta : \theta = \arg\min_{\theta\in\Theta} \hat{P}^\rho(\theta)\}$ (3.9).

Lemma E.4. *Suppose that Assumptions A.1 and A.2-GEL are satisfied. Then* $\tilde{\Theta}_{P_0}^{\rho,k} = \Theta_{P_0}$, $(k = a, b)$.

Proof. Let $\hat{P}^\rho(\theta, \lambda) = \mathbb{E}_{P_0}[\rho(\lambda' m(z, \theta))]$.

First, consider the alternative GEL population criterion $\tilde{P}^{\rho,a}(\theta, \lambda, \tau) = \mathbb{E}_{P_0}[\rho(\lambda'(m(z, \theta) - \tau))]$ corresponding to (E.1). Now $\tilde{P}^{\rho,a}(\theta) = \inf_{\tau \in \mathcal{T}} \sup_{\lambda \in \mathcal{R}^{d_m}} \tilde{P}^{\rho,a}(\theta, \lambda, \tau)$. The solution $\tau(\theta) \geq 0$ satisfies $\partial \tilde{P}^{\rho,a}(\theta, \lambda, \tau)/\partial \tau \geq 0$. Thus, since $\partial \tilde{P}^{\rho,a}(\theta, \lambda, \tau)/\partial \tau = -\mathbb{E}_{P_0}[\rho_1(\lambda'(m(z, \theta) - \tau))]\lambda$ and $\mathbb{E}_{P_0}[\rho_1(\lambda'(m(z, \theta) - \tau))] < 0$ by Assumption A.2-GEL(b), $\lambda^j \geq 0$, $(j = 1, ..., d_m)$, and $\lambda' \tau(\theta) = 0$. The solution $\lambda(\theta)$ satisfies $\partial \tilde{P}^{\rho,a}(\theta, \lambda, \tau)/\partial \lambda = 0$, i.e., $\mathbb{E}_{P_0}[\rho_1(\lambda(\theta)'(m(z, \theta) - \tau(\theta)))(m(z, \theta) - \tau(\theta))] = 0$ and, thus,

$$\begin{aligned} \tau(\theta) &= \frac{\mathbb{E}_{P_0}[\rho_1(\lambda(\theta)'(m(z, \theta) - \tau(\theta)))m(z, \theta)]}{\mathbb{E}_{P_0}[\rho_1(\lambda(\theta)'(m(z, \theta) - \tau(\theta)))]} \\ &= \frac{\mathbb{E}_{P_0}[\rho_1(\lambda(\theta)' m(z, \theta))m(z, \theta)]}{\mathbb{E}_{P_0}[\rho_1(\lambda(\theta)' m(z, \theta))]} \geq 0. \end{aligned}$$

Now $\tilde{P}^{\rho,a}(\theta, \lambda, \tau(\theta)) = \hat{P}^\rho(\theta, \lambda) - \mathbb{E}_{P_0}[\rho_1(\lambda'(m(z, \theta) - \tau_*))]\lambda' \tau(\theta) \geq \hat{P}^\rho(\theta, \lambda)$ for $\tau_* \in (0, \tau(\theta))$ since $\lambda \geq 0$ and $\mathbb{E}_{P_0}[\rho_1(\lambda'(m(z, \theta) - \tau_*))] < 0$ by Assumption A.2-GEL(b). Hence, as $\hat{P}^\rho(\theta, \lambda(\theta)) = \tilde{P}^{\rho,a}(\theta, \lambda(\theta), \tau(\theta)) \geq \tilde{P}^{\rho,a}(\theta, \lambda, \tau(\theta))$, $\hat{P}^\rho(\theta, \lambda(\theta)) \geq \hat{P}^\rho(\theta, \lambda)$. That is, $\lambda(\theta)$ also optimises the GEL criterion $\hat{P}^\rho(\theta, \lambda)$ (3.4) and therefore $\tilde{\Theta}_{P_0}^{\rho,a} = \hat{\Theta}_{P_0}^\rho$.

Secondly, the population criterion for the GEL criterion (E.3) is $\tilde{P}^{\rho,b}(\theta, \lambda, \tau) = \mathbb{E}_{P_0}[\rho(\lambda' m(z, \theta)) - \rho(\lambda' \tau)]$ with $\tilde{P}^{\rho,b}(\theta) = \min_{\tau \in \mathcal{T}} \sup_{\lambda \in \mathcal{R}^{d_m}} \tilde{P}^\rho(\theta, \lambda, \tau)$. The solution $\tau(\theta) \geq 0$ satisfies $\partial \tilde{P}^{\rho,b}(\theta, \lambda, \tau)/\partial \tau \geq 0$. Likewise, since $\partial \tilde{P}^{\rho,b}(\theta, \lambda, \tau)/\partial \tau = -\rho_1(\lambda' \tau)\lambda$, by Assumption A.2-GEL(b), $\lambda^j(\theta) \geq 0$, $(j = 1, ..., d_m)$, and $\lambda' \tau(\theta) = 0$ as above. The solution $\lambda(\theta)$ satisfies $\partial \tilde{P}^{\rho,b}(\theta, \lambda, \tau)/\partial \lambda = 0$, i.e., $\mathbb{E}_{P_0}[\rho_1(\lambda(\theta)' m(z, \theta))m(z, \theta)] - \rho_1(\lambda(\theta)' \tau(\theta))\tau(\theta) = 0$ or

$$\tau(\theta) = \frac{\mathbb{E}_{P_0}[\rho_1(\lambda(\theta)' m(z, \theta))m(z, \theta)]}{\rho_1(\lambda(\theta)' \tau(\theta))} \geq 0.$$

By similar reasoning $\tilde{P}^{\rho,b}(\theta, \lambda, \tau(\theta)) = \hat{P}^\rho(\theta, \lambda) - \rho_1(\lambda' \tau_*)\lambda' \tau(\theta) \geq \hat{P}^\rho(\theta, \lambda)$ for $\tau_* \in (0, \tau(\theta))$ since $\lambda \geq 0$ and $\rho_1(\cdot) < 0$ by Assumption A.2-GEL(b). Hence, as $\hat{P}^\rho(\theta, \lambda(\theta)) = \tilde{P}^{\rho,b}(\theta, \lambda(\theta), \tau(\theta)) \geq \tilde{P}^{\rho,b}(\theta, \lambda, \tau(\theta))$, $\hat{P}^\rho(\theta, \lambda(\theta)) \geq \hat{P}^\rho(\theta, \lambda)$, i.e., $\lambda(\theta)$ also optimises the GEL criterion $\hat{P}^\rho(\theta, \lambda)$ (3.4).Therefore, $\tilde{\Theta}_{P_0}^{\rho,b} = \hat{\Theta}_{P_0}^\rho$.■

Lemma E.5. *Suppose that Assumptions A.1 and A.2-GEL are satisfied. Then* $\tilde{\Theta}_{P_0}^\rho = \Theta_{P_0}$.

Proof. The population criterion corresponding to the alternative sample GEL criterion (E.6) is given by $\tilde{P}^{\rho}(\theta,\lambda,\tau)=\mathbb{E}_{P_0}[\rho(\lambda'm(z,\theta))-\rho(0)]+\lambda'\tau$ with $\tilde{P}^{\rho}(\theta)=\inf_{\tau\in\mathcal{T}}\sup_{\lambda\in\mathcal{R}^{d_m}}\tilde{P}^{\rho}(\theta,\lambda,\tau)$. The solution $\tau(\theta)$ satisfies $\partial\tilde{P}^{\rho}(\theta,\lambda,\tau)/\partial\tau\geq 0$. Thus, since $\partial\tilde{P}^{\rho}(\theta,\lambda,\tau)/\partial\tau=\lambda$, $\lambda^j\geq 0$, $(j=1,...,d_m)$, and $\lambda'\tau(\theta)=0$. The solution $\lambda(\theta)$ satisfies $\partial\tilde{P}^{\rho}(\theta,\lambda,\tau)/\partial\lambda=0$, i.e., $\mathbb{E}_{P_0}[\rho_1(\lambda'm(z,\theta))m(z,\theta)]+\tau(\theta)=0$ or

$$\tau(\theta)=-\mathbb{E}_{P_0}[\rho_1(\lambda'm(z,\theta))m(z,\theta)]\geq 0.$$

Now $\tilde{P}^{\rho}(\theta,\lambda,\tau(\theta))=\hat{P}^{\rho}(\theta,\lambda)+\lambda'\tau(\theta)\geq\hat{P}^{\rho}(\theta,\lambda)$ since $\lambda\geq 0$. Hence, as $\hat{P}^{\rho}(\theta,\lambda(\theta))=\tilde{P}^{\rho}(\theta,\lambda(\theta),\tau(\theta))\geq\tilde{P}^{\rho}(\theta,\lambda,\tau(\theta))$, $\hat{P}^{\rho}(\theta,\lambda(\theta))\geq\hat{P}^{\rho}(\theta,\lambda)$, i.e., $\lambda(\theta)$ also optimises the GEL criterion $\hat{P}^{\rho}(\theta,\lambda)$ (3.4).Therefore, $\tilde{\Theta}^{\rho}_{P_0}=\hat{\Theta}^{\rho}_{P_0}$.■

References

Andrews, D.W.K., and P.J. Barwick (2012). "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure." *Econometrica*, 80, 2805–2826.

Andrews, D.W.K., and P. Guggenberger (2009)."Validity of Subsampling and "Plug-In Asymptotic" Inference for Parameters Defined by Moment Inequalities." *Econometric Theory*, 25, 669-709.

Andrews, D.W.K., and G. Soares (2010). "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection." *Econometrica*, 78, 119–157.

Andrews, D.W.K., and X. Shi (2013). "Inference Based on Conditional Moment Inequalities." *Econometrica* 81, 609–666.

Andrews, D.W.K., and X. Shi (2014). "Nonparametric Inference Based on Conditional Moment Inequalities." *Journal of Econometrics*, 179, 31–45.

Armstrong, T.B. (2014). "Weighted KS Statistics for Inference on Conditional Moment Inequalities." *Journal of Econometrics*, 181, 92–116.

Armstrong, T.B. (2015). "Asymptotically Exact Inference in Conditional Moment Inequality Models." *Journal of Econometrics*, 186, 51–85.

Armstrong, T.B., and H.P. Chan (2016). "Multiscale Adaptive Inference on Conditional Moment Inequalities." *Journal of Econometrics*, 194, 24–43.

Back, K., and D.P. Brown (1993). "Implied Probabilities in GMM Estimators." *Econometrica*, 61, 971-975.

Brown, B.W., and W.K. Newey (1992). "Bootstrapping for GMM." Working paper, MIT.

Brown, B.W., and W.K. Newey (2002). "Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference." *Journal of Business and Economic Statistics*, 20, 507-51

Bugni, F.A., Canay, I.A., and P. Guggenberger (2012). "Distortions of Asymptotic Confidence Size in Locally Misspecified Moment Inequality Models." *Econometrica*, 80, 1741-1768.7.

Beresteanu, A., Molchanov, I., and F. Molinari (2011). "Sharp Identification Regions in Models with Convex Moment Predictions." *Econometrica*, 79, 1785–1821.

Beresteanu, A., and F. Molinari (2008). "Asymptotic Properties for a Class of Partially Identified Models.' *Econometrica*, 76, 763–814.

Canay, I.A. (2010). "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity." *Journal of Econometrics* 156, 408-425.

Chen, X., Christensen, T.M., and E. Tamer (2016). "MCMC Confidence Sets for Identified Sets." Working paper, Harvard University. Available at arXiv:1605.00499v1 [stat.ME] 2 May 2016

Chernozhukov, V., Hong, H., and E. Tamer (2007). "Estimation and Confidence Regions for Parameter Sets in Econometric Models." *Econometrica* 75, 1243-1284.

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

Han, C., and P.C.B. Phillips (2006). "GMM with Many Moment Conditions." *Econometrica* 74, 147-192.

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

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

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

Kaido, H. (2016). "A Dual Approach to Inference for Partially Identified Econometric Models." *Journal of Econometrics*, 192, 269-290.

Kaido, H., and A. Santos (2014). "Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities." *Econometrica*, 82, 387–413.

Khan, S., and E. Tamer (2009). "Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities." *Journal of Econometrics*, 152, 104-119.

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

Menzel, K. (2014). "Consistent Estimation with Many Moment Inequalities." *Journal of Econometrics*, 182, 329-350.

Moon, H.R., and F. Schorfheide (2009). "Estimation with Overidentifying Inequality Moment Conditions." *Journal of Econometrics*, 153, 136-154.

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

Parente, P.M.D.C., and R.J Smith (2011). "GEL Methods for Non-Smooth Moment Indicators." *Econometric Theory* 27, 74-113.

Ponomareva, M., and E. Tamer (2011). "Misspecification in Moment Inequality Models: Back to Moment Equalities?" *The Econometrics Journal*, 14, 186-203.

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

Romano, J.P., and A.M. Shaikh (2008). "Inference for Identifiable Parameters in Partially Identified Econometric Models." *Journal of Statistical Planning and Inference*, 138, 2786-2807.

Rosen, A.M. (2008). "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities." *Journal of Econometrics*, 146, 107-117.

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

Smith, R.J. (2007). "Local GEL Estimation with Conditional Moment Restrictions." Chapter 4 in *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, eds. G.D.A. Phillips and E. Tzavalis, 100-122. Cambridge University Press: Cambridge.

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

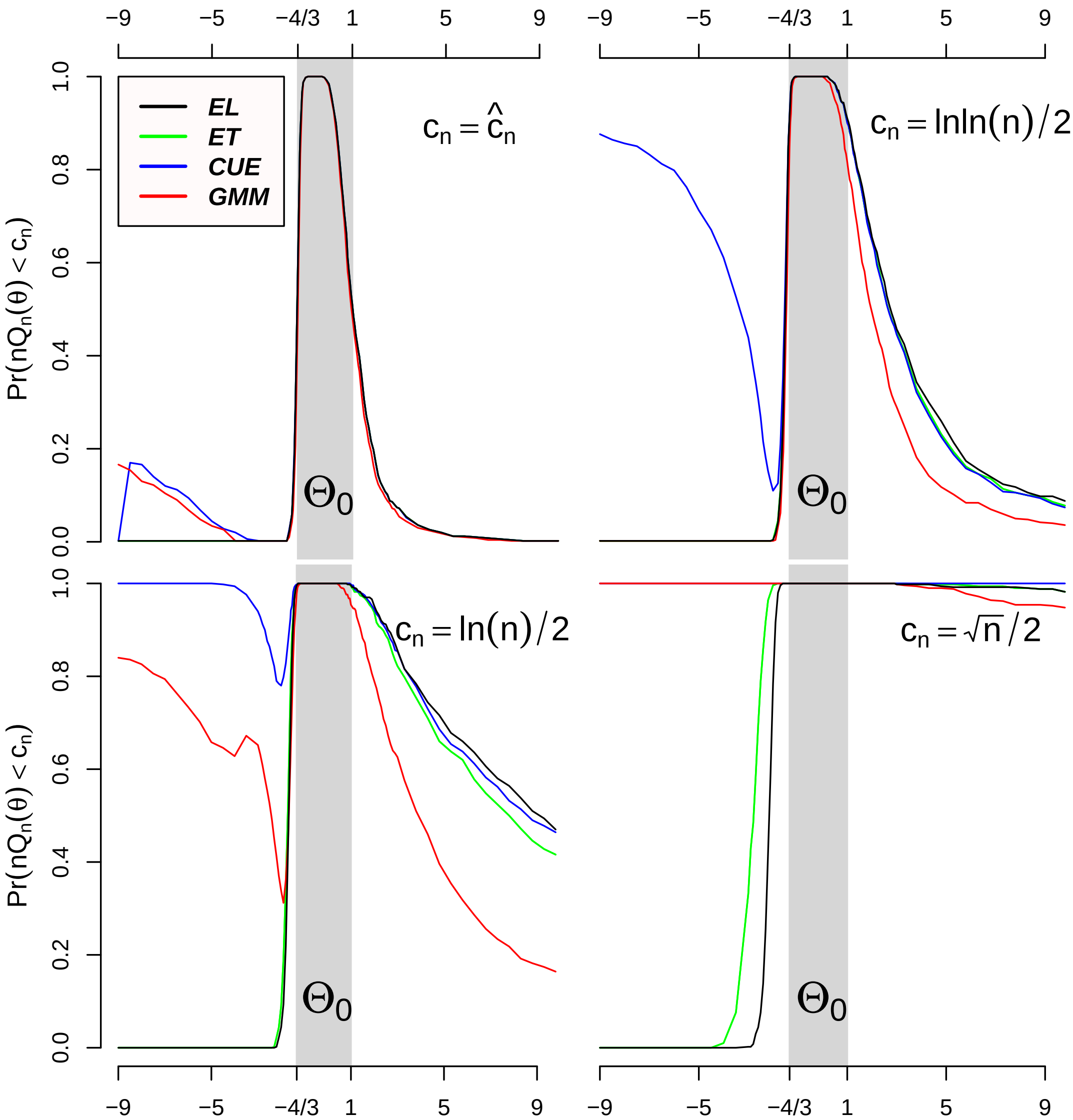

Figure 1: **Coverage Probabilities for** $\{n\hat{Q}_n^j(\theta) \leq c_n^j\}$, $n = 50$.

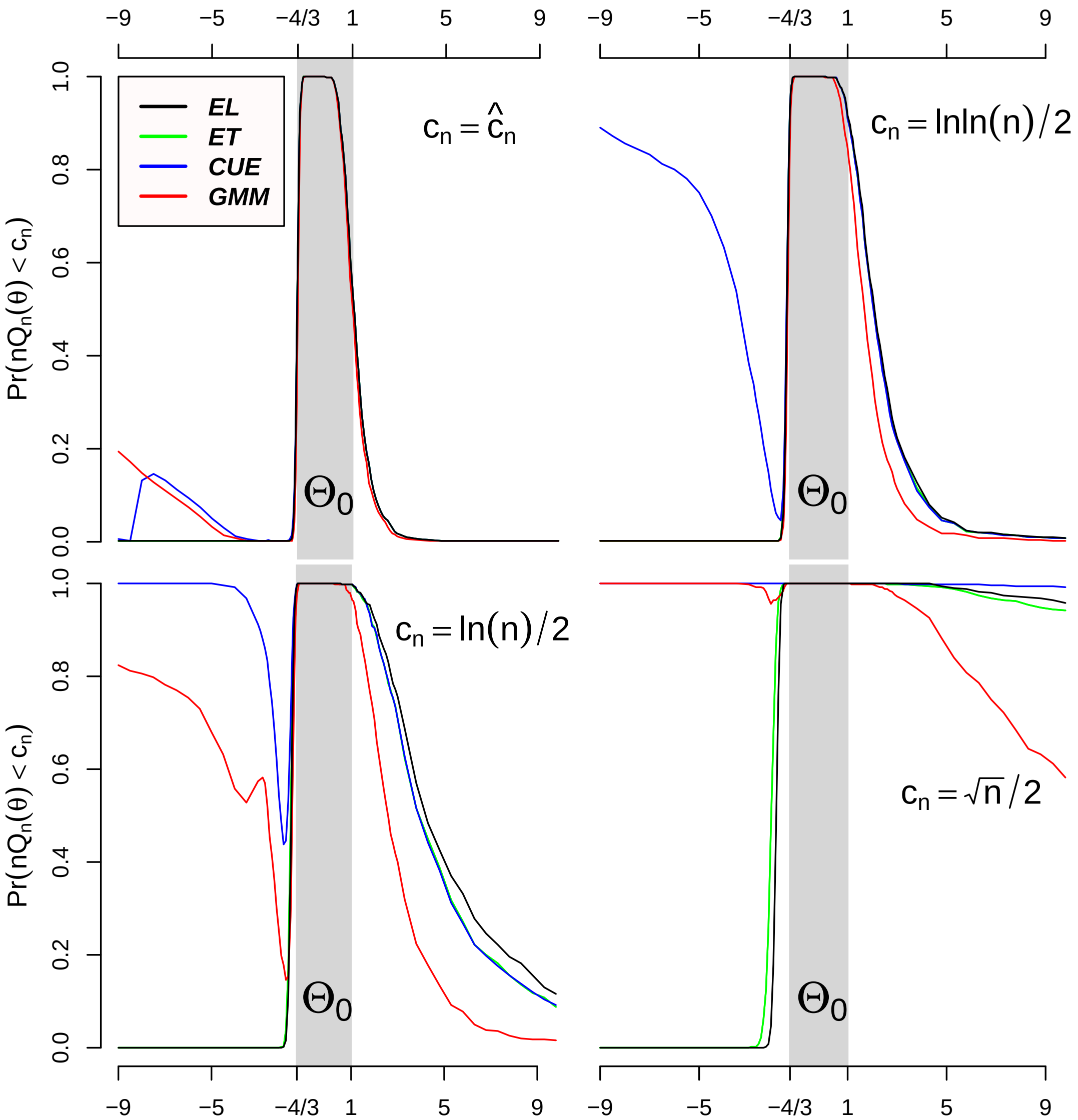

Figure 2: **Coverage Probabilities for** $\{n\hat{Q}_n^j(\theta) \leq c_n^j\}$, $n = 100$.

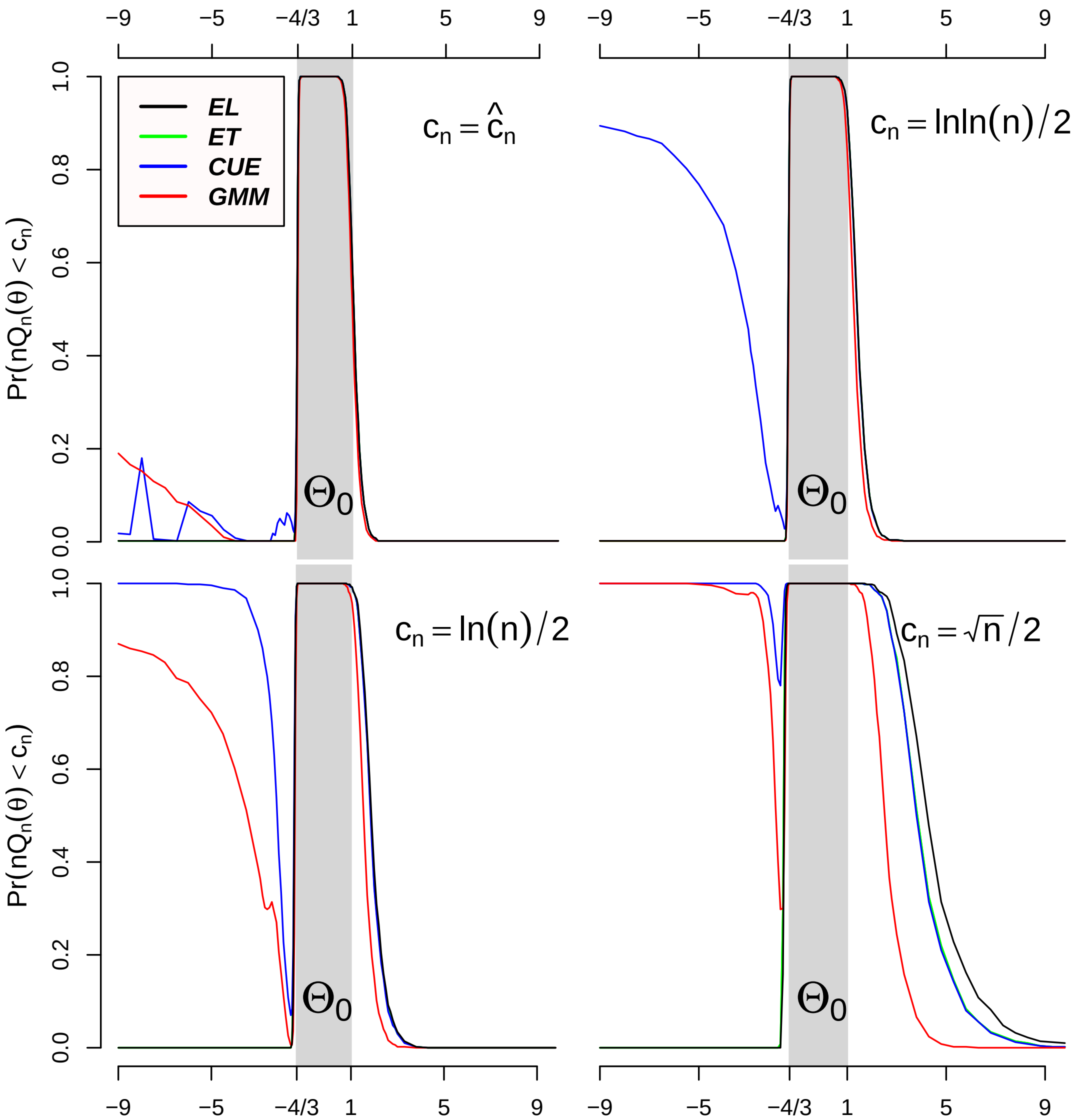

Figure 3: **Coverage Probabilities for** $\{n\hat{Q}_n^j(\theta) \leq c_n^j\}$, $n = 500$.

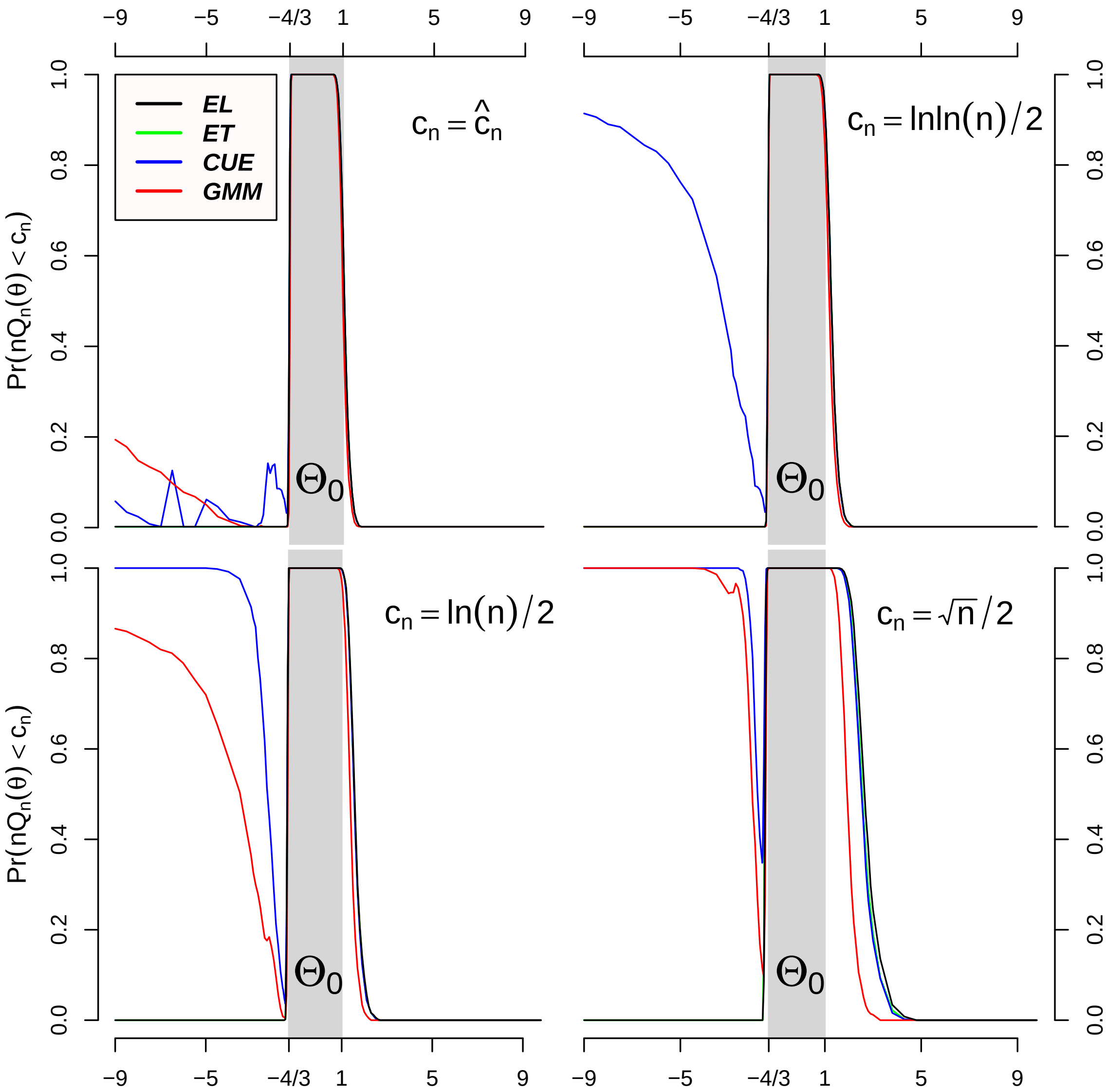

Figure 4: **Coverage Probabilities for** $\{n\hat{Q}_n^j(\theta) \le c_n^j\}$, $n = 1000$.

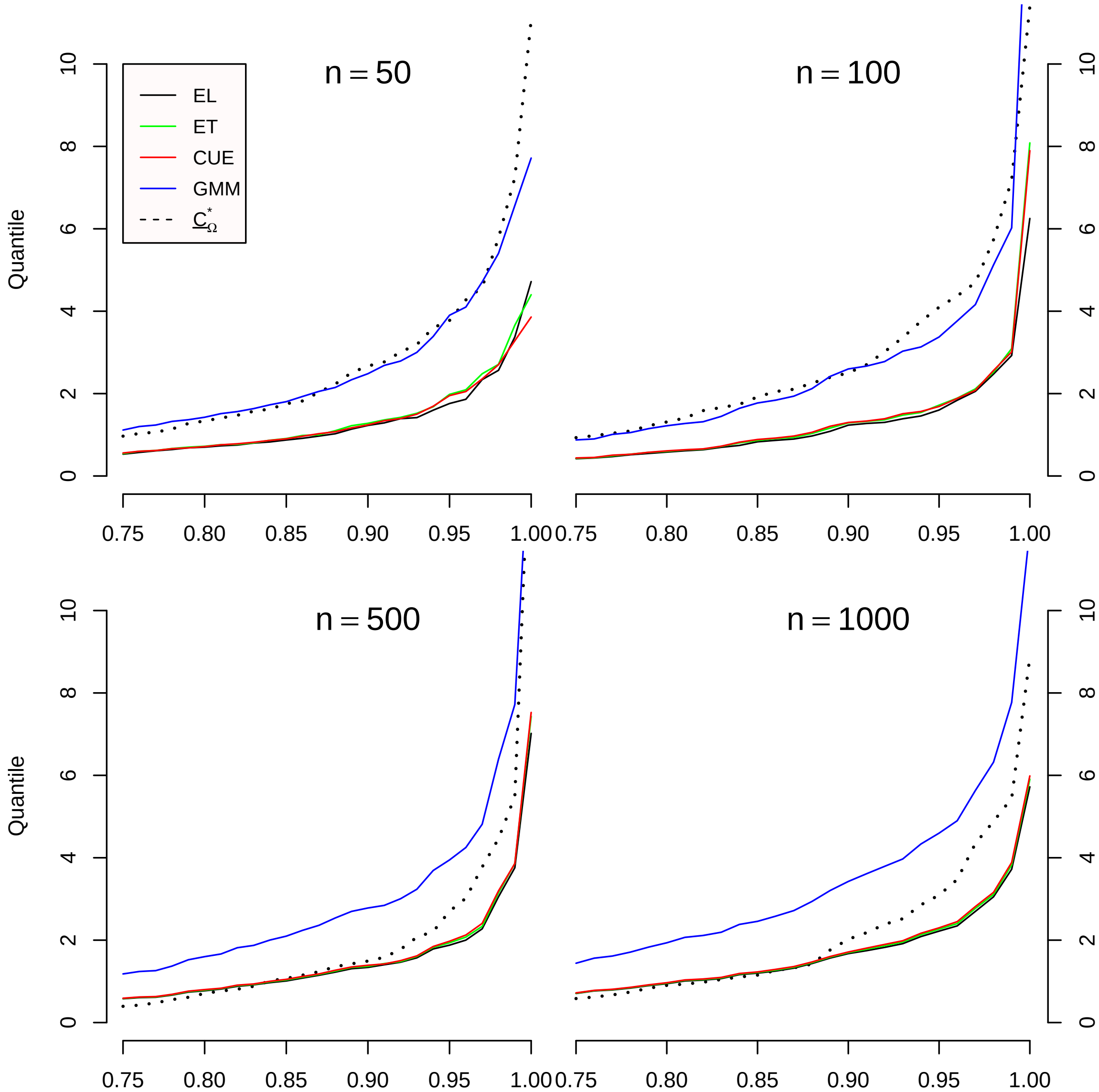

Figure 5: **Quantiles of** $\underline{\hat{C}}_n^{\Omega^{-1}*}$ **and** $\sup_{\theta\in\Theta_{P_0}} n\hat{Q}_n^j(\theta)$.