Journal of Econometrics 156 (2010) 408–425

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

EL inference for partially identified models: Large deviations optimality and bootstrap validity

Ivan A. Canay *

Department of Economics, Northwestern University, USA

ARTICLE INFO

Article history:
Received 28 October 2008
Received in revised form
14 November 2009
Accepted 17 November 2009
Available online 2 December 2009

JEL classification:
C12
C14
C21

Keywords:
Empirical likelihood
Partial identification
Large deviations
Empirical likelihood bootstrap
Asymptotic optimality

ABSTRACT

This paper addresses the issue of optimal inference for parameters that are partially identified in models with moment inequalities. There currently exists a variety of inferential methods for use in this setting. However, the question of choosing optimally among contending procedures is unresolved. In this paper, I first consider a canonical large deviations criterion for optimality and show that inference based on the empirical likelihood ratio statistic is optimal. Second, I introduce a new empirical likelihood bootstrap that provides a valid resampling method for moment inequality models and overcomes the implementation challenges that arise as a result of non-pivotal limit distributions. Lastly, I analyze the finite sample properties of the proposed framework using Monte Carlo simulations. The simulation results are encouraging.

1. Introduction

Recently, there have been many papers proposing methodologies for estimation and inference in models where the parameter of interest is not uniquely defined by the economic model and the distribution of the observed data (see, among others, Chernozhukov et al., 2007; Pakes et al., 2005; Romano and Shaikh, 2008, forthcoming; Imbens and Manski, 2004; Rosen, 2008; Beresteanu and Molinari, 2008). When this situation arises the model is said to be partially identified. Given this expanding literature on various inferential methods, it is natural to wonder which method is optimal. This paper addresses the question of optimal inference and contains the following contributions. First, I consider a canonical large deviations criterion for optimality and show that inference based on the empirical likelihood ratio (ELR) statistic is optimal. Second, I introduce a simple and natural modification of the empirical likelihood bootstrap introduced by Brown and Newey (2002) that provides a valid bootstrap method for moment inequality models. This modified empirical likelihood bootstrap is important to overcome the implementation challenges associated with non-pivotal limit distributions in partially identified models. Third, I conduct a Monte Carlo experiment which suggests a finite sample performance advantage to the new bootstrap. These results firmly ground empirical likelihood as an attractive method for inference in moment inequality models.

The problem of optimal inference can be interpreted as a problem of optimal choice of a criterion function. Partially identified models are usually represented via a population objective function $Q(\theta, P_0)$ which does not have a unique minimizer, so that

$$\Theta_0(P_0) = \arg\min_{\theta\in\Theta} Q(\theta, P_0)$$

represents a set containing all the values of θ consistent with the economic model and the distribution P_0. The primary goal is to use a sample analog $\widehat{Q}$ of $Q(\theta, P_0)$ to construct confidence regions that cover each of the elements of $\Theta_0(P_0)$ with a given probability. Most of these models involve a moment inequality condition of the form $\mathbb{E}[m(z, \theta)] \geq 0$ in which case $\Theta_0(P_0)$ is the set of all θ that satisfy the moment condition.[1] In such cases, there are many different choices of $Q(\theta, P_0)$ that have $\Theta_0(P_0)$ as the minimizer set and each choice could lead to different sample analogs and thus different confidence sets. The question of interest is whether there is an optimal criterion function $Q^*(\theta, P_0)$, where optimal means

* Corresponding address: Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA. Tel.: +1 847 491 2929.
E-mail address: iacanay@northwestern.edu.

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.11.009

[1] Some partially identified models cannot be represented as a moment inequality (see Santos, 2006; Chernozhukov et al., 2007).

that inference based on $\widehat{Q}^*$ is more *precise* than inference based on any other sample criterion function. This paper contributes to the growing literature on inference in partially identified models by giving an answer to this question. I introduce empirical likelihood (EL) as a new procedure for partially identified models and show that inference based on the empirical likelihood ratio (ELR) statistic is optimal in a large deviations sense.

The method of empirical likelihood is known to have several optimality properties for models with *equality* moment restrictions. In terms of point estimation, the EL estimator is semiparametrically efficient (i.e., attains the semiparametric efficiency bound derived by Chamberlain, 1987). In addition, this estimator exhibits desirable properties in terms of higher order comparisons (see Newey and Smith, 2004). Regarding inference, DiCiccioet al. (1991) proved that the ELR test admits Bartlett correction, which gives the same accuracy rate as the parametric case, while Kitamura (2001) showed that EL is uniformly most powerful in an Generalized Neyman–Pearson sense for testing moment restrictions. Additional optimality results are presented by Kitamura and Otsu (2005), Kitamura et al. (2009) and Canay and Otsu (2009). This is just a sample of the large list of papers that show some sort of optimality for EL. Kitamura (2006) and Owen (2001) provide additional discussions.

The search for an optimal test in partially identified models involves a number of complications that are not found in the point identified case. The fact that $\Theta_0(P_0)$ is not a singleton complicates the use of local asymptotic optimality notions since standard expansion tools are not as obviously available.[2] Another optimality notion that has been widely applied in point identified models is the large deviations approach. This approach has the virtue of translating naturally to the partially identified setting and is the criterion I pursue here.

The theory of large deviations deals with the behavior of estimators in a fixed neighborhood of the true value. Suppose that there is a statistic T_n that converges in probability to T and let A denote a set such that the closure of A does not contain T. For each n, $\Pr(T_n \in A) \to 0$. In typical cases, $\Pr(T_n \in A) \to 0$ at an exponential rate, i.e. there exists a constant $0 < \eta < \infty$ such that, $n^{-1} \log \Pr(T_n \in A) \to -\eta$. Notice the contrast with conventional local asymptotic theory where the focus is on the behavior of T_n in a shrinking neighborhood of the true parameter value, T. Here the neighborhood A is fixed. For example, let $X_1, \ldots, X_n$ be i.i.d. from $N(0, 1)$ and consider the sample mean $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. Since $\bar{X}_n$ is also normal with zero-mean and variance $1/n$, for any $\delta > 0$,

$$\Pr(|\bar{X}_n| \geq \delta) = 1 - (\sqrt{2\pi})^{-1} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} e^{-x^2/2} dx \to 0$$
$$\Longrightarrow \frac{1}{n} \log \Pr(|\bar{X}_n| \geq \delta) \to -\frac{\delta^2}{2}. \quad (1.1)$$

Eq. (1.1) is an illustration of a large deviations statement: the typical value of $\bar{X}_n$ is of order $1/\sqrt{n}$, but with small probability (of the order $e^{-n\delta^2/2}$), $\bar{X}_n$ takes values outside a fixed bound. The large deviations behavior of the type I and type II error probabilities associated with a given test procedure gives insight on the good performance of the test: the higher the rate of decrease of these errors, the better the inference. Thus, while standard definitions of efficiency (e.g., Pitman relative efficiency) make the testing problem harder by considering alternative hypothesis that get closer to the null hypothesis as the sample size increases, the efficiency criteria based on large deviations make the problem harder by letting the type I and type II error probabilities go to zero asymptotically. Precise statements are postponed to Section 3, where I show that the ELR test achieves the fastest rates of decrease.

The second main contribution of this paper is related to the actual implementation of the new optimal procedure. I show that under the null hypothesis the ELR statistic converges to a well defined asymptotic distribution. However, whether this limit distribution is useful to calculate critical values depends on the case under consideration. The number of binding constraints – the components of the vector $m(z, \theta)$ with zero expectation – depends crucially on θ and this causes the asymptotic distribution to be non-pivotal. The non-pivotalness is not a barrier in some cases and then one can compute valid critical values using a simple and straightforward approximation. In more complicated setups though, these approximations could be really slack (see Wolak, 1991; Gourieroux et al., 2008) so that using a resampling technique could be desirable. One alternative that many authors have adopted in these types of models is the use of subsampling for the construction of critical values. Subsampling would in fact be valid to approximate the limit distribution of the ELR statistic. Yet, the contribution in this paper lies in a different alternative. I first show that the empirical likelihood bootstrap proposed by Brown and Newey (2002) is not asymptotically valid when applied to moment inequality models.[3] Then I propose a slight modification of that bootstrap, along the lines of the modified parametric bootstrap in Andrews (2000), that does work asymptotically. The modification involves changing the set of inequalities $\mathbb{E}[m(z, \theta)] \geq 0$ by $\mathbb{E}[m(z, \theta)] \geq \varrho_n$, where ϱ_n is a positive sequence that goes to zero asymptotically.[4]

Before proceeding any further, I mention the recent literature that has introduced different techniques to deal with partially identified models and is closely related to the tools presented here. Horowitz and Manski (1998, 2000), Manski and Tamer (2002) and Imbens and Manski (2004) developed methods for estimation and inference for the case where the identification region is defined by lower and upper bounds that can be estimated from the data. For an excellent exposition of such cases see Manski (2003). Going beyond these particular cases, Chernozhukov et al. (2004, 2007), were the first to extend the methodologies to more general setups, defining the identified set as the solution of the minimization of a criterion function and providing several results on estimation and inference on both θ_0 and $\Theta_0(P_0)$ based on subsampling, simulation and the bootstrap. See also Andrews et al. (2004) and Pakes et al. (2005). Romano and Shaikh (2008, forthcoming) carry out a further analysis of the validity of subsampling and present conditions under which the confidence regions cover the parameter of interest uniformly. For additional results on uniform coverage see Soares (2006) and the recent papers by Andrews and Guggenberger (2009) and Andrews and Soares (forthcoming). Rosen (2008) presents a connection between moment inequality models and the literature on one-sided hypothesis testing. As it will be noted in the next section, his Gaussian quasi-likelihood ratio (QLR) statistic is closely related to the empirical likelihood ratio statistic proposed here. Using a different line of analysis Beresteanu and Molinari (2008) propose an inference procedure for partially identified models that can be written as a transformation of an expectation of a set valued random variable. Galichon and Henry (2006a,b) address the choice of the criterion function suggesting the use

2 This is related to the lack of an asymptotic distribution for sets. The methodology used by Beresteanu and Molinari (2008) is a promising direction for this type of analysis. Also, Andrews and Soares (forthcoming) use local asymptotics to compare the power properties of different critical values for a given criterion function.

3 Brown and Newey (2002) developed an EL bootstrap for models comprised of moment *equalities*.

4 This idea is related to the independently derived work by Andrews and Soares (forthcoming) and Bugni (forthcoming, 2009).

of a Kolmogorov-Smirnov-type statistic and present a bootstrap algorithm to make feasible inference in their setup. A general bootstrap procedure for moment inequality models was proposed by Bugni (forthcoming).[5] Finally, the works by Ciliberto and Tamer (forthcoming), Ho (2009) and Ishii (2005) are some of the papers that apply the tools mentioned above.

The remainder of the paper is organized as follows. Section 2 presents the basic notation and the two canonical examples used throughout the paper. Section 3 develops the empirical likelihood approach for unconditional moment inequalities and shows that inference based on the empirical likelihood ratio statistic is large deviations optimal. There are two main results in the section. The first one characterizes the asymptotic distribution of the ELR statistic. The second one deals with the so-called Generalized Neyman–Pearson optimality. Once the optimal properties are defined, Section 4 introduces the new empirical likelihood bootstrap for moment inequality models. Section 5 studies the finite sample behavior of the method proposed in this paper via Monte Carlo simulations. Finally, Section 6 concludes.

2. Setup and notation

The point of departure is a statistical model that imposes an inequality moment condition,

$$\mathbb{E}_{P_0}[m(z,\theta_0)] \equiv \int_{\mathcal{Z}} m(z,\theta_0)\mathrm{d}P_0 \geq 0, \tag{2.1}$$

where $\{z_i : i \leq n\}$ is an i.i.d. sample generated from some distribution P_0 with support on $\mathcal{Z} \subseteq \mathbb{R}^d$, θ_0 is the parameter of interest that lies in $\Theta \subseteq \mathbb{R}^k$ and $m(z,\theta)$ is a $q \times 1$ known function. Under the moment condition (2.1), the set

$$\Theta_0(P_0) \equiv \{\theta \in \Theta : \mathbb{E}_{P_0}[m(z,\theta)] \geq 0\}$$

denotes the so-called *identified set* while any $\theta \in \Theta_0(P_0)$ has been termed an *identifiable parameter* by Romano and Shaikh (2008). Thus, the true value of the parameter might not be uniquely identified by the distribution of the observed data P_0 and the economic model.

Example 1 (*Missing Data — Manski (1989)*). Let $\{(x_i; w_i) : i = 1, \ldots, n\}$ be a random sample with support $[0,1] \times \{0,1\}$, where x_i is only observed if $w_i = 1$. Let $\theta_0 = \mathbb{E}(x) \equiv p\mu_1 + (1-p)\mu_0$, where $\mu_j = \mathbb{E}(x|w=j), j = \{1,2\}$ and $p = E(w_i)$. μ_1 is identified by the data but μ_0 is only known to be in $[0,1]$. Thus, the model yields the following two moment inequalities,

$$\mathbb{E}[m_1(x,w,\theta_0)] = \mathbb{E}[\theta_0 - xw] \geq 0 \tag{2.2}$$

$$\mathbb{E}[m_2(x,w,\theta_0)] = \mathbb{E}[1 - w + xw - \theta_0] \geq 0, \tag{2.3}$$

and the identified set is simply $\Theta_0(P_0) = [p\mu_1, p\mu_1 + (1-p)]$.

Example 2 (*Entry Game — Tamer (2003)*). Suppose that firm $j \in \{1,2\}$ decides whether to enter ($z_{j,m} = 1$) a market $m \in \{1,\ldots,M\}$ or not ($z_{j,m} = 0$) based on the profit function $\pi_{j,m} = (\varepsilon_{j,m} - \theta_j z_{-j,m})1\{z_{j,m} = 1\}$, where $\varepsilon_{j,m}$ is firm's j benefit of entry in market m and $z_{-j,m}$ denotes the decision of the other firm. Let $\varepsilon_{j,m} \sim U(0,1)$ and $\theta_0 = (\theta_1,\theta_2) \in (0,1)^2$. There are four possible outcomes in this game: (i) $(z_{1,m}, z_{2,m}) = (1,1)$ is the unique Nash equilibrium (NE) if $\varepsilon_{j,m} > \theta_j$ for all j; (ii) $(z_{1,m}, z_{2,m}) = (1,0)$; is the unique NE if $\varepsilon_{1,m} > \theta_1$ and $\varepsilon_{2,m} < \theta_2$; (iii) $(z_{1,m}, z_{2,m}) = (0,1)$ is the unique NE if $\varepsilon_{1,m} < \theta_1$ and $\varepsilon_{2,m} > \theta_2$ and; (iv) there are multiple equilibria if $\varepsilon_{j,m} < \theta_j$ for all j as both $(z_{1,m}, z_{2,m}) = (1,0)$ and $(z_{1,m}, z_{2,m}) = (0,1)$ are NE. Without imposing additional assumptions this model implies

$$P(z_{1,m} = 1, z_{2,m} = 1) = (1-\theta_1)(1-\theta_2)$$
$$\theta_2(1-\theta_1) \leq P(z_{1,m} = 1, z_{2,m} = 0) \leq \theta_2,$$

so that there are two moment inequalities and one moment equality,

$$\begin{aligned} \mathbb{E}[m_1(z_m,\theta_0)] &= \mathbb{E}[z_{1,m}z_{2,m} - (1-\theta_1)(1-\theta_2)] = 0 \\ \mathbb{E}[m_2(z_m,\theta_0)] &= \mathbb{E}[z_{1,m}(1-z_{2,m}) - \theta_2(1-\theta_1)] \geq 0 \\ \mathbb{E}[m_3(z_m,\theta_0)] &= \mathbb{E}[\theta_2 - z_{1,m}(1-z_{2,m})] \geq 0. \end{aligned} \tag{2.4}$$

The identified set $\Theta_0(P_0)$ is given by,

$$\Theta_0(P_0) = \left\{ \theta \in \Theta : \begin{array}{l} \theta_2 \leq \dfrac{\mathbb{E}[z_{1,m}(1-z_{2,m})]}{\mathbb{E}[z_{1,m}(1-z_{2,m})] + \mathbb{E}[z_{1,m}z_{2,m}]} \\ \theta_2 \geq \mathbb{E}[z_{1,m}(1-z_{2,m})] \\ \theta_1 = 1 - \mathbb{E}[z_{1,m}z_{2,m}]/(1-\theta_2) \end{array} \right\}. \tag{2.5}$$

Given the moment inequalities, the goal is to construct confidence regions $\mathcal{C}_n$ that contain the parameter of interest with some pre-specified probability. Depending on the case, interest might center on the element θ_0 or the set $\Theta_0(P_0)$. That is, in partially identified models there is a distinction between inference on the identified set or on individual elements of that set (see Imbens and Manski, 2004; Chernozhukov et al., 2007; Romano and Shaikh, forthcoming). This paper focuses on inference about θ_0. This focus corresponds to interest in a particular "true value" of the parameter, which often has a particular economic interpretation. Another distinction of importance is whether the coverage of the confidence region is only valid for a fixed probability distribution P_0 or if the coverage is uniform over a large class of probability functions $\mathcal{P}$. Confidence regions in the former case satisfy the condition

$$\inf_{\theta \in \Theta_0(P_0)} \liminf_{n\to\infty} P_0(\theta \in \mathcal{C}_n) \geq 1 - \alpha, \tag{2.6}$$

while confidence regions in the latter case satisfy

$$\liminf_{n\to\infty} \inf_{(\theta,P)\in\Theta_0(P)\times\mathcal{P}} P(\theta \in \mathcal{C}_n) \geq 1 - \alpha. \tag{2.7}$$

In Section 4 I show that the EL confidence regions are uniformly valid over the class of null probabilities considered by Andrews and Soares (forthcoming).

Construction of $\mathcal{C}_n$ with the required coverage level typically proceeds as follows. One uses a sample analog $\hat{Q}(\theta)$ of a population criterion function $Q(\theta, P_0)$ and exploits the duality between hypothesis tests and confidence sets. The confidence region arises by inverting the test of each of the individual null hypotheses H_0^θ : $\theta \in \Theta_0(P_0)$ so that $\mathcal{C}_n \equiv \{\theta \in \Theta : \hat{Q}(\theta) \leq c\}$ for a given cut-off value c. In this paper I show that the use of the empirical likelihood ratio statistic $\mathcal{ELR}_n(\theta)$ results in optimal inference when testing the hypothesis H_0^θ : $\mathbb{E}_{P_0}[m(z,\theta)] \geq 0$, which is equivalent to H_0^θ : $\theta \in \Theta_0(P_0)$. I also show that $\mathcal{ELR}_n(\theta)$ has a well defined asymptotic distribution that allows one to compute an asymptotic critical value $c_{1-\alpha}^\theta$ for a given asymptotic size α. Thus, for each $\theta \in \Theta$ there is a $c_{1-\alpha}^\theta$ such that

$$\mathcal{C}_n \equiv \{\theta \in \Theta : \mathcal{ELR}_n(\theta) \leq c_{1-\alpha}^\theta\} \tag{2.8}$$

satisfies (2.6). As will be clear later on, in some cases the asymptotic distribution might not be that useful to compute $c_{1-\alpha}^\theta$. For such cases I introduce a modified empirical likelihood bootstrap that consistently estimates $c_{1-\alpha}^\theta$.

3. Optimal inference in partially identified models

An inferential method involves a test statistic based on a sample analog of $Q(\theta, P_0)$ together with a critical value c^θ. On the one hand, given a criterion function one might compute the critical

[5] The bootstrap procedure proposed in this paper and in Bugni (forthcoming, 2009) were independently derived and both are related to the GMS approach introduced, also independently, by Andrews and Soares (forthcoming).

value using an asymptotic approximation, a simulation technique or a resampling approach (e.g., subsampling or the bootstrap). On the other hand, a model that imposes $\mathbb{E}_{P_0}[m(z,\theta)] \geq 0$ has many different criterion functions $Q(\theta, P_0)$ that have $\Theta_0(P_0)$ as the minimizer set. This fact translates into consequences for the inferential procedure given that a standard method of confidence region construction is to invert a test based on a sample analog of the criterion function. Consider the next example for an illustration.

Example 3 (*Missing Data Cont.*). Consider the moments in (2.2) and (2.3) and let $m(z,\theta)' = [m_1(z,\theta) \ \ m_2(z,\theta)]$. The following criterion functions provide an equivalent representation of the identified set via $\Theta_0(P_0) = \arg\min_{\theta\in\Theta} Q_k(\theta, P_0)$, for $k = \{1, 2, 3\}$.

1. $Q_1(\theta, P_0) \equiv \left(\mathbb{E}_{P_0}[m(z,\theta)]\right)'_- W(\theta) \left(\mathbb{E}_{P_0}[m(z,\theta)]\right)_-$. Here $(x)_- = \min\{x, 0\}$ and $W(\theta)$ is a positive definite weighting matrix. This criterion function for the choice $W(\theta) = I_{2\times 2}$ has been used by Chernozhukov et al. (2007) and Romano and Shaikh (2008).
2. $Q_2(\theta, P_0) \equiv \min_{t\geq 0} \left(\mathbb{E}_{P_0}[m(z,\theta)] - t\right)' V(\theta)^{-1}(\mathbb{E}_{P_0}[m(z,\theta)] - t)$, where $V(\theta) = \text{var}\{m(z,\theta)\}$. This Gaussian quasi-likelihood ratio has been used by Rosen (2008).
3. $Q_3(\theta, P_0) \equiv \max_{\lambda\leq 0} \mathbb{E}_{P_0}[\log(1 + \lambda' m(z,\theta))]$. This criterion function defines the identified set (Lemma B.3 in Appendix B) and is the one associated with empirical likelihood.

It is useful then to think of the search for an optimal test in terms of choice of criterion function and choice of critical value. Unfortunately, optimizing over both Q and c^θ turns out to be very challenging in moment inequality models. Therefore, what this paper does is to search for the optimal criterion function given a fixed (not data-dependent) critical value and shows that appropriate choice of $Q(\theta, P_0)$ leads to a large deviations optimal inferential procedure.[6] In particular, I use a large deviations optimality criteria that is defined in terms of asymptotic power since more powerful tests reject more false hypotheses and generally lead to smaller confidence regions. Then I show that tests based on the empirical likelihood ratio statistic are optimal according to this criteria. The result firmly grounds the use of $\mathcal{ELR}_n(\theta)$, the statistic associated with the criterion function $Q_3(\theta, P_0)$ in Example 3.

One notion of optimality for confidence regions is to focus on the optimality of the corresponding tests that are inverted. To assess the relative or absolute performance of test procedures there exist several optimality criteria which, in order to make informative comparisons, usually consider problems that become harder as the sample size increases. A line of attack that is applicable to a wide range of cases is based on the theory of large deviations and has been used since the papers by Bahadur (1960), Chernoff (1952) and Hoeffding (1965), among others. Thus, test procedures are compared through their power functions and the various methods of comparison differ in the manner in which type I and type II error probabilities vary with the sample size, and also in the manner in which the alternatives under consideration are required to behave. Letting α_n and β_n denote the type I and type II error probabilities of a test, each performance criteria entails particular specifications regarding: (i) whether α_n goes to zero or not, (ii) whether β_n goes to zero or not, and (iii) whether the alternative hypotheses are fixed or get closer to the null with the sample size.[7]

In this paper I focus on large deviations cases where both types of errors decrease to zero as the sample size increases and the set of alternatives is held fixed. I follow Kitamura (2001) and use the so-called Generalized Neyman–Pearson approach, an extension of the idea introduced by Hoeffding (1965) for multinomial models. To put it in simple terms, consider competing tests that satisfy $\limsup_{n\to\infty} n^{-1}\log\alpha_n \leq -\eta$, for a given $\eta > 0$. Among such tests, a test is optimal if it minimizes $\limsup_{n\to\infty} n^{-1}\log\beta_n$ uniformly over all distributions in a given class. In Section 3.3 I show that the empirical likelihood-based test of the unconditional moment restrictions (2.1) is optimal in the above sense. In anticipation of such result and for ease of exposition, I first introduce empirical likelihood for moment inequalities.

3.1. The ELR statistic for moment inequalities

EL is a data-driven nonparametric method of estimation and inference for moment restriction models, which does not require weight matrix estimation like GMM and is invariant to nonsingular linear transformations of the moment conditions.[8] It was introduced by Owen (1988, 1990, 1991) and later studied in depth by Qin and Lawless (1994), Imbenset al. (1998), Kitamura (2001) and Newey and Smith (2004), among others.

The standard EL for moment *equalities* maximizes the nonparametric likelihood over distributions with an atom of probability on each z_i that impose the moment condition. In this paper I use this idea, but impose the moment *inequality* (2.1) into the optimization problem. The (restricted) empirical log-likelihood problem is,

$$l^r_{EL}(\theta) \equiv \max_{p_1,\ldots,p_n} \left\{ \sum_{i=1}^n \log(p_i) \,\middle|\, p_i > 0; \sum_{i=1}^n p_i = 1; \sum_{i=1}^n p_i m(z_i,\theta) \geq 0 \right\} \tag{3.1}$$

where p_i denotes the probability mass placed at z_i by a discrete distribution with support $\{z_1,\ldots,z_n\}$. Note that this differs from the usual EL for moment equalities only in the last restriction where $\sum_{i=1}^n p_i m(z_i,\theta)$ is now required to satisfy an inequality. The unrestricted empirical log-likelihood problem, $l^{ur}_{EL}(\theta)$, is similar to $l^r_{EL}(\theta)$ except that the moment restriction $\sum_{i=1}^n p_i m(z_i,\theta) \geq 0$ is not imposed. The solution in such case is simply $\tilde{p}_i = 1/n$ and then $l^{ur}_{EL}(\theta) = -n\log(n)$. The ELR statistic arises by computing the difference between the restricted and unrestricted log-likelihood,

$$\begin{aligned}\mathcal{ELR}_n(\theta) &\equiv 2\{l^{ur}_{EL}(\theta) - l^r_{EL}(\theta)\} \\ &= \min_{p_1,\ldots,p_n} \left\{ 2\sum_{i=1}^n \log\left(\frac{1/n}{p_i}\right) \,\middle|\, p_i > 0; \sum_{i=1}^n p_i = 1; \sum_{i=1}^n p_i m(z_i,\theta) \geq 0 \right\}.\end{aligned} \tag{3.2}$$

Hence, large values of this statistic suggest the restriction is not supported by the data.

A nice feature of EL is that imposing moment inequalities preserves the simplicity of the moment equality case. The only difference lies in the behavior of the Lagrange multipliers. To see this, note that (3.1) is solved by maximizing the Lagrangian

$$\mathbb{L}(\theta, \{p_i\}_{i=1}^n, \lambda, \varkappa) \equiv \sum_{i=1}^n \log(p_i) + \varkappa\left(1 - \sum_{i=1}^n p_i\right) - n\lambda' \sum_{i=1}^n p_i m(z_i,\theta) \tag{3.3}$$

6 As a way of comparison, Andrews and Soares (forthcoming) do basically the opposite. For a given criterion function Q, they search for the optimal way (in terms of size and power) of computing a critical value.

7 For a review on asymptotic comparisons see Serfling (1980, Ch. 10) and van der Vaart (1998, Ch. 14).

8 EL does not require weight matrix estimation even computationally when ones uses algorithms with numerical derivatives.

where $\varkappa$ is the Lagrange multiplier for the second constraint and $\lambda \leq 0$ is a $q \times 1$ vector of multipliers for moment inequality constraints. Solving this problem results in

$$\widehat{p}_i = \frac{1}{n(1+\widehat{\lambda}' m(z_i,\theta))}, \tag{3.4}$$

which looks identical to the standard EL solution. The difference lies in $\widehat{\lambda}$, which now solves the following three first order conditions[9]

$$\sum_{i=1}^{n} \frac{m(z_i,\theta)}{n(1+\widehat{\lambda}' m(z_i,\theta))} \geq 0; \qquad \widehat{\lambda} \leq 0;$$

$$\widehat{\lambda}' \sum_{i=1}^{n} \frac{m(z_i,\theta)}{n(1+\widehat{\lambda}' m(z_i,\theta))} = 0.$$

Since $\widehat{p}_i$ has a closed form solution, I can write a profiled likelihood and define $\widehat{\lambda}$ accordingly. Plugging (3.4) into (3.3) results in

$$l^r_{EL}(\theta) \equiv \mathbb{L}(\theta, \{\widehat{p}_i\}_{i=1}^n, \lambda, \widehat{\varkappa}) = \min_{\lambda \leq 0} \left\{ -n\log(n) - \sum_{i=1}^{n} \log(1+\lambda' m(z_i,\theta)) \right\}.$$

Finally, using $l^{ur}_{EL}(\theta) = -n\log(n)$ the statistic in (3.2) becomes,

$$\mathcal{ELR}_n(\theta) \equiv 2\{l^{ur}_{EL}(\theta) - l^r_{EL}(\theta)\} = \max_{\lambda \leq 0} 2\sum_{i=1}^{n} \log(1+\lambda' m(z_i,\theta)). \tag{3.5}$$

Therefore, a model represented through moment inequalities affects the empirical likelihood ratio statistic only via the Lagrange multiplier λ which is now required to be non-positive. This difference is important for two reasons. First, relative to the standard case, the computational difficulty is only trivially affected. Second, the restriction on λ affects the limit distribution of the statistic to a great extent and results in a non-pivotal asymptotic distribution, as the next section shows.

3.2. Asymptotic distribution

Before deriving the asymptotic distribution of the ELR statistic I introduce some additional notation. Let $m(z_i,\theta)$ be $q \times 1$ with the following partition: $m(z_i,\theta)' = [m_b(z_i,\theta)' m_s(z_i,\theta)']$ where $m_b(z_i,\theta)$ is a $b(\theta) \times 1$ vector of moments with zero mean, $\mathbb{E}_{P_0}[m_b(z,\theta)] = 0$, and $m_s(z_i,\theta)$ is a $s(\theta) \times 1$ vector of moments with positive mean, $\mathbb{E}_{P_0}[m_s(z,\theta)] > 0$. The asymptotic behavior of the test statistic will follow from the next standard assumption.

Assumption 3.1. (i) $\{z_i : i \leq n\}$ is a random sample where $z_i \in \mathcal{Z} \subseteq \mathbb{R}^d$. (ii) $\Theta \subseteq \mathbb{R}^k$ is compact and $m : \mathbb{R}^d \times \Theta \to \mathbb{R}^q$ is known. (iii) $\sup_{\theta\in\Theta_0(P_0)} \mathbb{E}_{P_0}[\|m(z,\theta)\|^r] < \infty$ for $r \geq 3$. (iv) For each $\theta \in \Theta_0(P_0)$, $\Sigma_b(\theta) = \mathbb{E}_{P_0}[m_b(z,\theta)m_b(z,\theta)']$ is positive definite.

Assumption 3.1(iii) implies that for all $\theta \in \Theta_0(P_0)$,

$$\sqrt{n}\{\bar{m}_n(\theta) - \mathbb{E}_{P_0}[m(z,\theta)]\} \rightsquigarrow N(0, V(\theta))$$

where $\bar{m}_n(\theta) = n^{-1}\sum_{i=1}^n m(z_i,\theta) \to^{P_0} \mathbb{E}_{P_0}[m(z,\theta)]$, $V(\theta) = \text{var}_{P_0}\{m(z,\theta)\}$ and $\rightsquigarrow$ denotes weak convergence. In particular, under the same assumption $\sqrt{n}\bar{m}_{n,b}(\theta) \rightsquigarrow N(0, \Sigma_b(\theta))$.

The following theorem provides a limit distribution for the statistic $\mathcal{ELR}_n(\theta)$ under the null hypothesis.

Theorem 3.1. *Under Assumption 3.1, for all $\theta \in \Theta_0(P_0)$ the statistic in (3.5) satisfies:*

$$\lim_{n\to\infty} P_0(\mathcal{ELR}_n(\theta) \geq c) = \sum_{j=0}^{b(\theta)} \varpi(b(\theta), b(\theta)-j, \Sigma_b(\theta)) \Pr(\chi^2_j \geq c) \tag{3.6}$$

where $\varpi(b(\theta), b(\theta)-j, \Sigma_b(\theta))$ is the weight function defined by Wolak (1987) and Kudo (1963).

Proof. See Appendix A. ∎

Theorem 3.1 shows that the limit distribution of $\mathcal{ELR}_n(\theta)$ is a chi-bar-square distribution, which I denote by $\overline{\chi}^2_{b(\theta)}(\Sigma_b(\theta))$. This distribution is non-pivotal since both $b(\theta)$ and $\varpi(b(\theta), b(\theta)-j, \Sigma_b(\theta))$ depend on θ. The set of $b(\theta)$ binding constraints has a significant discontinuous effect on the shape of the distribution. In the extreme case where θ is such that $b(\theta) = 0$ – no constraint is binding – the resulting distribution is degenerate at zero since $\mathcal{ELR}_n(\theta) = 0$ wp$\to$1. The weights $\varpi(b(\theta), b(\theta)-j, \Sigma_b(\theta))$ are called level probabilities and aside from cases where $b(\theta) \leq 4$, there are no closed-form expressions for these weights.[10] Rosen (2008) derived a QLR statistic for the model in (2.1) and showed that its limit distribution is the chi-bar-square distribution in (3.6). To overcome the problem caused by the non-pivotalness of the statistic Rosen uses two conservative approximations to the chi-bar-square that were proposed by Wolak (1987, 1991). These approximations require imposition of an upper bound b^* on $b(\theta)$, i.e., $\sup_{\theta\in\Theta_0(P_0)} b(\theta) \leq b^*$ – see Eq. (5.1) in Section 5.[11]

Theorem 3.1 shows that the ELR statistic is equivalent to the QLR statistic proposed by Rosen (2008) up to first order. However, EL has two clear advantages. First, the QLR statistic cannot outperform $\mathcal{ELR}_n(\theta)$ in terms of asymptotic power as the next subsection shows. Second, the computation of $\mathcal{ELR}_n(\theta)$ does not require an estimate of $V(\theta)$ as it is required for the QLR statistic. This feature reflects the internal Studentization property of empirical likelihood that often improves finite sample properties of the tests. In fact, avoiding estimation of $V(\theta)^{-1}$ is more important than usual in the present setup. For example, in the entry game of Section 5 this matrix is singular.[12] On the other hand, the one disadvantage of EL is additional computational time. Section 5 also addresses this.

3.3. Generalized Neyman–Pearson optimality

I now consider the problem of optimal inference in models that impose the moment inequalities (2.1) and show that the ELR statistic just described yields optimal inference from a large deviations point of view. The idea works as follows. A test for H_0^θ is a map from data into a decision. I denote this mapping by $r_n^\theta : \mathcal{Z} \to \{0, 1\}$, where $r_n^\theta = 0$ ($r_n^\theta = 1$) means acceptance (rejection) of the null. Characterizing data by the empirical measure $\hat{P}_n$ it induces, a test maps $\hat{P}_n$ into accept/reject. This is equivalent to partitioning the space of probability measures into acceptance or rejection regions. To write this formally denote by $\mathcal{M}$ the space of probability measures on the Borel σ-field $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$ endowed with the Lévy metric.[13] A test then induces a partition $\Omega_n^\theta \equiv (\Omega_{n,0}^\theta, \Omega_{n,1}^\theta)$ of $\mathcal{M}$ such that $r_n^\theta = 1(\hat{P}_n \in \Omega_{n,1}^\theta)$.

9 $\widehat{\lambda}$ is a shorthand for $\widehat{\lambda}(\theta)$ or $\widehat{\lambda}(\theta,\tau)$ depending on the context.

10 These issues and much more have been studied in depth in the literature on one-sided hypothesis testing, e.g. Wolak (1987), Silvapulle and Sen (2004) and Gourieroux et al. (1982).

11 If b^* is a good approximation to the supremum and the number of moments is small, these approximations can work well (see Example 1, Section 5). However, if the moments are non-linear and the number of binding moments is large, they could be slack as mentioned by Wolak (1991) and Gourieroux et al. (2008).

12 Also, $V(\theta)$ might be singular when using subsampling in the presence of missing data and small n.

13 The Lévy metric is compatible with the weak topology, Dembo and Zeitouni (1998, Theorem D.8).

Recall that the null hypothesis for each $\theta \in \Theta$ is given by

$$H_0^\theta : \mathbb{E}_{P_0}[m(z,\theta)] \equiv \int_{\mathcal{Z}} m(z,\theta)\mathrm{d}P_0 \geq 0, \tag{3.7}$$

which can be alternatively written as $H_0^\theta : P_0 \in \mathcal{P}_0(\theta)$, where

$\mathcal{P}_0(\theta) \equiv \{P \in \mathcal{M} : \mathbb{E}_P[m(z,\theta)] \geq 0\}$

denotes the subset of probability measures that satisfy the moment inequality restriction. Next, let $Q \ll P$ denote that Q is absolutely continuous with respect to P,[14]

$$I(Q \parallel P) \equiv \begin{cases} \int \log(\mathrm{d}Q/\mathrm{d}P)\mathrm{d}Q & \text{if } Q \ll P \\ \infty & \text{otherwise} \end{cases}$$

denote the relative entropy (or Kullback–Leibler divergence) for measures Q and P, and

$\mathcal{P}(Q,\theta) \equiv \{J \in \mathcal{P}_0(\theta) : J \ll Q, Q \ll J\}$,

denote the measures in $\mathcal{P}_0(\theta)$ that are equivalent to a given measure Q. This notation allows me to re-write the ELR statistic in (3.2) as,

$$\mathcal{ELR}_n(\theta) = \min_{P \in \mathcal{P}(\hat{P}_n,\theta)} 2nI(\hat{P}_n \parallel P).$$

Intuitively, EL picks the measure in the set $\mathcal{P}(\hat{P}_n,\theta)$ that is closest to the empirical measure, where closest is defined in terms of entropy. Finally, note that the test based on ELR,

$$\bar{r}_n^\theta \equiv 1\left(\frac{\mathcal{ELR}_n(\theta)}{2n} \geq \eta^\theta\right) = 1\left(\inf_{P \in \mathcal{P}(\hat{P}_n,\theta)} I(\hat{P}_n \parallel P) \geq \eta^\theta\right),$$

depends on the data solely through $\hat{P}_n$ and induces the following partition of $\mathcal{M}$,

$$\Lambda_0^\theta \equiv \left\{Q \in \mathcal{M} : \inf_{P \in \mathcal{P}(Q,\theta)} I(Q \parallel P) < \eta^\theta\right\},$$
$$\Lambda_1^\theta \equiv \left\{Q \in \mathcal{M} : \inf_{P \in \mathcal{P}(Q,\theta)} I(Q \parallel P) \geq \eta^\theta\right\}. \tag{3.8}$$

In a recent paper, Kitamura et al. (2009) established the Generalized Neyman–Pearson optimality of EL for testing moment equalities. They provide a counterexample to the results of Kitamura (2001), who was the first to consider this problem. The counterexample shows that if the class of null distributions is too rich, most commonly-used tests (including EL) cannot control the rate at which the type I error tends to zero. Taking this into consideration I define the space of null distributions as follows.

Definition 3.1 (*Null Parameter Space*). *Given $\epsilon > 0$, $\mathcal{F}_\epsilon$ denotes the space for null parameters $(\theta, P) \in \Theta \times \mathcal{M}$ such that:*

(i) *$\{z_i : i \leq n\}$ is i.i.d under P,*
(ii) *$\mathbb{E}_P[m(z,\theta)] \geq 0$,*
(iii) *$\sigma_{P,j}^2(\theta) = \mathrm{var}_P[m_j(z,\theta)] \in [\epsilon, \infty)$, for $j = 1, \ldots, q$,*
(iv) *$|\mathrm{Corr}_P[m(z,\theta)]| \geq \epsilon$, and*
(v) *$E_P|m_j(z,\theta)/\sigma_{P,j}(\theta)|^{2+a} \leq M$, for $a > 0$, $M < \infty$ and $j = 1, \ldots, q$,*

where Corr$_P$ *denotes correlation matrix and $|\cdot|$ the determinant of a square matrix. In addition, for a given $\theta \in \Theta$, $\mathcal{P}_{\epsilon,0}(\theta) = \{P \in \mathcal{M} : (\theta, P) \in \mathcal{F}_\epsilon\}$ is the set of null distributions in $\mathcal{F}_\epsilon$ and $\mathcal{F}_0$ denotes $\mathcal{F}_\epsilon$ with $\sigma_{P,j}^2(\theta) \in (0,\infty)$.*

Note that the space $\mathcal{F}_0$ is the same parameter space used by Andrews and Soares (forthcoming) to show uniform validity of their generalized moment selection (GMS) approach. The bootstrap method proposed in the next section is also uniformly valid over $\mathcal{F}_0$. $\mathcal{F}_\epsilon$ further restricts $\mathcal{F}_0$ by ruling out distributions that become arbitrarily close to degenerate distribution (i.e., distributions that put probability 1 on a strict subspace of $\mathcal{Z}$). This is similar in spirit to the approach used by Kitamura et al. (2009) although $\mathcal{F}_\epsilon$ imposes variance-covariance restrictions as opposed to directly controlling the minimum mass P puts on any half-space.[15]

The main result of this section follows from the two assumptions below, and uses the following notation. An open δ-blow up of a set $A \subset \mathcal{M}$ is given by $A^\delta \equiv \{Q \in \mathcal{M} : \inf_{P \in A} d(Q,P) < \delta\}$, where $d(Q,P)$ denotes the Lévy metric. Also, P^n denotes the n-fold product measure $\bigotimes_{i=1}^n P$ of a measure P.

Assumption 3.2. $\mathcal{Z}$ and Θ are compact subsets of $\mathbb{R}^d$ and $\mathbb{R}^k$.

Assumption 3.3. $m(z,\theta) : \mathcal{Z} \times \Theta \mapsto \mathbb{R}^q$ is continuous in z for each $\theta \in \Theta$.

Theorem 3.2. *Suppose Assumptions 3.2 and 3.3 hold. Let $\Lambda^\theta = (\Lambda_0^\theta, \Lambda_1^\theta)$ be defined as in (3.8) and $\mathcal{P}_{\epsilon,0}(\theta)$ denote the null set from Definition 3.1. For each $\theta \in \Theta$ there exists $\eta^\theta(\epsilon) > 0$ such that for any $0 < \eta^\theta \leq \eta^\theta(\epsilon)$ the statements below follow.*

(I) *The Empirical Likelihood Ratio test satisfies,*

$$\sup_{P \in \mathcal{P}_{\epsilon,0}(\theta)} \limsup_{n\to\infty} \frac{1}{n} \log P^n(\hat{P}_n \in \Lambda_1^\theta) \leq -\eta^\theta. \tag{3.9}$$

(II) *If an alternative test $\Omega_n \equiv (\Omega_{n,0}, \Omega_{n,1})$ satisfies,*

$$\sup_{P \in \mathcal{P}_{\epsilon,0}(\theta)} \limsup_{n\to\infty} \frac{1}{n} \log P^n(\hat{P}_n \in \Omega_{n,1}^{\theta,\delta}) \leq -\eta^\theta \tag{3.10}$$

for any $\delta > 0$, it follows that,

$$\limsup_{n\to\infty} \frac{1}{n} \log P_1^n(\hat{P}_n \in \Omega_{n,0}) \geq \limsup_{n\to\infty} \frac{1}{n} \log P_1^n(\hat{P}_n \in \Lambda_0^\theta)$$

for any $P_1 \in \mathcal{A}_{\eta,\epsilon}(\theta) \equiv \{Q \notin \mathcal{P}_0(\theta) : d(Q,P) \geq \sqrt{\eta/2}, \forall P \in \mathcal{P}_0(\theta) \setminus \mathcal{P}_{\epsilon,0}(\theta)\}$.

Proof. See Appendix A. ■

Theorem 3.2 says that EL *uniformly* controls the rate of decay of type I error probabilities over distributions in $\mathcal{P}_{\epsilon,0}(\theta)$ and that there is no test for the null (3.7) satisfying the rate restriction (3.10) that outperforms EL in term of asymptotic power rate. This part of the Theorem also holds uniformly over alternative distributions that are far from degenerate null distributions, i.e., alternatives in $\mathcal{A}_{\eta,\epsilon}(\theta)$.

Remark 3.1. As in Kitamura (2001, Theorem 2), Theorem 3.2 uses δ-smoothing. To get a similar result without the need of smoothing, the alternative test has to be regular, i.e.,

$$\lim_{\delta\to 0} \sup_{P \in \mathcal{P}_{\epsilon,0}(\theta)} \limsup_{n\to\infty} n^{-1} \log P^n(\hat{P}_n \in \Omega_{n,1}^{\theta,\delta})$$
$$= \sup_{P \in \mathcal{P}_{\epsilon,0}(\theta)} \limsup_{n\to\infty} n^{-1} \log P^n(\hat{P}_n \in \Omega_{n,1}^{\theta}).$$

Remark 3.2. Theorem 3.2 is an analog result to those in Kitamura (2001) and Kitamura et al. (2009), although with important

[14] Q is absolutely continuous with respect to P if $P(B) = 0$ implies $Q(B) = 0$ for every measurable set B.

[15] A completely different approach would be to use the notion of δ-optimality as in Zeitouni and Gutman (1991). A Supplementary Appendix available upon request shows that EL is also optimal in this sense.

differences. First, the type I error rate in (3.9) is controlled by imposing restrictions on the covariance matrix of $m(z,\theta)$. Second, the set of relevant alternative distributions $\mathcal{A}_{\eta,\epsilon}(\theta)$ is here defined in terms of the Lévy distance which is the metric of $\mathcal{M}$. Third, the proof of the Theorem exploits the Donsker–Varadham variational formula of the relative entropy and does not use duality of linear programs.

Remark 3.3. Theorem 3.2 employs a "fixed" (i.e., not data dependent) critical value η^θ that depends on θ and so η^θ is not necessarily conservative. However, in many situations computing such a critical value might not be feasible without using a data-dependent rule.

Remark 3.4. Having good power against distant alternatives is important in moment inequality models. This is so because confidence regions are constructed by exploring Θ pointwise. Note, however, that alternatives in $\mathcal{A}_{\eta,\epsilon}(\theta)$ can be close to null distributions in $\mathcal{P}_{\epsilon,0}(\theta)$.

The large deviation optimality of EL is tightly related to its connection to the relative entropy. The two key elements that help explain this connection are the Large Deviation Principle (LDP) and Sanov's Theorem, defined below.

Definition 3.2 (*Large Deviation Principle*). *Let $I : S \to [0,\infty]$ be a function such that $I^{-1}([0,a]) \subset S$ is compact for each $a > 0$. The sequence of probability measures $\{Q_n\}_{n\geq 1}$ is said to obey the LDP with a rate function I if for any set $G \subset S$,*

$$-\inf_{y\in G^o} I(y) \leq \liminf_{n\to\infty} \frac{1}{n}\log Q_n(G^o) \leq \limsup_{n\to\infty}\frac{1}{n}\log Q_n(\bar{G}) \leq -\inf_{y\in\bar{G}} I(y),$$

where G^o and $\bar{G}$ denote the interior and closure of G, respectively.

The LDP characterizes the limiting behavior of a family of probability measures in terms of a rate function $I(\cdot)$. For example, the rate function that controls the probability measure of the sample mean in Eq. (1.1) is $I(y) = y^2/2$,

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n(\bar{X}_n \in [\delta,\infty)) \leq -\inf_{y\in[\delta,\infty)}\frac{y^2}{2} = -\frac{\delta^2}{2}.$$

Cramér showed that the LDP holds for sample means of i.i.d. random variables. Since the empirical measure can be viewed as a mean (of Dirac measures), it is a potential candidate for a large-deviation theorem. The first version of such a theorem was proved by Sanov.

Theorem 3.3 (*Sanov*). *Let $\mathcal{M}(\Psi)$ denote the space of probability measures on a Polish space Ψ equipped with the Lévy metric and take $P_0 \in \mathcal{M}(\Psi)$. Then for any set $G \in \mathcal{M}(\Psi)$,*

$$-\inf_{v\in G^o} I(v \parallel P_0) \leq \liminf_{n\to\infty} n^{-1}\log P_0^n(\hat{P}_n \in G^o) \leq \limsup_{n\to\infty} n^{-1}\log P_0^n(\hat{P}_n \in \bar{G}) \leq -\inf_{v\in\bar{G}} I(v \parallel P_0).$$

Intuitively, the ELR test uses the minimum Kullback–Leibler divergence between the empirical measure $\hat{P}_n$ and the set $\mathcal{P}(\hat{P}_n,\theta)$ as a statistical criterion. By Sanov's Theorem, the sequence of empirical measures $\hat{P}_n$ satisfies the LDP with rate function $I(Q \parallel P)$, and then $I(Q \parallel P)$ controls the limit behavior of the probability that $\hat{P}_n$ falls into the set $\mathcal{P}(\hat{P}_n,\theta)$.

Before moving to the next section it is important to note that the good power properties of EL have direct implications on the associated confidence regions since $\mathcal{C}_n$ is constructed by test inversion. Therefore, if $\theta^* \in \Theta_0^c$ is such that $P_0 \in \mathcal{A}_{\eta,\epsilon}(\theta^*)$, $\Lambda_0^{\theta^*}$ is most powerful and then the EL confidence region $\mathcal{C}_n$ is asymptotically most accurate at θ^*, meaning that it is the procedure with the smallest chance of covering such incorrect values of θ.

4. Implementation: A new empirical likelihood bootstrap

This sections concerns implementation issues for the ELR statistic. From (2.8) it is clear that once a critical value $c^\theta_{1-\alpha}$ is available, the construction of the confidence region only involves the evaluation of $\mathcal{ELR}_n(\theta)$ which is straightforward. The analysis in Section 3 shows that $\mathcal{ELR}_n(\theta)$ has a well defined asymptotic distribution that can be used to compute critical values in some simple cases. More generally, computation of fixed asymptotic critical values is infeasible. In the present EL setting the empirical likelihood bootstrap introduced by Brown and Newey (2002) for moment *equality* models seems to be an appealing alternative. Since when inequality constraints are present such approach does not produce a consistent approximation to the limit distribution of $\mathcal{ELR}_n(\theta)$, as the next section illustrates, I show that a slight modification of this bootstrap is first order valid. The modification is simple enough to preserve the computational tricks that typically make EL straightforward to use.

4.1. EL bootstrap invalidity: A canonical example

To motivate the modified empirical likelihood bootstrap, I use a simple example from Romano and Shaikh (2008) to show that the standard empirical likelihood bootstrap does not work. The argument follows the one used by Andrews (2000) to show that the standard i.i.d. bootstrap is inconsistent when the parameter is on the boundary of the parameter space. The example also illustrates intuitively why the new bootstrap does work. The general case is developed following this illustrative example.

Suppose the economic model imposes $\mathbb{E}_{P_0}(X) \geq \theta$ where $X_i \sim P_0 = N(0,1)$. Without loss of generality and to make the example simpler, I use the criterion function $Q(\theta,P_0) = (\mathbb{E}_{P_0}(X)-\theta)^2_-$ where $(a)_- = \min\{a,0\}$. The identified set here is $\Theta_0(P) = [\theta_l, 0]$ where θ_l is some lower bound of the parameter space. The sample analog satisfies,

$$n\hat{Q}_n(\theta) = (\sqrt{n}(\bar{X}_n-\theta))^2_- \rightsquigarrow (\mathbb{Z})^2_- \times 1\{\theta=0\}, \quad \mathbb{Z}\sim N(0,1).$$

The EL bootstrap works as follows. Denote by $\tilde{P}_n = (\tilde{p}_1,\ldots,\tilde{p}_n)$ the EL probabilities that solve (3.1) with $m(X_i,\theta) = X_i - \theta$ and let $\tilde{\mu}^\theta_n \equiv \sum \tilde{p}_i(X_i-\theta) = \max\{\bar{X}_n-\theta,0\}$. The bootstrap samples $\{X_i^* : i \leq n\}$ are i.i.d. according to $\tilde{P}_n$ and the bootstrap criterion function is given by $n\tilde{Q}^*_n(\theta) = (\sqrt{n}(\bar{X}^*_n-\theta))^2_-$.

Following Andrews (2000, page 401) let $B_c \equiv \{\omega : \limsup_{n\to\infty}\sqrt{n}\bar{X}_n > c\}$ for $0 < c < \infty$ and note that by the law of iterated logarithm $P_0(B_c) = 1$. For $\omega \in B_c$ consider a subsequence $\{n_k : k \geq 1\}$ of n such that $n_k^{1/2}\bar{X}_n(\omega) \geq c$ for all k. Then, for such a subsequence,

$$\begin{aligned}
n_k\tilde{Q}^*_{n_k}(0) &= (\sqrt{n_k}(\bar{X}^*_{n_k}-\tilde{\mu}^0_{n_k}) + \sqrt{n_k}\tilde{\mu}^0_{n_k})^2_- \\
&= (\sqrt{n_k}(\bar{X}^*_{n_k}-\tilde{\mu}^0_{n_k}) + \max\{\sqrt{n_k}\bar{X}_{n_k},0\})^2_- \\
&\leq (\sqrt{n_k}(\bar{X}^*_{n_k}-\tilde{\mu}^0_{n_k}) + c)^2_- \\
&\rightsquigarrow (\mathbb{Z}+c)^2_- \quad \text{as } k\to\infty \text{ conditional on } \{\tilde{P}_{n_k} : k\geq 1\} \\
&\leq (\mathbb{Z})^2_-,
\end{aligned}$$

meaning that with probability one (with respect to the randomness in $\{\tilde{P}_n : n \geq 1\}$) the EL bootstrap fails to approximate the asymptotic distribution of $n\hat{Q}_n(\theta)$ at $\theta = 0$.

4.2. EL bootstrap validity: Modification for the canonical example

Next I show that a simple modification to the EL bootstrap yields first order validity for the canonical example and in the next section I discuss the general case. The modification of the EL bootstrap I propose affects both the EL probabilities and the bootstrap criterion function. Intuitively, instead of imposing $\mathbb{E}_{P_0}(X-\theta) \geq 0$, I impose $\mathbb{E}_{P_0}(X-\theta) \geq \varrho_n$ where ϱ_n is a sequence of positive random (or deterministic) variables satisfying

$$P_0\left(\lim_{n\to\infty} \varrho_n = 0 \text{ and } \liminf_{n\to\infty} \varrho_n (n/(2\log\log n))^{1/2} \geq 1\right) = 1. \quad (4.1)$$

The modified EL bootstrap for this example requires two steps. The first step involves computing the modified EL probabilities $\bar{P}_n = (\bar{p}_1, \ldots, \bar{p}_n)$ by solving

$$\tilde{l}^r_{EL} = \sup_{p_1,\ldots,p_n} \left\{ \sum_{i=1}^n \log(p_i) \middle| p_i > 0; \sum_{i=1}^n p_i = 1; \sum_{i=1}^n p_i(X_i - \theta) \geq \varrho_n \right\}.$$

Once $\bar{P}_n$ is known, the new empirical likelihood mean is denoted by $\bar{\mu}^\theta_n = \sum \bar{p}_i(X_i - \theta) = \max\{\bar{X}_n - \theta, \varrho_n\}$. The second step generates i.i.d. bootstrap samples $\{X^*_i : i \leq n\}$ from $\bar{P}_n$ and computes the modified bootstrap criterion function, $n\hat{Q}^*_n(\theta) = (\sqrt{n}(\bar{X}^*_n - \theta - \varrho_n))^2_-$.

Remark 4.1. When the parameter of interest is on the boundary of the parameter space, Andrews (2000, page 403) proposes a parametric bootstrap procedure in which the parameter estimator used to generate the bootstrap shrinks to the boundary when $\bar{X}_n$ is below ϱ_n. The EL bootstrap I propose here does basically the opposite.[16] This is, when $\theta = 0$ the EL mean is given by $\bar{\mu}^0_n = \max\{\bar{X}_n, \varrho_n\}$ so that it keeps the mean far from zero when $\bar{X}_n$ is too small.

Note that by (4.1) and the law of iterated logarithm,

$$P_0\left(\limsup_{n\to\infty}(\bar{X}_n - \theta - \varrho_n) \leq 0\right) = \begin{cases} 0 & \text{if } \theta < 0 \\ 1 & \text{if } \theta = 0, \end{cases}$$

meaning that $\bar{\mu}^\theta_n > \varrho_n$ for n large enough with probability one (w.p.1) when $\theta < 0$ and $\bar{\mu}^\theta_n = \varrho_n$ for n large enough w.p.1 when $\theta = 0$. Hence, for all $\theta \leq 0$,

$$\begin{aligned} n\hat{Q}^*_n(\theta) &= (\sqrt{n}(\bar{X}^*_n - \theta - \varrho_n))^2_- \\ &= (\sqrt{n}(\bar{X}^*_n - \theta - \bar{\mu}^\theta_n) + \sqrt{n}(\bar{\mu}^\theta_n - \varrho_n))^2_- \\ &\rightsquigarrow (\mathbb{Z})^2_- \times 1\{\theta = 0\}, \quad \text{conditional on } \{\bar{P}_n : n \geq 1\}, \end{aligned}$$

and the new bootstrap is asymptotically valid to approximate the distribution of $n\hat{Q}_n(\theta)$.

4.3. Modified EL bootstrap validity: The general case

Now consider the general case where the economic model imposes $\mathbb{E}_{P_0}[m(z,\theta)] \geq 0$. The modified empirical likelihood function is,

$$\tilde{l}^r_{EL}(\theta, \varrho^\theta_n) \equiv \max_{p_1,\ldots,p_n} \left\{ \sum_{i=1}^n \log(p_i) \middle| p_i > 0; \sum_{i=1}^n p_i = 1; \sum_{i=1}^n p_i m(z_i, \theta) \geq \varrho^\theta_n \right\} \quad (4.2)$$

where $\varrho^\theta_n \geq 0$ is a $q \times 1$ vector of random (or deterministic) variables that satisfies

$$P_0\left(\lim_{n\to\infty} \varrho^\theta_{j,n} = 0 \text{ and } \liminf_{n\to\infty} \varrho^\theta_{j,n}(n/(2\log\log n))^{1/2} \geq V^{1/2}_{jj}(\theta)\right) = 1 \quad (4.3)$$

where $V_{jj}(\theta) = \mathrm{Var}_{P_0}\{m_j(z_i,\theta)\}$ for all $j \in \{1, \ldots, q\}$. By the law of iterated logarithm,

$$P_0\left(\limsup_{n\to\infty}(\bar{m}_{j,n}(\theta) - \varrho^\theta_{j,n}) \leq 0\right) = \begin{cases} 0 & \text{if } \mathbb{E}_{P_0}[m_j(z,\theta)] > 0 \\ 1 & \text{if } \mathbb{E}_{P_0}[m_j(z,\theta)] = 0. \end{cases} \quad (4.4)$$

Denote the modified EL probabilities that solve (4.2) by $\bar{P}_n(\theta) = (\bar{p}_1, \ldots, \bar{p}_n)$ and define $\bar{\mu}_n(\theta) = \sum_{i=1}^n \bar{p}_i m(z_i, \theta)$. Let $\{Z^*_i : i \leq n\}$ be i.i.d. according to $\bar{P}_n(\theta)$ so that $\mathbb{E}_{\bar{P}_n}[m(z^*, \theta)] = \bar{\mu}_n(\theta)$. The modified bootstrap empirical likelihood function is,

$$\tilde{l}^{r*}_{EL}(\theta, \varrho^\theta_n) \equiv \max_{p_1,\ldots,p_n} \left\{ \sum_{i=1}^n \log(p_i) \middle| p_i > 0; \sum_{i=1}^n p_i = 1; \sum_{i=1}^n p_i(m(z^*_i, \theta) - \varrho^\theta_n) \geq 0 \right\}$$

and the modified bootstrap ELR statistic is,

$$\mathcal{ELR}^*_n(\theta, \varrho^\theta_n) \equiv \max_{\lambda \leq 0} 2 \sum_{i=1}^n \log[1 + \lambda'(m(z^*_i, \theta) - \varrho^\theta_n)]. \quad (4.5)$$

Letting N_n be the number of bootstrap replications, the $(1-\alpha)$-quantile of $\{\mathcal{ELR}^*_{n,k}(\theta, \varrho^\theta_n) : k \leq N_n\}$ is given by,

$$\bar{c}^\theta_{n,1-\alpha} \equiv \inf\left\{ x : \frac{1}{N_n}\sum_{k=1}^{N_n} 1\{\mathcal{ELR}^*_{n,k}(\theta, \varrho^\theta_n) \leq x\} \geq 1 - \alpha \right\},$$

so that the bootstrap confidence region is just,

$$\mathcal{C}_n \equiv \{\theta \in \Theta : \mathcal{ELR}_n(\theta) \leq \bar{c}^\theta_{n,1-\alpha}\}. \quad (4.6)$$

The next theorem shows that the confidence set $\mathcal{C}_n$ defined in (4.6) satisfies the pointwise coverage requirement for all $\Theta_0(P_0)$ and it is uniformly valid over $\mathcal{F}_0$.

Theorem 4.1. *Suppose Assumption 3.1 holds and let $\varrho^\theta_n \geq 0$ be the vector satisfying (4.3). Then, for $0 < \alpha < 1/2$, the statements below follow.*

I. *For all $\theta \in \Theta_0(P_0)$ the modified bootstrap ELR statistic defined in (4.5) satisfies*

$$\mathcal{ELR}^*_n(\theta, \varrho^\theta_n) \rightsquigarrow \overline{\chi}^2_{b(\theta)}(\Sigma_b(\theta))$$

where the convergence is conditional on $\{\bar{P}_n(\theta) : n \geq 1\}$ for almost every sample path.

II. *$\mathcal{C}_n$ as defined in (4.6) satisfies*

$$\liminf_{n\to\infty} P_0(\theta \in \mathcal{C}_n) \geq 1 - \alpha \quad (4.7)$$

for all $\theta \in \Theta_0(P_0)$.

III. *$\mathcal{C}_n$ as defined in (4.6) satisfies*

$$\liminf_{n\to\infty} \inf_{(\theta,P)\in\mathcal{F}_0} P(\theta \in \mathcal{C}_n) = 1 - \alpha. \quad (4.8)$$

Proof. See Appendix A. ■

Remark 4.2. The statements in (4.7) can be decomposed into two parts. When θ is on the boundary of $\Theta_0(P_0)$ it is the case that $\liminf_{n\to\infty} P_0(\theta \in \mathcal{C}_n) = 1-\alpha$. However, when θ is in the interior of $\Theta_0(P_0)$, $\mathcal{C}_n$ includes θ w.p. $\to 1$.

Remark 4.3. The sequence $\varrho^\theta_{j,n}$ provides a rule to determine whether the jth moment is binding or not. Thus, this sequence is similar to $\kappa_n n^{-1/2}$ in Andrews and Soares (forthcoming, Assumption GMS4), to $\tau_n n^{-1/2}$ in Bugni (forthcoming) and to $c_j\sqrt{\log n/n}$ in Chernozhukov et al. (2007, Remarks 4.5 and 5.1).

---

16 A supplementary appendix shows how the new EL bootstrap works for the example in Andrews (2000).

Remark 4.4. The uniform result in Theorem 4.1 follows almost immediately from Andrews and Soares (forthcoming, Theorem 1) since it turns out that the modified EL bootstrap falls into the class of GMS procedures. I include it here for completeness and in the Appendices I show the adjustments to make the modified EL bootstrap fit into the GMS family.

Remark 4.5. Andrews and Soares (forthcoming) recently proved that, keeping a given test statistic fixed, tests based on a generalized moment selection (GMS) critical values have greater power than tests based on subsampling or fixed asymptotic critical values. The modified EL bootstrap falls into the class of GMS tests and so inherits this power property.

As is the case for the asymptotic distribution and subsampling, the application of this bootstrap procedure is not free of tuning parameters. Although any sequence ϱ_n^θ that satisfies the rate in (4.3) will work asymptotically, the finite sample properties of the confidence region might be sensitive to this choice. One advantage here is that there is a value of ϱ_n^θ that can serve as a natural benchmark. This value is given by,

$$\bar{\varrho}_{j,n}^\theta \equiv n^{-1}(2\log\log n \times V_{jj}(\theta))^{1/2}, \tag{4.9}$$

which represents a lower bound for $\varrho_{j,n}^\theta$, since $\varrho_{j,n}^\theta = C \times \bar{\varrho}_{j,n}^\theta$ for $C < 1$ does not satisfy (4.3). There is no reason to believe that $\bar{\varrho}_n^\theta$ is optimal in any sense but having a reference value might in fact be useful in practice. Optimal choice of ϱ_n^θ is beyond the scope of this paper.[17] In the examples of Section 5 I find that the benchmark value performs well, and that coverage appears reasonably robust to this choice.

Finally, it is worth mentioning that the results from Romano and Shaikh (2008) and Andrews and Guggenberger (2009) can be applied to show that subsampling is a uniformly valid approach to approximate the distribution of $\mathcal{ELR}_n(\theta)$ and thus construct confidence regions for θ_0. The main practical problem in using subsampling lies in choosing the block size $a_n = o(n)$. The simulations of the next section suggest that finite sample coverage of the confidence regions might be more sensitive to the choice of a_n than the choice of ρ_n^θ.

5. Monte Carlo simulations

5.1. Missing data example

This section takes the setup from Example 1 to evaluate the finite sample performance of the tools developed in the previous sections. Each simulation experiment depends on three parameters: the sample size n, the size of the test α and the propensity score p. I set $p = 0.7$ and I consider $n = \{100, 500, 1000\}$ and $1 - \alpha = \{0.85, 0.90, 0.95\}$.[18] I then take independent draws of $x_i \sim$ Uniform(0,1) and $w_i \sim$ Bernoulli(p) to construct the simulated missing data as $\{(x_i w_i, w_i) : i \le n\}$, resulting in the identified set $\Theta_0(P_0) = [\theta_L, \theta_H] = [0.35, 0.65]$. The number of Monte Carlo replications is equal to 3000.

I compute three empirical likelihood confidence regions that use different critical values. The first one uses an asymptotic critical value. Since the maximum number of binding constraints is one, i.e. $b^* = 1$, I can use the limit distribution in Theorem 3.1 without the need of any approximation,

$$\lim_{n\to\infty} P_0(\mathcal{ELR}_n(\theta) \ge c) = \frac{1}{2}P(\chi_1^2 \ge c).$$

A confidence region for θ_0 simply uses a critical value that solves $P\{\chi_1^2 \ge c\} = 2\alpha$ in this case. The second confidence region uses subsampling to compute the critical value. I use four different subsample sizes, $a_n = \{n^{0.95}/10, n^{0.95}/8, n^{0.95}/6, n^{0.95}/4\}$ with $N_n = 200$ subsamples. Finally, the last set of confidence regions uses the modified empirical likelihood bootstrap introduced in Section 4. In this case I set $N_n = 200$ and use four different values for the parameter $\varrho_n^\theta = \{\bar{\varrho}_n^\theta, 1.5\bar{\varrho}_n^\theta, 2\bar{\varrho}_n^\theta, 3\bar{\varrho}_n^\theta\}$, where $\bar{\varrho}_n^\theta$ is defined in (4.9).

Table 1 shows the empirical coverage and the average power over the set of points {0.27, 0.29, 0.31, 0.33, 0.67, 0.69, 0.71, 0.73} outside $\Theta_0(P_0)$. The results show that the modified empirical likelihood bootstrap performs well. A comparison between the bootstrap and subsampling shows that for small sample sizes the bootstrap represents a better approximation. Subsampling does really well in some cases (e.g., $n = 500$ and $\alpha = 0.15$) although given a value of a_n its performance varies as n varies. In the case of the bootstrap on the other hand, if we just focus on the benchmark case bootstrap_1 where $\varrho_n^\theta = \bar{\varrho}_n^\theta$, the bootstrap performs very well across n.[19] Note also that in most cases the modified bootstrap approach performs as well as the asymptotic approximation. This is worth noticing since the bootstrap requires no knowledge on the number of binding constraints. Finally, the average power of the bootstrap dominates almost uniformly both subsampling and the asymptotic approximation. This is consistent with Remark 4.5. The gains in size and power over subsampling cost additional computational time. While each bootstrap case takes 2 ($n = 100$), 6.5 ($n = 500$) and 32.5 ($n = 1000$) s to compute the size for each MC round, each subsampling case takes 1, 1.5 and 2 s respectively.[20]

Table 2 shows the finite sample power (not adjusted for size) of EL versus alternative criterion functions. In this case $n = 100$, $\alpha = \{0.15, 0.05\}$, with 10,000 replications. The columns labeled Asymptotic compare the performance of the ELR statistic versus the Gaussian quasi-likelihood ratio (QLR) statistic associated with $Q_2(\theta, P_0)$ in Example 3. Since both of these statistics have the same asymptotic distribution, I use a common critical value from $P\{\chi_1^2 \ge c\} = 2\alpha$. The columns under the label Subsampling compare the ELR statistic and the sample analog of $Q_1(\theta, P_0) \equiv \|(\mathbb{E}_{P_0}[m(z,\theta)])_-\|^2$ from Example 3. Here, the critical value is computed using subsampling with $a_n = 25$ and $N_n = 200$. Given the 10,000 Monte Carlo replications it follows that the gains in power by EL are marginal for values close to 0.65, but significant for most cases when the alternative is far from the null.[21] This fact is consistent with the results in Theorem 3.2. As before, the associated cost is computational time. In the asymptotic case, EL takes 0.05 s to test the 11 points in the table while QLR takes 0.01 s. For subsampling EL takes 10 s to go over the 11 points 200 times, while the statistic Q_1 (which does not involve any optimization) takes 0.03 s.

5.2. Entry game example

This section takes the simultaneous entry game from Example 2. I set $\theta_0 = (0.3, 0.5)$ and consider $n = 1000$ and $\alpha = \{0.85, 0.95\}$.[22] The four possible Nash Equilibria (NE) situations are

17 Andrews and Jia (2008) recently proposed a way to choose these type of tuning parameters.

18 Different values of p yield similar results so they are not reported.

19 Note that Bugni (2009) shows that while the rate of approximation of subsampling depends on a_n, the rate of approximation of his bootstrap approach does not depend on τ_n, a sequence which has a similar role to ρ_n^θ in this paper.

20 All the computations were carried out using R 2.8 on a Mac Pro computer with two 2.8 GHz quad-core processors and 8 GB 800 MHz of ram. The extra time the bootstrap takes is greatly affected by the fact that I use a loop to resample data with probability $\bar{P}_n$. Computational time could be reduced by avoiding this loop.

21 Note that in this case the maximal simulation standard errors are $\sqrt{0.5 \times 0.5/10{,}000} = 0.005$.

22 Results for $n = 500$ are similar and not reported to save space.

Table 1
Missing data example. Coverage and average power. Coverage is computed as $\min\{P(\theta_L \in \mathcal{C}_n), P(\theta_H \in \mathcal{C}_n)\}$. Four values of a_n for subsampling and four values of ρ_n for the Bootstrap. 200 bootstrap/subsampling replications. 3000 MC replications.

Coverage	0.85			0.90			0.95		
Sample size	100	500	1000	100	500	1000	100	500	1000
Asymptotic	0.8380	0.8434	0.8510	0.8900	0.8884	0.8910	0.9443	0.9450	0.9440
Subsampling_1	0.9073	0.8703	0.8323	0.9540	0.9070	0.8583	0.9837	0.9454	0.8953
Subsampling_2	0.9073	0.8667	0.8273	0.9527	0.9027	0.8513	0.9803	0.9380	0.8837
Subsampling_3	0.9120	0.8577	0.8073	0.9530	0.8964	0.8303	0.9793	0.9317	0.8623
Subsampling_4	0.9194	0.8517	0.7907	0.9590	0.8817	0.8067	0.9833	0.9154	0.8373
Bootstrap_1	0.8250	0.8377	0.8443	0.8780	0.8860	0.8805	0.9360	0.9404	0.9420
Bootstrap_2	0.8270	0.8373	0.8477	0.8773	0.8840	0.8843	0.9337	0.9380	0.9387
Bootstrap_3	0.8274	0.8330	0.8487	0.8777	0.8843	0.8847	0.9343	0.9387	0.9370
Bootstrap_4	0.8223	0.8283	0.8443	0.8807	0.8780	0.8830	0.9337	0.9330	0.9407
Average power over {0.27, 0.29, 0.31, 0.33, 0.67, 0.69, 0.71, 0.73}									
Asymptotic	0.6564	0.8928	0.9428	0.5822	0.8613	0.9315	0.4713	0.8088	0.8964
Subsampling_1	0.5325	0.8433	0.9178	0.4268	0.8017	0.9038	0.2730	0.7446	0.8779
Subsampling_2	0.5217	0.8365	0.9137	0.4167	0.7953	0.9017	0.2815	0.7423	0.8793
Subsampling_3	0.5036	0.8265	0.9115	0.3996	0.7898	0.9015	0.2707	0.7395	0.8807
Subsampling_4	0.4662	0.8102	0.9105	0.3609	0.7751	0.8968	0.2370	0.7278	0.8778
Bootstrap_1	0.6647	0.8959	0.9418	0.5942	0.8638	0.9321	0.4876	0.8128	0.8998
Bootstrap_2	0.6708	0.8942	0.9453	0.6000	0.8645	0.9327	0.4933	0.8133	0.9002
Bootstrap_3	0.6694	0.8950	0.9453	0.5980	0.8651	0.9328	0.4910	0.8156	0.9007
Bootstrap_4	0.6700	0.9002	0.9451	0.5987	0.8727	0.9329	0.4891	0.8254	0.9004

Table 2
Missing data example. Power across different criterion functions. 200 subsampling replications with $a_n = 25$. 10,000 MC replications.

Theta	Asymptotic $\alpha = 0.15$		Subsampling $\alpha = 0.15$		Asymptotic $\alpha = 0.05$		Subsampling $\alpha = 0.05$	
	ELR	QLR	ELR	Q_1	ELR	QLR	ELR	Q_1
0.65	0.1508	0.1467	0.0576	0.0559	0.0505	0.0473	0.0097	0.0085
0.66	0.2278	0.2239	0.0947	0.0908	0.0932	0.0869	0.0204	0.0191
0.67	0.3213	0.3149	0.1479	0.1460	0.1518	0.1425	0.0354	0.0343
0.74	0.9606	0.9580	0.8208	0.8043	0.8786	0.8643	0.5274	0.4879
0.75	0.9799	0.9785	0.8839	0.8686	0.9353	0.9246	0.6439	0.5924
Average Power over ten points in [0.65, 0.75]								
	0.6789	0.6732	0.4760	0.4647	0.5161	0.5009	0.2501	0.2324

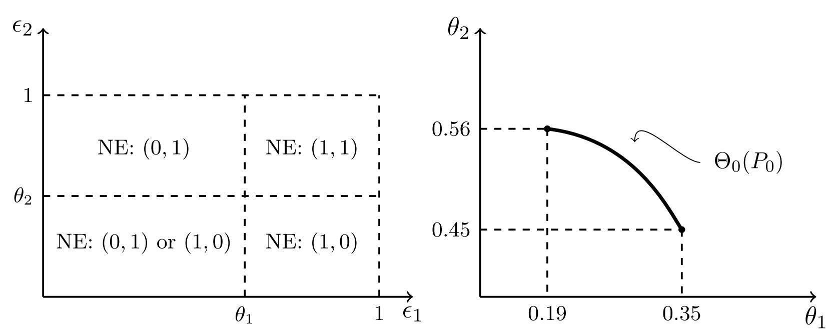

Fig. 1. Left: four possible outcomes of the entry game. Right: The identified set when $\theta_0 = (0.5, 0.3)$.

presented in Fig. 1 (left panel). This model imposes one moment equality and two moment inequalities, as described in Eq. (2.4).[23] In addition, the identified set $\Theta_0(P_0)$ is given in Eq. (2.5). The set $\Theta_0(P_0)$ has no interior, as Fig. 1 (right panel) shows, for the parameter values used in the simulation.

I simulate the data by taking independent draws of $\varepsilon_{j,m} \sim U(0, 1)$ for $j = \{1, 2\}$ and computing the equilibrium according to the region in which $\varepsilon_m = (\varepsilon_{1,m}, \varepsilon_{2,m})$ falls. In the region of multiple equilibria, I select $(1, 0)$ with probability 0.7. Note that in order to simulate the data I need an equilibrium selection mechanism, but this information is not used for inference. This is precisely why the model is not point identified. Intuitively, each θ in the identified set is consistent with some selection mechanism in the region of uncertainty.

As before, I compute the ELR statistic with three alternative critical values. For the asymptotic critical value I use the following approximation from Rosen (2008) to deal with the non-pivotalness of the asymptotic distribution,

$$\sup_{\theta\in\Theta_0(P_0)} \lim_{n\to\infty} P_0(\mathcal{ELR}_n(\theta) \geq c)$$

$$\leq \frac{1}{2}\Pr(\chi^2_{b^*} \geq c) + \frac{1}{2}\Pr(\chi^2_{b^*-1} \geq c), \quad (5.1)$$

where $b^* = 2$ since there is no θ such that both inequalities are

[23] Handling moment equalities with the approach this paper proposes is straightforward: moment equalities have unrestricted Lagrange multipliers and $\varrho_n^\theta = 0$ for the bootstrap implementation.

Table 3
Entry game example. Coverage for a grid of values in $\Theta_0(P_0)$. 3000 MC replications. 200 bootstrap/subsampling replications.

θ_1	θ_2	$n = 1000$ — coverage = 0.85			$n = 1000$ — coverage = 0.95		
		Asymp	Boot	SubSam	Asymp	Boot	SubSam
0.3578	0.4550	0.8530	0.8497	0.8027	0.9460	0.9377	0.8793
0.3519	0.4600	0.8893	0.8810	0.8270	0.9653	0.9587	0.8973
0.3458	0.4650	0.9090	0.9037	0.8330	0.9703	0.9677	0.8977
0.3396	0.4700	0.9180	0.9093	0.8317	0.9720	0.9717	0.8983
0.3333	0.4750	0.9200	0.9107	0.8273	0.9727	0.9673	0.8933
0.3269	0.4800	0.9203	0.9083	0.8190	0.9727	0.9653	0.8887
0.3204	0.4850	0.9203	0.8950	0.8140	0.9727	0.9650	0.8830
0.3137	0.4900	0.9203	0.8887	0.8080	0.9727	0.9597	0.8793
0.3069	0.4950	0.9203	0.8770	0.8020	0.9727	0.9540	0.8753
0.3000	0.5000	0.9203	0.8720	0.7997	0.9727	0.9510	0.8713
0.2929	0.5050	0.9203	0.8620	0.7960	0.9727	0.9497	0.8687
0.2857	0.5100	0.9203	0.8560	0.7943	0.9727	0.9437	0.8660
0.2784	0.5150	0.9203	0.8543	0.7907	0.9727	0.9443	0.8660
0.2708	0.5200	0.9203	0.8507	0.7900	0.9727	0.9453	0.8653
0.2632	0.5250	0.9203	0.8557	0.7897	0.9727	0.9433	0.8640
0.2553	0.5300	0.9203	0.8560	0.7913	0.9727	0.9450	0.8657
0.2473	0.5350	0.9203	0.8593	0.7940	0.9727	0.9457	0.8663
0.2391	0.5400	0.9203	0.8703	0.7977	0.9727	0.9517	0.8690
0.2308	0.5450	0.9203	0.8863	0.8020	0.9727	0.9573	0.8757
0.2222	0.5500	0.9203	0.9010	0.8110	0.9727	0.9643	0.8837
0.2135	0.5550	0.9183	0.9093	0.8247	0.9723	0.9697	0.8903
0.2045	0.5600	0.9103	0.9040	0.8243	0.9693	0.9640	0.8933
0.1954	0.5650	0.8570	0.8530	0.7910	0.9520	0.9467	0.8757
Min. coverage		0.8530	0.8497	0.7897	0.9460	0.9377	0.8640
Ave. coverage		0.9122	0.8788	0.8070	0.9700	0.9552	0.8788

binding simultaneously. Thus, I compute c by solving:

$$\frac{1}{2}\Pr(\chi_2^2 \geq c) + \frac{1}{2}\Pr(\chi_1^2 \geq c) = \alpha.$$

For the modified empirical likelihood bootstrap I use the benchmark value $\varrho_n^\theta = \bar{\varrho}_n^\theta$ while for subsampling I use $a_n = 100$.[24] In both of these cases $N_n = 200$.

Table 3 shows the coverage over a grid of values of $\theta \in \Theta_0(P_0)$ equally spaced between (0.19, 0.56) and (0.35, 0.45) — see Fig. 1. Note that at the true parameter value $\theta_0 = (0.3, 0.5)$, the asymptotic approximation that uses $b^* = 2$ is too conservative (0.9203) since at that point only one moment is binding. The bootstrap captures this difference and so it gives a better approximation (0.872). For this model, the bootstrap takes 16 min to compute coverage for the entire grid while subsampling takes 1 min.

6. Concluding remarks

This paper presents results about the optimal choice of criterion function in moment inequality models and introduces empirical likelihood as a new inference tool for such models. Inference based on the empirical likelihood ratio statistic is shown to be optimal in a Generalized Neyman–Pearson sense. The paper also addresses implementation of the new optimal procedure by proposing an empirical likelihood bootstrap for moment inequality models.

There are a number of directions for future research. First, it would be interesting to explore alternative notions of asymptotic optimality to differentiate between test statistics in these models. Second, the current large deviation optimality result requires a fixed (non data-dependent) critical value η while the recommended procedure employs the bootstrap. It would be interesting then to extend the result to allow for a data-dependent η. Third, the proposed EL bootstrap involves choosing the tuning parameter ϱ_n for its application. Optimal choice of such a sequence would represent an important contribution for practitioners. The recent paper by Andrews and Jia (2008) addresses this issue for GMS tests in general.

[24] 10% of the sample size is a common value of a_n in applications (e.g., Ciliberto and Tamer, forthcoming).

Acknowledgements

I am deeply grateful to Jack Porter and Bruce Hansen for thoughtful discussions and constant encouragement. I also want to acknowledge helpful conversations with Don Andrews, Steven Durlauf, Dennis Kristensen, Elie Tamer, Azeem Shaikh and Ken West, as well as the good feedback provided by seminar participants at the 2007 Summer Meeting of the Econometric Society, UW-Madison, UCL, Northwestern, Brown, Boston University, Columbia, Harvard, Duke, Chicago GSB, Penn State, UC Berkeley and Michigan. Finally, I want to thank two anonymous referees whose comments have led to an improved version of this paper.

Appendix

Throughout the appendix I partition a given vector v_n as $v_n' = [v_{n,b}', v_{n,s}']$, where $v_{n,b}$ is a $b \times 1$ vector associated with binding moments and $v_{n,s}$ is a $s \times 1$ vector associated with slack moments. Typical elements are denoted by $v_{n,b,j}$ and $v_{n,s,j}$. I use $\mu_P(\theta)$ to denote $\mathbb{E}_P[m(z,\theta)]$ and $\sigma_P^2(\theta)$ to denote $\text{var}_P\{m(z,\theta)\}$. When P_0 is understood, the subscript is omitted. Finally, $\bar{P}_n(\theta)$-a.e. denotes "conditional on $\{\bar{P}_n(\theta) : n \geq 1\}$ for almost every sample path".

Appendix A. Proof of theorems

Proof of Theorem 3.1. To derive the asymptotic distribution of $\mathcal{ELR}_n(\theta)$ it is convenient to write an alternative parametrization as follows,

$$\mathcal{ELR}_n(\theta) = \min_{\tau \geq 0} \max_{\lambda(\theta,\tau) \in \mathbb{R}^q} \left\{ 2 \sum_{i=1}^{n} \log(1 + \lambda(\theta,\tau)' m(z_i,\theta)) - 2\lambda(\theta,\tau)' n\tau \right\} \equiv \min_{\tau \geq 0} \max_{\lambda(\theta,\tau) \in \mathbb{R}^q} R_n(\theta,\lambda,\tau).$$

Note that the derivative with respect to τ is non-negative, $\frac{\partial}{\partial\tau}R_n(\theta,\lambda,\tau)\geq 0$, and this in turn imposes $\lambda\leq 0$ and the moment condition. Note also that $\widehat{\tau}=\sum_{i=1}^n\widehat{p}_i m(z_i,\theta)\geq 0$.

Now consider an element j of $m_s(z,\theta)$. For such moment $\exists N$ s.t. $\forall n\geq N$, $\bar{m}_{n,s,j}(\theta)>0$ with probability one (w.p.1). This implies that $\forall n\geq N$, $\hat{\tau}_{n,s,j}>0$ w.p.1 since,

$$\hat{\tau}_{n,s,j}=\frac{1}{n}\sum_{i=1}^n\left(\frac{m_{s,j}(z_i,\theta)}{1+\hat{\lambda}_n(\theta)'m(z_i,\theta)}\right)\geq\frac{\bar{m}_{n,s,j}(\theta)}{1+\hat{\lambda}_n(\theta)'\bar{m}_n(\theta)} \tag{A.1}$$

by Jensen's inequality, $\hat{\lambda}_n(\theta)\leq 0$ and $\hat{\lambda}_n(\theta)'\bar{m}_n(\theta)\geq 0$. The latter follows from

$$0\leq\max_{\lambda\leq 0}2\sum_{i=1}^n\log[1+\lambda(\theta)'m(z_i,\theta)]$$
$$\leq 2\log\left[1+\hat{\lambda}_n(\theta)'\sum_{i=1}^n m(z_i,\theta)\right].$$

Given that $\hat{\tau}_{n,s,j}\times\hat{\lambda}_{n,s,j}(\theta)=0$ for all n, it follows that $\hat{\lambda}_{n,s,j}(\theta)=0$ for n large enough with probability one. Conclude that $\hat{\lambda}_{n,s}(\theta)=0$ for n large enough w.p.1. On the other hand, Lemma B.2 shows that $\|\hat{\lambda}_{n,b}(\theta)\|=O_p(n^{-1/2})$. Hence, $\exists N$ s.t. $\forall n\geq N$ the following equalities hold w.p.1.

$$\begin{aligned}\mathcal{ELR}_n(\theta)&=2\sum_{i=1}^n\log(1+\hat{\lambda}_{n,b}(\theta)'m_b(z_i,\theta)+\hat{\lambda}_{n,s}(\theta)'m_s(z_i,\theta))\\&=2\sum_{i=1}^n\log(1+\hat{\lambda}_{n,b}(\theta)'m_b(z_i,\theta))\\&=\max_{\lambda_b(\theta)\leq 0}2\sum_{i=1}^n\log(1+\lambda_b(\theta)'m_b(z_i,\theta))\\&=\min_{\tau_b\geq 0}\max_{\lambda_b(\theta,\tau_b)\in\mathbb{R}^b}R_n(\theta,\lambda_b,\tau_b).\end{aligned}$$

Now I can make a similar expansion to that in Owen (1990, pages 100–102) for the expression given above. For an arbitrary sequence $0\leq\tau_{n,b}=O_p(n^{-1/2})$, the first order condition for $\hat{\lambda}_{n,b}(\theta,\tau_{n,b})$ is given by:

$$-\sum_{i=1}^n\frac{m_b(z_i,\theta)}{(1+\hat{\lambda}_{n,b}(\theta,\tau_{n,b})'m_b(z_i,\theta))}+n\tau_{n,b}=0. \tag{A.2}$$

Let $\gamma_i\equiv\hat{\lambda}_{n,b}(\theta,\tau_{n,b})'m_b(z_i,\theta)$ and note that by Lemmas B.1 and B.2 we have:

$$\max_{1\leq i\leq n}|\gamma_i|=O_p(n^{-1/2})o_p(n^{1/2})=o_p(1). \tag{A.3}$$

Expanding (A.2):

$$\begin{aligned}0&=-\sum_{i=1}^n m_b(z_i,\theta)(1-\gamma_i+\gamma_i^2/(1+\gamma_i))+n\tau_{n,b}\\&=-\bar{m}_{n,b}(\theta)+\Sigma_{n,b}(\theta)\hat{\lambda}_{n,b}(\theta,\tau_{n,b})\\&\quad-\frac{1}{n}\sum_{i=1}^n m_b(z_i,\theta)\gamma_i^2/(1+\gamma_i)+\tau_{n,b}\\&=-\bar{m}_{n,b}(\theta)+\Sigma_{n,b}(\theta)\widehat{\lambda}_{n,b}(\theta,\tau_{n,b})+\tau_{n,b}+r_{1,n}\end{aligned}$$

where $\Sigma_{n,b}(\theta)\equiv n^{-1}\sum_{i=1}^n m_b(z_i,\theta)m_b(z_i,\theta)'$ and $r_{1,n}\equiv-\frac{1}{n}\sum_{i=1}^n m_b(z_i,\theta)\gamma_i^2/(1+\gamma_i)$. Then,

$$\hat{\lambda}_{n,b}(\theta,\tau_{n,b})=\Sigma_{n,b}^{-1}(\theta)\left(\bar{m}_{n,b}(\theta)-\tau_{n,b}\right)-\Sigma_{n,b}^{-1}(\theta)r_{1,n} \tag{A.4}$$

where

$$\begin{aligned}\left\|r_{1,n}\right\|&\leq\frac{1}{n}\sum_{i=1}^n\|m_b(z_i,\theta)\|^3\left\|\hat{\lambda}_{n,b}(\theta,\tau_{n,b})\right\|^2|1+\gamma_i|^{-1}\\&=O_p(1)O_p(n^{-1})O_p(1)=O_p(n^{-1})\end{aligned} \tag{A.5}$$

by Assumption 3.1. Next, use $\log(1+\gamma_i)=\gamma_i-\gamma_i^2/2+r_{2,i}$ where for some finite $C>0$

$$P\left(\left|r_{2,i}\right|\leq C\left|\gamma_i\right|^3,1\leq i\leq n\right)\to 1 \tag{A.6}$$

as $n\to\infty$. Now I can approximate the likelihood ratio and then use (A.4).

$$\begin{aligned}\mathcal{ELR}_n(\theta)&=\min_{\tau_b\geq 0}2\sum_{i=1}^n\log(1+\hat{\lambda}_{n,b}(\theta,\tau_b)'m_b(z_i,\theta))\\&\quad-2\hat{\lambda}_{n,b}(\theta,\tau_b)'n\tau_b\\&=\min_{\tau_b\geq 0}\Bigg\{2\hat{\lambda}_{n,b}(\theta,\tau_b)'\sum_{i=1}^n m_b(z_i,\theta)\\&\quad-\hat{\lambda}_{n,b}(\theta,\tau_b)'\sum_{i=1}^n m_b(z_i,\theta)m_b(z_i,\theta)'\hat{\lambda}_{n,b}(\theta,\tau_b)\\&\quad-2\hat{\lambda}_{n,b}(\theta,\tau_b)'n\tau_b+2\sum_{i=1}^n r_{2,i}\Bigg\}\\&=\min_{\tau_b\geq 0}\Bigg\{n\hat{\lambda}_{n,b}(\theta,\tau_b)'(\bar{m}_{n,b}(\theta)-\tau_b)\\&\quad+n\hat{\lambda}_{n,b}(\theta,\tau_b)'r_{1,n}+2\sum_{i=1}^n r_{2,i}\Bigg\}\\&=\min_{\tau_b\geq 0}\Bigg\{n(\bar{m}_{n,b}(\theta)-\tau_b)'\Sigma_{n,b}^{-1}(\theta)(\bar{m}_{n,b}(\theta)-\tau_b)\\&\quad-nr_{1,n}'\Sigma_{n,b}^{-1}(\theta)r_{1,n}+2\sum_{i=1}^n r_{2,i}\Bigg\}\\&=T_n^\theta+O_p(n^{-1/2})\end{aligned} \tag{A.7}$$

where

$$T_n^\theta\equiv\min_{\tau_b\geq 0}\left\{n(\bar{m}_{n,b}(\theta)-\tau_b)'\Sigma_{n,b}^{-1}(\theta)(\bar{m}_{n,b}(\theta)-\tau_b)\right\}$$

$$nr_{1,n}'\Sigma_{n,b}^{-1}(\theta)r_{1,n}=nO_p(n^{-1})O_p(1)O_p(n^{-1})=O_p(n^{-1})$$

$$\begin{aligned}\left|\sum_{i=1}^n r_{2,i}\right|&\leq\sum_{i=1}^n|\gamma_i|^3=n\left\|\hat{\lambda}_{n,b}(\theta,\tau_b)'\right\|^3\frac{1}{n}\sum_{i=1}^n\|m_b(z_i,\theta)\|^3\\&=nO_p(n^{-3/2})O_p(1)=O_p(n^{-1/2}).\end{aligned}$$

Next, define

$$\varsigma_{n,b}(\theta)\equiv\sqrt{n}\bar{m}_{n,b}(\theta)\to^p\varsigma_b(\theta)\sim N(0,\Sigma_b(\theta))$$

so that for all $\theta\in\Theta_0(P_0)$

$$\begin{aligned}T_n^\theta&=\min_{\tau_b\geq 0}\left(\varsigma_{n,b}(\theta)-\sqrt{n}\tau_b\right)'\Sigma_{n,b}^{-1}(\theta)\left(\varsigma_{n,b}(\theta)-\sqrt{n}\tau_b\right)\\&\rightsquigarrow\min_{u_b\geq 0}(\varsigma_b(\theta)-u_b)'\Sigma_b^{-1}(\theta)(\varsigma_b(\theta)-u_b)\end{aligned}$$

where $u_b=\sqrt{n}\tau_b$ and $\Sigma_{n,b}(\theta)\to^p\Sigma_b(\theta)$ by Assumption 3.1.

The statistic $T^\theta=\min_{u_b\geq 0}(\varsigma_b(\theta)-u_b)'\Sigma_b^{-1}(\theta)(\varsigma_b(\theta)-u_b)$ measures the distance of the normal random variable $\varsigma_b(\theta)\sim N(0,\Sigma_b(\theta))$ from the nonnegative orthant and by Wolak (1991)

$$P_0\left\{T^\theta>c\right\}=\sum_{j=0}^{b(\theta)}\varpi(b(\theta),b(\theta)-j,\Sigma_b(\theta))\Pr\{\chi_j^2\geq c\}. \quad\blacksquare$$

Proof of Theorem 3.2. Fix $\theta \in \Theta$ and denote by Π the class of all finite measurable partitions of $\mathcal{X}$, where $\mathcal{X} \equiv \{m(z,\theta) : z \in \mathcal{Z}\}$ denotes the support of $m(z,\theta)$. Also, let $\mathcal{Z}_Q$ denote the support of Q. The first step then involves proving that there exists $\eta(\epsilon) > 0$ such that for all $0 < \eta \leq \eta(\epsilon)$, $\Lambda_1^{\theta} \subseteq \Lambda_1^{\theta,*}$ where $\Lambda_1^{\theta,*} \equiv \{Q \in \mathcal{M} : \inf_{P\in\mathcal{P}_{\epsilon,0}(\theta)} I(Q|P) \geq \eta\}$. To this end note that since $Q \in \Lambda_1^{\theta}$, we have $\mu_{Q,j}(\theta) < 0$ for some $j = 1,\ldots,q$ and any probability that has entropy less than η must put mass outside $\mathcal{Z}_Q$. Suppose that indeed $\mathcal{Z}_Q \subset \mathcal{Z}$. By Lemma 1.4.3(g) in Dupuis and Ellis (1997),

$$I(Q|P) = \sup_{\pi\in\Pi} \sum_{A\in\pi} \log\left(\frac{Q(A)}{P(A)}\right) Q(A). \tag{A.8}$$

Taking the partition $\mathcal{X} = \{A, A^c\}$, with $A \equiv \{m(z,\theta) : z \in \mathcal{Z}_Q\}$ it follows that,

$$I(Q|P) \geq Q(A)\log\left(\frac{Q(A)}{P(A)}\right)$$

since $Q(A^c) = 0$. Therefore,

$$\inf_{P\in\mathcal{P}_{\epsilon,0}(\theta)} I(Q|P) \geq -\sup_{P\in\mathcal{P}_{\epsilon,0}(\theta)} \log(P(A)), \tag{A.9}$$

and the more the mass we assign to A, the closer P is to Q. Also by (A.8), the closer $P(B)$ is to $Q(B)$ for B any Borel measurable set in A, the smaller the entropy. Thus, let $c > 0$ and define P_c as the distribution that satisfies $P_c(B) = (1-c)Q(B)$ for any Borel set $B \subset A$ and assigns mass c to a point m_0 in A^c. To be in $\mathcal{P}_{\epsilon,0}(\theta)$, P_c must satisfy the following conditions for $j = 1,\ldots,q$,

$$\mu_{P_c,j}(\theta) \equiv \int m_j(z,\theta)\mathrm{d}P_c = (1-c)\mu_{Q,j}(\theta) + cm_{0,j} \geq 0, \tag{A.10}$$

$$\begin{aligned}\sigma^2_{P_c,j}(\theta) &\equiv \int m_j^2(z,\theta)\mathrm{d}P_c - \mu^2_{P_c,j}(\theta)\\ &= (1-c)\sigma^2_{Q,j}(\theta) + c(1-c)(m_{0,j}-\mu_{Q,j})^2 \geq \epsilon,\end{aligned} \tag{A.11}$$

$$|\mathrm{Corr}_{P_c}[m(z,\theta)]| \geq \epsilon. \tag{A.12}$$

We wish to show that $c \geq \bar{c} > 0$. From condition (A.10) it follows that $m_{0,j} \geq m^*_{0,j} \equiv -(1-c)\mu_{Q,j}(\theta)/c$ and for $m^*_{0,j} \in \mathcal{X}$, it must be that

$$c \geq c_j(\mu) \equiv -\mu_{Q,j}(\theta)/(m_{H,j} - \mu_{Q,j}(\theta)),$$

where $m_{H,j} = \sup_{z\in\mathcal{Z}} m_j(z,\theta)$. This restricts c from below only when $\mu_{Q,j}(\theta) < 0$ and places no restriction otherwise. Now, from condition (A.11),

$$c(1-c)(m_{0,j} - \mu_{Q,j}(\theta))^2 \geq (1-c)(\epsilon - \sigma^2_{Q,j}(\theta)) + c\epsilon. \tag{A.13}$$

Set $m_{0,j} = m_{H,j}1\{\mu_{Q,j}(\theta) < m_{H,j}/2\}$ and suppose that $\sigma^2_{Q,j}(\theta) = 0$. Then, $c(1-c)(m_{0,j} - \mu_{Q,j}(\theta))^2 \geq c(1-c)m^2_{H,j}/4 > \epsilon$, for some $c > 0$ and small ϵ (i.e., $\epsilon < m^2_{H,j}/8$). The same analysis holds for $\sigma^2_{Q,j}(\theta) < \epsilon$. Thus, letting $c_j(\sigma) \equiv \inf\{c \in (0,1) : \sigma^2_{P_c,j}(\theta) \geq \epsilon\}$ condition (A.11) imposes $c \geq c_j(\sigma)$ only when $\sigma^2_{Q,j}(\theta) < \epsilon$ and places no restriction otherwise (when $\sigma^2_{Q,j}(\theta) \geq \epsilon$ the right hand side of (A.13) is $\leq c\epsilon$ and $c_j(\sigma) = 0$). Finally, condition (A.12) imposes the restriction $c \geq c^*$, where $c^* \equiv \inf\{c \in (0,1) : |\mathrm{Corr}_{P_c}[m(z,\theta)]| \geq \epsilon\}$ and this represents a restriction to c only if $|\mathrm{Corr}_Q[m(z,\theta)]| < \epsilon$. Putting all restrictions together,

$$c \geq c_Q \equiv \max_j \max\left\{c_j(\mu), c_j(\sigma), c^*\right\}.$$

If it turns out that $c(\epsilon) \equiv \inf\{c_Q : Q \in \Lambda_1\} > 0$, then

$$\inf_{P\in\mathcal{P}_{\epsilon,0}(\theta)} I(\hat{P}_n|P) \geq -\sup_{P\in\mathcal{P}_{\epsilon,0}(\theta)} \log(P(A)) = -\log(1-c(\epsilon)),$$

and the results follows for $\eta(\epsilon) = -\log(1-c(\epsilon))$. To see that this is actually the case, suppose by way of contradiction that $c(\epsilon) = 0$. Then there would exists a sequence Q_n in Λ_1 such that $Q_n \rightsquigarrow Q^*$ where $\mathbb{E}_{Q^*}[m(z,\theta)] = 0$, $\int m_j^2(z,\theta)\mathrm{d}Q^* = \epsilon$ and $|\mathrm{Corr}_{Q^*}[m(z,\theta)]| = \epsilon$. By the Portmanteau Lemma, for every $\kappa > 0$ there exists $N(\kappa) \in \mathbb{N}$ such that for all $n > N(\kappa)$,

$$\left|\int m_j^2(z,\theta)\mathrm{d}Q_n - \epsilon\right| < \kappa \quad \forall j = 1,\ldots,q.$$

Pick $\kappa > 0$ small enough such that $\epsilon' \equiv \epsilon - \kappa > 0$. It then follows that for n sufficiently large, $\int m_j^2(z,\theta)\mathrm{d}Q_n \geq \epsilon'$. A similar argument shows that for n sufficiently large, $|\mathrm{Corr}_{Q_n}[m(z,\theta)]| \geq \epsilon'$. Since $\mathbb{E}_{Q_n}[m(z,\theta)] \to 0$, it then follows that zero eventually belongs to the interior of the convex hull of $\mathcal{X}$ and then $\mathcal{P}(Q_n) \neq \emptyset$ for n large enough. Hence, $\inf_{P\in\mathcal{P}(Q_n)} I(Q_n|P) < \infty$ and we can let P_n^* denote the measure that minimizes the entropy. It follows that $I(Q_n|P_n^*) \to 0$ under the sequence Q_n and then for $E_{Q_n}(X)$ small enough $I(Q_n|P_n^*) < \eta$, which in turn violates $Q_n \in \Lambda_1$ for all n. It follows that $c(\epsilon) > 0$ and so for all $\eta \leq \eta(\epsilon) = -\log(1-c(\epsilon))$,

$$\inf_{P\in\mathcal{P}_{\epsilon,0}(\theta)} I(Q|P) \geq \eta. \tag{A.14}$$

The second step is to prove that $\Lambda_1^{\theta,*}$ and $\mathcal{P}_{\epsilon,0}(\theta)$ are compact in the weak topology. Since $\mathcal{M}$ is compact by Assumption 3.2 and Prohorov's Theorem, any subset of $\mathcal{M}$ is sequentially compact and it is enough to prove that these sets are closed. To see that $\mathcal{P}_{\epsilon,0}(\theta)$ is closed define $f_j^a(z,\theta) = |m(z,\theta)|^a$ for $a \geq 1$ and take any sequence in $\mathcal{P}_{\epsilon,0}(\theta)$ such that $P_n \rightsquigarrow P$. We wish to show that $P \in \mathcal{P}_{\epsilon,0}(\theta)$. By Assumption 3.3, $f_j^a(z,\theta) \in \mathcal{C}_b(\mathcal{Z})$ where $\mathcal{C}_b(\mathcal{Z})$ denotes the set of bounded continuous functions on $\mathcal{Z}$ and then the sequence P_n is uniformly integrable. Thus, by Theorem 2.20 in van der Vaart (1998),

$$\mathbb{E}_P[f_j^a(z,\theta)] = \lim_{n\to\infty} \mathbb{E}_{P_n}[f_j^a(z,\theta)],$$

and $\mathcal{P}_{\epsilon,0}(\theta)$ is immediately closed. Now consider $\Lambda_1^{\theta,*}$. Take a sequence $Q_n \in \Lambda_1^{\theta,*}$ such that $Q_n \rightsquigarrow Q$. We wish to show that $Q \in \Lambda_1^*$. Suppose by way of contradiction that $\inf_{P\in\mathcal{P}_{\epsilon,0}(\theta)} I(Q|P) < \eta$. By Dupuis and Ellis (1997, Lemma 1.4.3(b)), $I(Q|\cdot)$ is lower semi continuous so the minimum is attained on $\mathcal{P}_{\epsilon,0}(\theta)$ compact. It then follows that there exits $\kappa > 0$ and $\bar{P} \in \mathcal{P}_{\epsilon,0}(\theta)$ such that $I(Q|\bar{P}) < \eta - \kappa$. By the Donsker–Varadhan variational formula, we can express the relative entropy as

$$I(Q|\bar{P}) = \sup_{g\in\mathcal{C}_b(\mathcal{X})} \left\{\int g\mathrm{d}Q_n - \log\int e^g\mathrm{d}\bar{P}\right\}. \tag{A.15}$$

Take any $g \in \mathcal{C}_b(\mathcal{Z})$ and note that,

$$\begin{aligned}\int g\mathrm{d}Q - \log\int e^g\mathrm{d}\bar{P} &= \int g\mathrm{d}Q_n + \int g(\mathrm{d}Q - \mathrm{d}Q_n)\\ &\quad - \log\int e^g\mathrm{d}\bar{P} < \eta - \kappa,\end{aligned}$$

so that,

$$\int g\mathrm{d}Q_n - \log\int e^g\mathrm{d}\bar{P} < \eta - \left(\kappa + \int g(\mathrm{d}Q - \mathrm{d}Q_n)\right).$$

By Prohorov's Theorem $\int g(\mathrm{d}Q - \mathrm{d}Q_n) \to 0$ as $n \to \infty$ so that for n large enough $\kappa'_g = \kappa + \int g(\mathrm{d}Q - \mathrm{d}Q_n)$ is positive. This means that for all $g \in \mathcal{C}_b(\mathcal{Z})$ there exists $\kappa'_g > 0$ such that

$$\int g\mathrm{d}Q_n - \log\int e^g\mathrm{d}\bar{P} < \eta - \kappa'_g,$$

for n large enough and then

$$\inf_{P\in\mathcal{P}_{\epsilon,0}(\theta)} I(Q_n|P) \le I(Q_n|\bar{P}) = \sup_{g\in\mathcal{C}_b(\mathbb{Z})}\left\{\int g\mathrm{d}Q_n - \log\int e^g\mathrm{d}\bar{P}\right\} < \eta - \inf_{g\in\mathcal{C}_b(\mathbb{Z})}\kappa'_g < \eta,$$

which contradicts $Q_n \in \Lambda_1^{\theta,*}$. Thus, $\Lambda_1^{\theta,*}$ is closed in the weak topology.

The third step is to prove part (I) of the Theorem using Sanov's Theorem. Define,

$$\eta^\theta(P) \equiv \inf_{Q\in\Lambda_1^{\theta,*}} I(Q|P), \quad \forall P\in\mathcal{P}_{\epsilon,0}(\theta),$$

and note that,

$$\begin{aligned}
&\sup_{P\in\mathcal{P}_{\epsilon,0}(\theta)}\limsup_{n\to\infty}\frac{1}{n}\log P^n(\hat{P}_n\in\Lambda_1^\theta)\\
&\le \sup_{P\in\mathcal{P}_{\epsilon,0}(\theta)}\limsup_{n\to\infty}\frac{1}{n}\log P^n(\hat{P}_n\in\Lambda_1^{\theta,*})\\
&\le \sup_{P\in\mathcal{P}_{\epsilon,0}(\theta)} -\inf_{Q\in\Lambda_1^{\theta,*}} I(Q|P)\\
&= -\inf_{P\in\mathcal{P}_{\epsilon,0}(\theta)}\eta^\theta(P),
\end{aligned}$$

where the first inequality follows from $\Lambda_1^\theta\subseteq\Lambda_1^{\theta,*}$, and the second inequality by Sanov's Theorem and the fact that $\Lambda_1^{\theta,*}$ is closed. The result would then follow by proving that,

$$\inf_{P\in\mathcal{P}_{\epsilon,0}(\theta)}\eta^\theta(P)\ge\eta. \tag{A.16}$$

Take any $P\in\mathcal{P}_{\epsilon,0}(\theta)$. Since $I_P(Q)$ is lower semi continuous and $\Lambda_1^{\theta,*}$ is compact, there exists $J_P\in\Lambda_1^{\theta,*}$ such that $\eta^\theta(P)=I(J_P|P)\ge\eta$. Since $\mathcal{P}_{\epsilon,0}(\theta)$ is closed, (A.16) follows.

Part (II) follows from two steps. First, there exists $N_1\in\mathbb{N}$ such that

$$\Lambda_0^{\theta,*}\subseteq\Omega_{n,0}^\theta$$

for all $n>N_1$, where $\Lambda_0^{\theta,*}\equiv\{Q\in\mathcal{M}:\inf_{P\in\mathcal{P}_{\epsilon,0}(\theta)}I(Q|P)<\eta\}$. Suppose it is not so. Then there exists an infinite sequence of measures $\{Q_n\}_{n\in\mathbb{N}}$ such that $Q_n\in\Lambda_0^{\theta,*}$ and $Q_n\in\Omega_{n,1}^\theta$. Since $\mathcal{M}$ is compact in the weak topology by Assumption 3.2, there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ such that $Q_{n_k}\rightsquigarrow\bar{Q}\in\mathcal{M}$. For such $\bar{Q}$ and for $\delta>0$ take the open ball $B(\bar{Q},\delta/2)$. It follows that there exists k_0 such that for all $k\ge k_0$,

$$Q_{n_k}\in B(\bar{Q},\delta/2)\subset\Omega_{n_k,1}^{\theta,\delta}. \tag{A.17}$$

We can then use Sanov's Theorem to note that,

$$\begin{aligned}
&\sup_{P\in\mathcal{P}_{\epsilon,0}(\theta)}\limsup_{n\to\infty}n^{-1}\log P^n(\hat{P}_n\in\Omega_{n,1}^{\theta,\delta})\\
&\ge\sup_{P\in\mathcal{P}_{\epsilon,0}(\theta)}\liminf_{n_k\to\infty}\frac{1}{n_k}\log P^{n_k}(\hat{P}_{n_k}\in\Omega_{n_k,1}^{\theta,\delta})\\
&\overset{(1)}{\ge}\sup_{P\in\mathcal{P}_{\epsilon,0}(\theta)}\liminf_{n_k\to\infty}\frac{1}{n_k}\log P^{n_k}(\hat{P}_{n_k}\in B(\bar{Q},\delta/2))\\
&\overset{(2)}{\ge}\sup_{P\in\mathcal{P}_{\epsilon,0}(\theta)} -\inf_{J\in B(\bar{Q},\delta/2)}I(J|P)\\
&\ge -\inf_{P\in\mathcal{P}_{\epsilon,0}(\theta)}I(Q_{n_{k_0}}|P)>\eta,
\end{aligned} \tag{A.18}$$

where $\overset{(1)}{\ge}$ follows from (A.17), $\overset{(2)}{\ge}$ follows from Sanov's Theorem and (A.18) follows from $Q_{n_{k_0}}\in\Lambda_0^{\theta,*}$. This contradicts

$$\limsup_{n\to\infty}n^{-1}\log P^n(\hat{P}_n\in\Omega_{n,1}^{\theta,\delta})\le-\eta.$$

Therefore, $\Lambda_0^{\theta,*}\subseteq\Omega_{n,0}^\theta$ for all $n\ge N_1$ so that $\forall P\in\mathcal{M}$,

$$\limsup_{n\to\infty}n^{-1}\log P^n(\hat{P}_n\in\Omega_{n,0}^\theta)\ge\limsup_{n\to\infty}n^{-1}\log P^n(\hat{P}_n\in\Lambda_0^{\theta,*}). \tag{A.19}$$

The last step involves showing that $\exists N_2\in\mathbb{N}$ such that for all $n\ge N_2$,

$$P_1^n(\hat{P}_n\in\Lambda_0^{\theta,*})\ge P_1^n(\hat{P}_n\in\Lambda_0^\theta), \tag{A.20}$$

for $P_1\in\mathcal{A}_{\eta,\epsilon}(\theta)$. To see this note that $\hat{P}_n\in\Lambda_0^\theta$ means that there exist P_n^* and $\kappa_1>0$ such that $I(\hat{P}_n|P_n^*)=\inf_{P\in\mathcal{P}(\hat{P}_n)}I(\hat{P}_n|P)=\eta-\kappa_1$. Since $\hat{P}_n\rightsquigarrow P_1$, it follows that for all $\kappa_2>0$ there exists N_2 such that $d(P_1,\hat{P}_n)<\kappa_2$ for all $n\ge N_2$ and then for κ_2 small,

$$d(P_1,P_n^*)\le d(P_1,\hat{P}_n)+d(\hat{P}_n,P_n^*)<\kappa_2+\sqrt{(\eta-\kappa_1)/2}<\sqrt{\eta/2}, \tag{A.21}$$

where the first inequality follows by the triangle inequality and the second one by $d(Q_1,Q_2)\le(I(Q_1|Q_2)/2)^{1/2}$ for two measures Q_1 and Q_2 in $\mathcal{M}$. As $P_1\in\mathcal{A}_{\eta,\epsilon}(\theta)$ and $P_n^*\in\mathcal{P}_0(\theta)$, (A.21) implies $P_n^*\in\mathcal{P}_{\epsilon,0}(\theta)$ for $n\ge N_1$ so (A.19) follows. Part (II) is a consequence of (A.19) and (A.20). ■

Proof of Theorem 4.1. To prove part (I) consider the jth element of $m_s(z,\theta)$. For such moment,

$$\begin{aligned}
\hat{\tau}_{n,s,j}^* &= \frac{1}{n}\sum_{i=1}^n\left(\frac{m_{s,j}(z_i^*,\theta)-\varrho_{n,s,j}^\theta}{1+\hat{\lambda}_n^*(\theta)'(m(z_i^*,\theta)-\varrho_n^\theta)}\right)\\
&\ge\frac{\bar{m}_{n,s,j}^*(\theta)-\varrho_{n,s,j}^\theta}{1+\hat{\lambda}_n^*(\theta)'(\bar{m}_n^*(\theta)-\varrho_n^\theta)}\\
&=\frac{\bar{m}_{n,s,j}^*(\theta)-\bar{\mu}_{n,s,j}(\theta)}{1+\hat{\lambda}_n^*(\theta)'(\bar{m}_n^*(\theta)-\varrho_n^\theta)}+\frac{\bar{\mu}_{n,s,j}(\theta)-\varrho_{n,s,j}^\theta}{1+\hat{\lambda}_n^*(\theta)'(\bar{m}_n^*(\theta)-\varrho_n^\theta)},
\end{aligned} \tag{A.22}$$

where the first term is $O_p(n^{-1/2})\bar{P}_n(\theta)$-a.e. and the second term is positive for n large enough w.p.1 by Lemma B.4. Thus, $\hat{\tau}_{n,s,j}^*>0$ for n sufficiently large $\bar{P}_n(\theta)$-a.e. Furthermore, since $\hat{\tau}_{n,s,j}^*\times\lambda_{n,s,j}^*=0$ for all n, we can conclude that $\lambda_{n,s,j}^*=0$ for n sufficiently large $\bar{P}_n(\theta)$-a.e. for all j such that $\mathbb{E}[m_j(z_i,\theta)]>0$. In addition, by Lemma B.2 it follows $\|\lambda_{n,b}^*\|=O_p(n^{-1/2})\bar{P}_n(\theta)$-a.e. Therefore, similar arguments as those in the proof of Theorem 3.1 show that,

$$\mathcal{ELR}_n^*(\theta,\varrho_n^\theta)\equiv\min_{\tau_b\ge0}\max_{\lambda_b\in\mathbb{R}^b}2\sum_{i=1}^n\log[1+\lambda'_b(m_b(z_i^*,\theta)-\varrho_{b,n}^\theta)]-2n\lambda'_b\tau_b$$

so that the corresponding first order condition is,

$$\sum_{i=1}^n\frac{(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)}{(1+\gamma_i^*)}-n\tau_b=0$$

where $\|\hat{\lambda}_{n,b}^*\|=O_p(n^{-1/2})$, $\gamma_i^*\equiv\hat{\lambda}_{n,b}^{*\prime}(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)$ and $\max_{1\le i\le n}|\gamma_i^*|=o_p(1)$. Therefore, using the same expansion as in (A.7), the statistic $\mathcal{ELR}_n^*(\theta,\varrho_n^\theta)$ is equivalent to a QLR statistic of the form,

$$\begin{aligned}
\mathcal{ELR}_n^*(\theta,\varrho_{n,b}^\theta)&=\min_{\tau_b\ge0}\left\{n\left(\bar{m}_{n,b}^*(\theta)-\varrho_{n,b}^\theta-\tau_b\right)'\tilde{\Sigma}_{n,b}^*(\theta)^{-1}\times(\bar{m}_{n,b}^*(\theta)-\varrho_{n,b}^\theta-\tau_b)\right\}+o_p(1)\\
&=T_n^{\theta*}(\varrho_{n,b}^\theta)+o_p(1),
\end{aligned}$$

where $\tilde{\Sigma}_{n,b}^*(\theta)=n^{-1}\sum_{i=1}^n(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)'$

and $\bar{m}^*_{n,b}(\theta) = n^{-1}\sum_{i=1}^n m_b(z_i^*,\theta)$. Now define the random variable $\varsigma^*_{b,n}(\theta)$ and note that,

$$\varsigma^*_{n,b}(\theta) \equiv \sqrt{n}(\bar{m}^*_{n,b}(\theta) - \bar{\mu}_{n,b}(\theta)) \to^p \varsigma_b(\theta) \sim N(0, \Sigma_b(\theta)) \quad \text{(A.23)}$$

where $\bar{\mu}_{n,b}(\theta) = \sum_{i=1}^n \bar{p}_i m_b(z_i,\theta)$ and the convergence is for all $\theta \in \Theta_0(P_0)\bar{P}_n(\theta)$-a.e. Then, letting $u_b = \sqrt{n}(\tau_b - (\bar{\mu}_{n,b}(\theta) - \varrho^\theta_{n,b}))$

$$T_n^{\theta*}(\varrho^\theta_{n,b}) = \min_{\tau_b \geq 0}\left\{\left(\varsigma^*_{n,b}(\theta) - \sqrt{n}(\varrho^\theta_{n,b} + \tau_b - \bar{\mu}_{n,b}(\theta))\right)' \tilde{\Sigma}^*_{n,b}(\theta)^{-1} \times \left(\varsigma^*_{n,b}(\theta) - \sqrt{n}(\varrho^\theta_{n,b} + \tau_b - \bar{\mu}_{n,b}(\theta))\right)\right\}$$
$$= \min_{u_b \geq -\sqrt{n}(\bar{\mu}_{b,n}(\theta) - \varrho^\theta_{b,n})} \left(\varsigma^*_{n,b}(\theta) - u_b\right)' \tilde{\Sigma}^*_{n,b}(\theta)^{-1} \times \left(\varsigma^*_{n,b}(\theta) - u_b\right).$$

By Lemma B.4, $-\sqrt{n}(\bar{\mu}_{n,b}(\theta) - \varrho^\theta_{n,b}) = 0$ w.p.1 for n sufficiently large, meaning that for n large enough the following equality holds w.p.1:

$$T_n^{\theta*}(\varrho^\theta_{n,b}) = \min_{u_b \geq 0}\left(\varsigma^*_{n,b}(\theta) - u_b\right)' \tilde{\Sigma}^*_{n,b}(\theta)^{-1}\left(\varsigma^*_{n,b}(\theta) - u_b\right).$$

Since $\tilde{\Sigma}^*_{n,b}(\theta) \to^p \Sigma_b(\theta)\bar{P}_n(\theta)$-a.e., it follows that $T_n^{\theta*}(\varrho^\theta_{n,b}) = T^\theta + o_p(1)\bar{P}_n(\theta)$-a.e. where

$$T^\theta = \min_{u_b \geq 0}(\varsigma_b(\theta) - u_b)'\Sigma_b(\theta)^{-1}(\varsigma_b(\theta) - u_b). \quad \text{(A.24)}$$

This is exactly the same asymptotic distribution of $\mathcal{ELR}_n(\theta)$ and the result follows.

Part (II) follows by noting that $T_n^\theta = 0$ wp→1 for any θ in the interior of $\Theta_0(P_0)$, and so,

$$\liminf_{n\to\infty} P_0(\mathcal{ELR}_n(\theta) \leq c) = 1,$$

for $\theta \in int(\Theta_0(P_0))$ and any $c \geq 0$; in particular, for $\bar{c}^\theta_{n,1-\alpha}$ as defined in (4.5). On the other hand, when $\theta \in \partial\Theta_0(P_0)$ it follows from the result in Part I that,

$$\liminf_{n\to\infty} P_0(\mathcal{ELR}_n(\theta) \leq \bar{c}^\theta_{n,1-\alpha}) = 1 - \alpha.$$

Part (III) uses the notation from Appendix C where the EL Bootstrap critical value $\bar{c}_n(\theta, 1-\alpha)$ is the $1-\alpha$ quantile of $T_n^*(\theta) = S(\varsigma_n^*(\theta) + \bar{\varphi}(\xi_n(\theta)), \bar{\Omega}_n^*(\theta))$. Now let $\{\gamma_{n,h} = (\gamma_{n,h,1}, \gamma_{n,h,2}, \gamma_{n,h,3}) : n \geq 1\}$ be a sequence of points in Γ that satisfies (i) $n^{1/2}\gamma_{n,h,1} \to h_1$ for some $h_1 \in R^q_{+,\infty}$, (ii) $\kappa_n^{-1}n^{1/2}\gamma_{n,h,1} \to \pi_1$ for some $\pi_1 \in R^q_{+,\infty}$ and (iii) $\gamma_{n,h,2} \to h_2$ for some $\gamma_{n,h,2} = (\theta_{n,h}, vech_*(\Omega(\theta_{n,h}, P_{n,h}))) \in R^p_{[\pm\infty]}$. Let $h = (h_1, h_2)$, $\pi = (\pi_1, \pi_2)$ and $\pi_2 = h_2$. We wish to show that,

(a) $\bar{c}_n(\theta_{n,h}, 1-\alpha) \geq \bar{c}_n^*$ a.s. for all n for a sequence of r.v. $\{\bar{c}_n^* : n \geq 1\}$ that satisfies $\bar{c}_n^* \to_p c_{\pi^*}(1-\alpha)$ under $\{\gamma_{n,h} : n \geq 1\}$,
(b) $\liminf_{n\to\infty} P_{\gamma_{n,h}}\left(T_n(\theta_{n,h}) \leq \bar{c}_n(\theta_{n,h}, 1-\alpha)\right) \geq 1-\alpha$,
(c) for any subsequence $\{w_n : n \geq 1\}$ of $\{n\}$, the results of parts (a) and (b) hold with w_n in place of n provided conditions (i)–(iii) above hold with w_n in place of n.

Let's first prove (a). If $c_{\pi^*}(1-\alpha) = 0$ define $\bar{c}_n^* = 0$ and so $\bar{c}_n(\theta_{n,h}, 1-\alpha) \geq \bar{c}_n^* \to_p c_{\pi^*}(1-\alpha)$. Next, suppose $c_{\pi^*}(1-\alpha) > 0$. For $\xi \in R^q_{[+\infty]}$, let $\bar{\varphi}^*(\xi)$ denote the k-vector whose jth element is

$$\bar{\varphi}_j^*(\xi) = \begin{cases} \bar{\varphi}_j(\xi) & \text{if } \pi_{1,j} = 0 \text{ and } j = 1,\ldots,q \\ \infty & \text{if } \pi_{1,j} > 0 \text{ and } j = 1,\ldots,q. \end{cases}$$

Now let $\bar{c}_n^*$ denote the $1-\alpha$ quantile of the df of $S(\varsigma_n^*(\theta_{n,h}) + \bar{\varphi}^*(\xi_n(\theta_{n,h})), \bar{\Omega}_n^*(\theta_{n,h}))$, where ς_n^* is random and $\xi_n(\theta_{n,h})$ is fixed. Since $\bar{\varphi}^*(\xi_n(\theta_{n,h})) \geq \bar{\varphi}(\xi_n(\theta_{n,h}))$, a.s. $[Z^*]$ by construction, it follows that,

$$\bar{c}_n(\theta_{n,h}, 1-\alpha) \geq \bar{c}_n^* \quad \text{a.s. for all } n.$$

Now I show that $\bar{c}_n^* \to^p c_{\pi^*}(1-\alpha) > 0$. Under $\{\gamma_{n,h} : n \geq 1\}$,

$$(\xi_n(\theta_{n,h}), \bar{\Omega}_n^*(\theta_{n,h})) \to_p ((\pi_1, 0_v), \Omega_{h_{2,2}}), \bar{P}_n(\theta_{n,h})\text{-a.e.}. \quad \text{(A.25)}$$

Consider j for which $\pi_{1,j} = 0$. Then, as $\xi \to \pi_1$ it follows $\bar{\varphi}_j^*(\xi) = \bar{\varphi}_j(\xi) \to \bar{\varphi}_j(\pi_1) = 0$ a.s. $[Z^*]$ since the function $\bar{\varphi}_j^*(\xi)$ satisfies Assumption GMS1 in AS. Next consider j for which $\pi_{1,j} > 0$. In such case $\bar{\varphi}_j^*(\xi) = \infty$. This result plus the fact that $S(m,\Omega)$ is continuous in m imply that for $x > 0$, as $(\xi, \Omega) \to (\pi, \Omega_0)$,

$$S(\Omega^{1/2}Z^* + \bar{\varphi}^*(\xi), \Omega) \to S(\Omega_0^{1/2}Z^* + \bar{\varphi}^*(\pi_1), \Omega_0) \quad \text{a.s. } [Z^*],$$
$$1\left(S(\Omega^{1/2}Z^* + \bar{\varphi}^*(\xi), \Omega) \leq x\right) \to 1\left(S(\Omega_0^{1/2}Z^* + \bar{\varphi}^*(\pi_1), \Omega_0) \leq x\right) \quad \text{a.s. } [Z^*],$$
$$P\left(S(\Omega^{1/2}Z^* + \bar{\varphi}^*(\xi), \Omega) \leq x\right) \to P\left(S(\Omega_0^{1/2}Z^* + \bar{\varphi}^*(\pi_1), \Omega_0) \leq x\right), \quad \text{(A.26)}$$

following the same arguments as those in AS. This shows that $P\left(S(\Omega^{1/2}Z^* + \bar{\varphi}^*(\xi), \Omega) \leq x\right)$ is a continuous function of (ξ, Ω) at (π_1, Ω_0). This result plus conditions (vi′)–(ix′) from Appendix C imply that under $\{\gamma_{n,h} : n \geq 1\}$ for all $x > 0$,

$$L_n(x) = P\left(S(\varsigma_n^*(\theta_{n,h}) + \bar{\varphi}^*(\xi_n(\theta_{n,h})), \bar{\Omega}_n^*(\theta_{n,h})) \leq x\right) \to P\left(S(\Omega_0^{1/2}Z^* + \bar{\varphi}^*(\pi_1), \Omega_0) \leq x\right) = L(x), \bar{P}_n(\theta_{n,h})\text{-a.e.} \quad \text{(A.27)}$$

where $\bar{\Omega}_n^*(\theta_{n,h}) \to_p \Omega_{h_{2,2}} = \Omega_0$, $\bar{P}_n(\theta_{n,h})$-a.e. by (viii′). By definition $\bar{c}_n^*$ is the $1-\alpha$ quantile of $L_n(x)$ and $c_{\pi^*}(1-\alpha)$ is the $1-\alpha$ quantile of $L(x)$. By Andrews and Guggenberger (forthcoming, Lemma 5), $\bar{c}_n^* \to_p c_{\pi^*}(1-\alpha)$ completing the proof. Parts (b) and (c) follow by the same arguments in those in AS once part (a) holds for $\bar{c}_n(\theta_{n,h}, 1-\alpha)$. It then follows that Lemmas 2 and 3 in AS holds for $\bar{c}_n(\theta_{n,h}, 1-\alpha)$ in place of $\hat{c}_n(\theta_{n,h}, 1-\alpha)$, and Theorem 1 in AS holds for the Modified EL Bootstrap. ■

Appendix B. Auxiliary lemmas

Lemma B.1. *Let $z_1,\ldots,z_n$ be i.i.d. If $\sup_{\theta\in\Theta_0(P_0)}\mathbb{E}[\|m(z_i,\theta)\|^a] < \infty$ for some $a > 0$ then for all $\theta \in \Theta_0(P_0)$ and n large enough,*

$$P_0\{\max_{1\leq i\leq n}\|m(z_i,\theta)\| = o(n^{1/a})\} = 1.$$

Proof. This proof follows Owen (1990, Lemma 3). Let $\epsilon > 0$. Since $\{z_i : i \leq n\}$ is i.i.d. and

$$\sup_{\theta\in\Theta_0(P_0)}\sum_{n=1}^\infty P_0\{\|m(z_1,\theta)\|^a/\epsilon \geq n\} \leq \sup_{\theta\in\Theta_0(P_0)}\mathbb{E}[\|m(z_1,\theta)\|^a/\epsilon] < \infty,$$

we have $\sum_{n=1}^\infty P_0\{A_n^\theta\} < \infty$ for all $\theta \in \Theta_0(P_0)$ where $A_n^\theta \equiv \{\|m(z_n,\theta)\| \geq \epsilon^{1/a}n^{1/a}\}$. By the Borel–Cantelli lemma $P_0\{A_n^\theta \ i.o.\} = 0$ and this implies $P_0\{\max_{1\leq i\leq n}\|m(z_i,\theta)\| \geq \epsilon^{1/a}n^{1/a} \ i.o.\} = 0$ for all $\theta \in \Theta_0(P_0)$ so that $\limsup_{n\to\infty}\max_{1\leq i\leq n}\|m(z_i,\theta)\|\,n^{1/a} < \epsilon^{1/a}$ with probability 1. Since ϵ is arbitrarily small it follows that $P_0\left\{\max_{1\leq i\leq n}\|m(z_i,\theta)\| = o(n^{1/a})\right\} = 1$ for all $\theta \in \Theta_0(P_0)$ and n large enough. ■

Lemma B.2. *Consider the set of binding moments, $\mathbb{E}_{P_0}[m_b(z_i,\theta)] = 0$, and define $\hat{\lambda}^*_{n,b}(\theta, \tau_{n,b})$ as,*

$$\hat{\lambda}^*_{n,b}(\theta,\tau_{n,b}) \equiv \operatorname*{argmax}_{\lambda_b\in\mathbb{R}^b} 2\sum_{i=1}^n \log[1 + \lambda_b'(m_b(z_i^*,\theta) - \varrho^\theta_{n,b})] - 2n\lambda_b'\tau_{n,b} \quad \text{(B.1)}$$

where $z_i^* \sim \bar{P}_n(\theta)$ as defined in (4.2), $0 \le \tau_{n,b} = O_p(n^{-1/2})$ and $\varrho_{n,b}^\theta$ is the sequence in (4.3). Then, under Assumption 3.1, $\|\hat{\lambda}_{n,b}^*(\theta,\tau_{n,b})\| = O_p(n^{-1/2})$ conditional on $\{\bar{P}_n : n \ge 1\}$ for almost every sample path. In addition, letting $\hat{\lambda}_{n,b}(\theta,\tau_{n,b})$ be defined as in (B.1) but setting $\varrho_{n,b}^\theta = 0$ and replacing z_i^* with $z_i \sim P_0$. Under the same assumption, $\|\hat{\lambda}_{n,b}(\theta,\tau_{n,b})\| = O_p(n^{-1/2})$.

Proof. The first order condition for $\hat{\lambda}_{n,b}^* \equiv \hat{\lambda}_{n,b}^*(\theta,\tau_{n,b})$ is,

$$0 = \frac{1}{n}\sum_{i=1}^n \frac{(m_b(z_i^*,\theta) - \varrho_{n,b}^\theta)}{1+\hat{\lambda}_{n,b}^{*\prime}(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)} - \tau_{n,b} \equiv g(\hat{\lambda}_{n,b}^*).$$

Let $\hat{\lambda}_{n,b}^* = c_n a_n$, where $c_n \ge 0$ and $\|a_n\| = 1$. Now,

$$\begin{aligned}
0 &= \|g(c_n a_n)\| \ge |a_n' g(c_n a_n)| \\
&= \frac{1}{n}\left| a_n'\left(\sum_{i=1}^n (m_b(z_i^*,\theta)-\varrho_{n,b}^\theta) \right.\right. \\
&\qquad \left.\left. - c_n \sum_{i=1}^n \frac{(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)a_n'(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)}{1+c_n a_n'(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)} - \tau_{n,b}\right)\right| \\
&\ge \frac{c_n}{n}\sum_{i=1}^n \frac{a_n'(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)'a_n}{1+c_n a_n'(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)} \\
&\qquad - \frac{1}{n}\left|\sum_{j=1}^b e_j' \sum_{i=1}^n (m_b(z_i^*,\theta)-\varrho_{n,b}^\theta) - \tau_{n,b}\right|,
\end{aligned}$$

where e_j is the unit vector in the jth coordinate direction. Now let $A_n = \max_i a_n'(m_b(z_i^*,\theta) - \varrho_{n,b}^\theta)$ and note that $A_n = o(n^{1/2})$ by similar arguments to those in Lemma B.1. Note in addition that $\tilde{\Sigma}_{n,b}^*(\theta) = n^{-1}\sum_{i=1}^n (m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)(m_b(z_i^*,\theta)-\varrho_{n,b}^\theta)' \to^p \Sigma_b(\theta)$, $\bar{P}_n(\theta)$-a.e. By Assumption 3.1, $\sigma_{(1)} > 0$, where $\sigma_{(1)}$ is the smallest eigenvalue of $\Sigma_b(\theta)$ so that $a_n'\tilde{\Sigma}_{n,b}^*(\theta)a_n \ge \sigma_{(1)} + o_p(1)$ and

$$\begin{aligned}
0 &\ge \frac{c_n a_n'\tilde{\Sigma}_{n,b}^*(\theta)a_n}{1+c_nA_n} - \frac{1}{n}\left|\sum_{j=1}^b e_j'\sum_{i=1}^n (m_b(z_i^*,\theta)-\varrho_{n,b}^\theta) - \tau_{n,b}\right| \\
&\ge \frac{c_n(\sigma_{(1)}+o_p(1))}{1+c_nA_n} - \frac{1}{n}\left|\sum_{j=1}^b e_j'\sum_{i=1}^n (m_b(z_i^*,\theta)-\varrho_{n,b}^\theta) - \tau_{n,b}\right|
\end{aligned}$$

so that,

$$\begin{aligned}
\frac{c_n(\sigma_{(1)}+o_p(1))}{1+c_nA_n} &\le \frac{1}{n}\left|\sum_{j=1}^b e_j'\sum_{i=1}^n (m_b(z_i^*,\theta) - \bar{\mu}_{n,b}(\theta))\right| \\
&\quad + \left|\sum_{j=1}^b e_j'(\bar{\mu}_{n,b}(\theta) - \varrho_{n,b}^\theta)\right| + |\tau_{n,b}| \le O_p(n^{-1/2}),
\end{aligned}$$

conditional on $\{\bar{P}_n(\theta) : n \ge 1\}$ for almost every sample path, where $\bar{\mu}_{n,b}(\theta) = \sum_{i=1}^n \bar{p}_i m_b(z_i,\theta)$. The last inequality follows because the first term obeys a triangular CLT conditional on the data and the second term satisfies $\sqrt{n}(\bar{\mu}_{n,b}(\theta) - \varrho_{n,b}^\theta) = 0$ for n large enough w.p.1 by Lemma B.4. We can then conclude that,

$$c_n = \|\hat{\lambda}_{n,b}^*(\theta,\tau_{n,b})\| \le \frac{O_P(n^{-1/2})}{\sigma_{(1)}+o_p(1)}. \tag{B.2}$$

Finally, note that setting $\varrho_{n,b}^\theta = 0$ and replacing z_i^* with $z_i \sim P_0$, we have $\|\hat{\lambda}_{n,b}(\theta,\tau_{n,b})\| = O_p(n^{-1/2})$ following the same steps as above without re-centering by $\bar{\mu}_{n,b}(\theta)$ in the last part. ■

Lemma B.3. *Define the criterion function $Q(\theta,P_0) \equiv \max_{\lambda\le 0}\mathbb{E}_{P_0}[\log(1+\lambda' m(z,\theta))]$. Then,*

$$\Theta_0(P_0) \equiv \{\theta \in \Theta : \mathbb{E}_{P_0}[m(z,\theta)] \ge 0\} = \operatorname*{argmin}_{\theta\in\Theta} Q(\theta,P_0).$$

Proof. First, note that $Q(\theta,P_0) \ge \mathbb{E}_{P_0}[\log(1)] = 0$. Next fix θ and consider $j \in \{1,\dots,q\}$ such that $\mathbb{E}_{P_0}[m_j(z,\theta)] > 0$. Then the multiplier λ_j associated with $m_j(z,\theta)$ has to be zero. To see this, note that $\lambda_j < 0$ would induce a contradiction since

$$0 \overset{(1)}{=} \mathbb{E}_{P_0}\left[\frac{m_j(z,\theta)}{1+\lambda' m(z,\theta)}\right] \overset{(2)}{\ge} \frac{\mathbb{E}_{P_0} m_j(z,\theta)}{1+\lambda'\mathbb{E}_{P_0} m(z,\theta)} > 0$$

where $\overset{(1)}{=}$ follows from the FOC when $\lambda_j < 0$, $\overset{(2)}{\ge}$ follows from Jensen's inequality and the last inequality follows from $\lambda'\mathbb{E}_{P_0} m(z,\theta) \ge 0$. Now use Jensen's inequality to conclude,

$$\begin{aligned}
Q(\theta,P_0) &= \max_{\lambda\le 0}\mathbb{E}_{P_0}[\log(1+\lambda' m(z,\theta))] \\
&\le \max_{\lambda\le 0}\log(1+\lambda_s'\mathbb{E}_{P_0}[m_s(z,\theta)] + \lambda_b'\mathbb{E}_{P_0}[m_b(z,\theta)]) = 0,
\end{aligned}$$

since $\lambda_s = 0$ by the previous argument and $\mathbb{E}_{P_0}[m_b(z,\theta)] = 0$. Therefore, $Q(\theta,P_0) = 0$ for all $\theta \in \Theta_0(P_0)$. To complete the argument note that $Q(\theta,P_0) > 0$ if θ is such that $\mathbb{E}_{P_0}[m_j(z,\theta)] < 0$ for some $j \in \{1,\dots,q\}$. To see this note that $\mathbb{E}_{P_0}[\log(1+\lambda' m(z,\theta))]_{\lambda=0} = 0$ and

$$\left.\frac{\partial \mathbb{E}_{P_0}[\log(1+\lambda' m(z,\theta))]}{\partial\lambda_j}\right|_{\lambda=0} < 0.$$

Let $\tilde{\lambda}$ be such that $\tilde{\lambda}_j = -\epsilon$ for some small $\epsilon > 0$ and $\tilde{\lambda}_k = 0$ for $k \ne j$. By continuity of the objective function in λ we have that $Q(\theta,P_0) \ge \mathbb{E}_{P_0}[\log(1+\tilde{\lambda}' m(z,\theta))] > \mathbb{E}_{P_0}[\log(1)] = 0$. ■

Lemma B.4. *Let $\bar{\mu}_n(\theta) = \sum_{i=1}^n \bar{p}_i m(z_i,\theta)$ denote the Modified EL mean and ϱ_n^θ be the sequence in (4.3). Under Assumption 3.1 the following two statements hold.*

(a) *$\bar{\mu}_{n,s}(\theta) > \varrho_{n,s}^\theta$ for n large enough with probability one.*
(b) *$\bar{\mu}_{n,b}(\theta) = \varrho_{n,b}^\theta$ for n large enough with probability one.*

Proof. For part (a) recall that $\bar{\mu}_{s,j,n}(\theta)$ comes from problem (4.2) and note that from

$$\begin{aligned}
0 &\le \max_{\lambda\le 0} 2\sum_{i=1}^n \log[1+\lambda'(m(z_i,\theta)-\varrho_n^\theta)] \\
&\le 2\log\left[1+\hat{\lambda}_n'\sum_{i=1}^n (m(z_i,\theta)-\varrho_n^\theta)\right],
\end{aligned}$$

it follows that $\hat{\lambda}_n'\sum_{i=1}^n (m(z_i,\theta)-\varrho_n^\theta) \ge 0$. Now, from the FOC of $\lambda_{s,j}$,

$$\frac{1}{n}\sum_{i=1}^n\left(\frac{m_{s,j}(z_i,\theta)-\varrho_{n,s,j}^\theta}{1+\hat{\lambda}_n'(m(z_i,\theta)-\varrho_n^\theta)}\right) \overset{(1)}{\ge} \frac{\bar{m}_{n,s,j}(\theta)-\varrho_{n,s,j}^\theta}{1+\hat{\lambda}_n'(\bar{m}_n(\theta)-\varrho_n^\theta)} \tag{B.3}$$

where $\overset{(1)}{\ge}$ follows from Jensen's inequality. Since $\hat{\lambda}_n'(\bar{m}_n(\theta)-\varrho_n^\theta) \ge 0$ for all n and $\bar{m}_{n,s,j}(\theta) - \varrho_{n,s,j}^\theta > 0$ for n large enough w.p.1 by (4.4) we have,

$$P\left(\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left(\frac{m_{s,j}(z_i,\theta)-\varrho_{n,s,j}^\theta}{1+\hat{\lambda}_n'(m(z_i,\theta)-\varrho_n^\theta)}\right) > 0\right) = 1$$

meaning that $\bar{\mu}_{n,s,j}(\theta) \equiv \frac{1}{n}\sum_{i=1}^n\left(\frac{m_j(z_i,\theta)}{1+\hat{\lambda}_n'(m(z_i,\theta)-\varrho_n^\theta)}\right) > \varrho_{n,s,j}^\theta$ w.p.1 for n sufficiently large.

For part (b) set $\lambda = 0$ and consider the FOC for $\lambda_{b,j}$,

$$\left.\frac{1}{n}\sum_{i=1}^n\left(\frac{m_j(z_i,\theta)-\varrho_{n,b,j}^\theta}{1+\lambda'(m(z_i,\theta)-\varrho_n^\theta)}\right)\right|_{\lambda=0} = \bar{m}_{n,b,j}(\theta) - \varrho_{n,b,j}^\theta. \tag{B.4}$$

By (4.4) we know $\bar{m}_{n,b,j}(\theta) - \varrho^{\theta}_{n,b,j} \leq 0$ for n large w.p.1. If $\bar{m}_{n,b,j}(\theta) - \varrho^{\theta}_{n,b,j} = 0$, we are done since $\bar{\mu}_{n,b,j}(\theta) = \bar{m}_{n,b,j}(\theta) = \varrho^{\theta}_{n,b,j}$ and $\lambda_{b,j} = 0$ is optimal. If $\bar{m}_{n,b,j}(\theta) - \varrho^{\theta}_{n,b,j} < 0$, then the optimal value of $\lambda_{b,j}$ has to decrease (so it will be negative) by continuity of the objective function in $\lambda_{b,j}$. Since the optimal solution has to satisfy $\lambda_{b,j}(\bar{\mu}_{n,b,j}(\theta) - \varrho^{\theta}_{n,b,j}) = 0$, it follows that $\bar{\mu}_{n,b,j}(\theta) = \varrho^{\theta}_{n,b,j}$. ■

Appendix C. Auxiliary notation

This appendix uses notation from Andrews and Soares (forthcoming) (AS in what follows) and Andrews and Guggenberger (2009) (AG) to write the Modified EL Bootstrap as a GMS procedure. Consider the space $\mathcal{F}_0$ of null parameters (θ, P) from Definition 3.1. This space can be alternatively parametrized using the vector $\gamma = (\gamma_1, \gamma_2, \gamma_3) \in \Gamma$, with elements,

$$\gamma_{1,j} = \sigma^{-1}_{P,j}(\theta)\mathbb{E}_P m_j(z, \theta) \quad \text{for } j = 1, \ldots, q, \qquad \gamma_2 = (\theta, vech_*(\Omega(\theta, P))), \qquad \gamma_3 = P, \tag{C.1}$$

where $\sigma^2_{P,j}(\theta) = \lim_{n\to\infty} \mathrm{var}_P(n^{1/2}\bar{m}_n(\theta))$, $\Omega(\theta, P) = \lim_{n\to\infty} \mathrm{Corr}_P(n^{1/2}\bar{m}_n(\theta))$, $\bar{m}_n(\theta) = n^{-1}\sum_{i=1}^n m(z_i, \theta)$, and $vech_*$ gives the elements of a matrix below the diagonal. The parameter space for γ is $\Gamma = \{\gamma = (\gamma_1, \gamma_2, \gamma_3) : \text{for some } (\theta, P) \in \mathcal{F}_0\}$. Given that the data is i.i.d., the space Γ satisfies the conditions (i) to (x) in AS (see AS, page 35). Finally, define the QLR test statistic as

$$T_n(\theta) = S(n^{1/2}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)) = S(n^{1/2}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta), \hat{\Omega}_n(\theta))$$

where $S(m, \Sigma) = \inf_{t\in R^q_{+,\infty}} (m-t)'\Sigma^{-1}(m-t)$, $\hat{\Sigma}_n(\theta) = n^{-1}\sum_{i=1}^n (m(z_i, \theta) - \bar{m}_n(\theta))(m(z_i, \theta) - \bar{m}_n(\theta))'$, $\hat{D}_n(\theta) = \mathrm{diag}(\hat{\Sigma}_n(\theta))$, $\hat{\Omega}_n(\theta) = \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta)$ and $\hat{\sigma}_{n,j}(\theta) = [\hat{D}_n(\theta)]_j$.

In addition, let $\bar{\varrho}^{\theta}_n = n^{-1/2}\kappa_n \times \bar{D}_n^{1/2}(\theta)$, where $\kappa_n = \sqrt{2\log\log n}$ and $\bar{D}_n(\theta) = \mathrm{diag}(\bar{\Sigma}_n(\theta))$, $\bar{\Sigma}_n(\theta) = \sum_{i=1}^n \bar{p}_i(\theta)(m(z_i, \theta) - \bar{\mu}_n(\theta))(m(z_i, \theta) - \bar{\mu}_n(\theta))'$, and $\bar{\Sigma}^*_n(\theta) = n^{-1}\sum_{i=1}^n (m(z^*_i, \theta) - \bar{\mu}_n(\theta))(m(z^*_i, \theta) - \bar{\mu}_n(\theta))'$. As before, $\bar{P}_n(\theta)$ and $\bar{\mu}_n(\theta)$ denote Modified EL probabilities and mean.

Under Assumption GEL in AG, for any sequence $\{\gamma_{n,h} : n \geq 1\}$ that satisfies $n^{1/2}\gamma_{n,h,1} \to h_1$, $(\theta_{n,h}, vech_*(\Omega(\theta_{n,h}, F_{n,h}))) \to h_2$ for some $(h_1, h_2) \in R^q_{+,\infty} \times R^p_{[\pm\infty]}$, $p = k + q(q-1)/2$,

$$\mathcal{ELR}_n(\theta_{n,h}) - T_n(\theta_{n,h}) = o_p(1).$$

This is an extension of the result in Eq. (A.7) and so $\mathcal{ELR}_n(\theta)$ satisfies Assumption B0 in AG and I can focus on $T_n(\theta)$ in the following analysis.

AS define a function $\varphi(\cdot)$ that replaces the vector h_1 in the asymptotic distribution. To write the Modified EL Bootstrap using such a notation, define $\bar{\varphi}(\xi_n(\theta))$ as the k-vector with elements

$$\bar{\varphi}_j(\xi_n(\theta)) = \begin{cases} 0 & \text{if } \xi_{n,j}(\theta) \leq 1 \text{ and } j = 1, \ldots, q \\ e_n(\xi_{n,j}(\theta)) & \text{if } \xi_{n,j}(\theta) > 1 \text{ and } j = 1, \ldots, q, \end{cases} \tag{C.2}$$

where $e_n(\xi_{n,j}(\theta))$ is a sequence that satisfies $e_n(\xi_{n,j}(\theta)) \geq \kappa_n(\xi_{n,j}(\theta) - 1)$ for all n and $\xi_{n,j}(\theta) = \kappa_n^{-1}n^{1/2}\bar{\sigma}^{-1}_{n,j}(\theta)\bar{m}_{n,j}(\theta)$. The function $\bar{\varphi}(\cdot)$ satisfies Assumptions GMS1, GMS3 and GMS6 in AS. The Modified EL Bootstrap replaces h_1 with the function $n^{1/2}\bar{D}_n^{-1/2}(\theta)(\bar{\mu}_n(\theta) - \bar{\varrho}^{\theta}_n)$, which is a particular case of $\bar{\varphi}(\xi)$. Indeed, similar arguments to those in Lemma B.4 show that when $\xi_{n,j}(\theta) \leq 1$, $(\bar{\mu}_{n,j}(\theta) - \bar{\varrho}^{\theta}_{n,j}) = 0$, while $n^{1/2}\bar{\sigma}^{-1}_{n,j}(\theta)(\bar{\mu}_{n,j}(\theta) - \bar{\varrho}^{\theta}_{n,j}) \geq \kappa_n(\xi_{n,j}(\theta) - 1) \to \infty$ when $\xi_{n,j}(\theta) > 1$.

In order to extend the uniformity results to the Modified EL Bootstrap I need to impose additional conditions on the space Γ. Let $\varsigma^*_n(\theta)$ be the k-vector whose jth element is

$$\varsigma^*_{n,j}(\theta) = n^{1/2}(\bar{m}^*_{n,j}(\theta) - \bar{\mu}_{n,j}(\theta))/\bar{\sigma}_{n,j}(\theta)$$

where $\bar{m}^*_{n,j}(\theta) = n^{-1}\sum_{i=1}^n m_j(z^*_i, \theta)$, $\bar{\sigma}^2_{n,j}(\theta) = [\bar{D}_n(\theta)]_{jj}$, and $z^*_i \sim \bar{P}_n(\theta)$ denotes the bootstrap data. Then, under any sequence $\{\gamma_{n,h} = (\gamma_{n,h,1}, (\theta_{n,h}, vech_*(\Omega(\theta_{n,h}, P_{n,h}))), P_{n,h}) : n \geq 1\}$ in Γ that satisfies $n^{1/2}\gamma_{n,h,1} \to h_1$ and $(\theta_{n,h}, vech_*(\Omega(\theta_{n,h}, P_{n,h}))) \to h_2$, the following conditions are needed.

- (vi′) $\varsigma^*_n(\theta_{n,h}) \rightsquigarrow Z_{h_{2,2}} \sim N(0_q, \Omega_{h_{2,2}})$, $\bar{P}_n(\theta_{n,h})$-a.e.
- (vii′) $\bar{\sigma}_{n,j}(\theta_{n,h})/\sigma_{P_{n,h},j}(\theta_{n,h}) \to_p 1$, $\bar{P}_n(\theta_{n,h})$-a.e.
- (viii′) $\bar{D}_n^{-1/2}(\theta_{n,h})\bar{\Sigma}^*_n(\theta_{n,h})\bar{D}_n^{-1/2}(\theta_{n,h}) = \bar{\Omega}^*_n(\theta_{n,h}) \to_p \Omega_{h_{2,2}}$, $\bar{P}_n(\theta_{n,h})$-a.e.
- (ix′) Conditions (vi′)–(viii′) hold for all subsequences $\{w_n\}$ in place of $\{n\}$.

The conditions in Definition 3.1 imply condition (vi′) above by the Lyapunov CLT for row-wise i.i.d. random variables with mean zero and variance one. Conditions (vii′) and (viii′) follow by standard arguments using the WLLN for row-wise i.i.d. random variables. Condition (iv′) holds by the same type of arguments.

Finally, given a vector $\pi = (\pi_1, \pi_2)$ define $\pi^* = (\pi^*_1, \pi_2)$ as $\pi^*_{1,j} = \infty$ if $\pi_{1,j} > 0$ and $\pi^*_{1,j} = 0$ if $\pi_{1,j} = 0$ for $j = 1, \ldots, q$. Also, $c_{\pi^*}(1-\alpha)$ denotes the $1-\alpha$ quantile of $S(\Omega^{1/2}_{\pi_{2,2}}Z^* + \pi^*_1, \Omega_{\pi_{2,2}})$ where $Z^* \sim N(0_q, I_q)$ and by definition if $\pi^*_{1,j} = \infty$ then the jth element of $\Omega^{1/2}_{\pi_{2,2}}Z^* + \pi^*_1$ equals ∞.

References

Andrews, D.W.K., 2000. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. Econometrica 68 (2), 399–405.

Andrews, D.W.K., Berry, S.T., Jia, P., 2004. Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store locations. Yale University. Manuscript.

Andrews, D.W.K., Guggenberger, P., 2009. Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. Econometric Theory 25, 669–709.

Andrews, D.W.K., Guggenberger, P., 2010. Asymptotic size and a problem with subsampling and with the m out of n bootstrap. Econometric Theory (forthcoming).

Andrews, D.W.K., Jia, P., 2008. Inference for parameters defined by moment inequalities: A recommended moment selection procedure. Yale University. Manuscript.

Andrews, D.W.K., Soares, G., 2009. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica (forthcoming).

Bahadur, R., 1960. On the asymptotic efficiency of test and estimators. Sankhya 22, 229–252.

Beresteanu, A., Molinari, F., 2008. Asymptotic properties for a class of partially identified models. Econometrica 76 (4), 763–814.

Brown, B.W., Newey, W., 2002. Generalized method of moments, efficient bootstrapping, and improved inference. Journal of Business and Economics Statistics 20 (4), 507–517.

Bugni, F.A., 2009. Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the elements of the identified set. Duke University. Manuscript.

Bugni, F.A., 2009. Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. Econometrica (forthcoming).

Canay, I.A., Otsu, T., 2009. Hodges–Lehmann optimality for testing moment condition models. Northwestern University. Manuscript.

Chamberlain, G., 1987. Asymptotic efficiency in estimation with conditional moment restrictions. Journal of Econometrics 34 (1), 305–334.

Chernoff, H., 1952. A measure of asymptotic efficiency for test of a hypothesis based on the sum of observations. The Annals of Mathematical Statistics 23, 497–507.

Chernozhukov, V., Hong, H., Tamer, E., 2004. Inference for identified parameter sets in econometric models. M.I.T., Manuscript.

Chernozhukov, V., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models. Econometrica 75 (5), 1243–1284.

Ciliberto, F., Tamer, E., 2009. Market structure and multiple equilibria in airline industry. Econometrica (forthcoming).

Dembo, A., Zeitouni, O., 1998. Large Deviations Techniques and Applications, second edition. Springer, New York.

DiCiccio, J.D., Hall, P., Romano, J., 1991. Empirical likelihood is Bartlett-correctable. Annals of Statistics 19, 1053–1061.

Dupuis, P., Ellis, R.S., 1997. A Weak Convergence Approach to the Theory of Large Deviations. In: Wiley Series in Probability and Statistics. New York.

Galichon, A., Henry, M., 2006a. Dilation bootstrap: A natural approach to inference in incomplete models. Harvard University. Manuscript.

Galichon, A., Henry, M., 2006b. Inference in incomplete models. Harvard University. Manuscript.

Gourieroux, C., Holly, A., Monfort, A., 1982. Likelihood ratio test, Wald test, and Khun–Tucker test in linear models with inequality constraints on the regression parameters. Econometrica 50 (1), 63–80.
Guggenberger, P., Hahn, J., Kim, K., 2008. Specification and testing under moment inequalities. Economics Letters 99 (2), 375–378.
Ho, K., 2009. Insurer-provider networks in the medical care market. American Economic Review 99 (1), 393–430.
Hoeffding, W., 1965. Asymptotically optimal test for multinomial disrtibutions (with discussion). Annals of Mathematical Statistics 36, 369–408.
Horowitz, J.L., Manski, C.F., 1998. Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations. Journal of Econometrics 84 (1), 37–58.
Horowitz, J.L., Manski, C.F., 2000. Nonparametric analysis of randomized experiments with missing covariate and outcome data. Journal of the American Statistical Association 95 (449), 77–84.
Imbens, G., Manski, C., 2004. Confidence intervals for partially identified parameters. Econometrica 72, 1845–1857.
Imbens, G., Spady, R.H., Johnson, P., 1998. Information theoretic approaches to inference in moment condition models. Econometrica 66 (2), 333–357.
Ishii, J., 2005. Interconnection pricing, compatibility, and investment in network industries: An empirical study of atm surcharging in the retail banking industry. Harvard University. Manuscript.
Kitamura, Y., 2001. Asymptotic optimality of empirical likelihood for testing moment restrictions. Econometrica 69 (6), 1661–1672.
Kitamura, Y., 2006. Empirical likelihood methods in econometrics: Theory and practice. Cowles Foundation Discussion Paper 1569.
Kitamura, Y., Otsu, T., 2005. Minimax estimation and testing for moment condition models via large deviations. Yale University. Manuscript.
Kitamura, Y., Santos, A., Shaikh, A., 2009. On the asymptotic optimality of empirical likelihood for testing moment restrictions. University of Chicago. Manuscript.
Kudo, A., 1963. A multivariate analog of a one-sided test. Biometrica 59 (3), 403–418.
Manski, C., 1989. Anatomy of the selection problem. Journal of Human Resources 24 (3), 343–360.
Manski, C., 2003. Partial Identification of Probability Distributions. Springer-Verlag, New York.
Manski, C., Tamer, E., 2002. Inference on regressions with interval data on a regressor or outcome. Econometrica 70, 519–546.
Newey, W.K., Smith, R.J., 2004. Higher order properties of gmm and generalized empirical likelihood estimators. Econometrica 72 (1), 219–255.
Owen, A., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.
Owen, A., 1990. Empirical likelihood for confidence regions. Annals of Statistics 18 (1), 90–120.
Owen, A., 1991. Empirical likelihood for linear models. Annals of Statistics 19 (4), 1725–1747.
Owen, A., 2001. Empirical Likelihood. Chapman and Hall, CRC.
Pakes, A., Porter, J., Ho, K., Ishii, J., 2005. Moment inequalities and their applications. Harvard University. Manuscript.
Qin, B.J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Annals of Statistics 22 (1), 300–325.
Romano, J.P., Shaikh, A.M., 2009. Inference for the identified set in partially identified econometric models. Econometrica (forthcoming).
Romano, J.P., Shaikh, A.M., 2008. Inference for identifiable parameters in partially identified econometric models. Journal of Statistical Planning and Inference 138 (9), 2786–2807.
Rosen, A., 2008. Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. Journal of Econometrics 146 (1), 107–117.
Santos, A., 2006. Inference in nonparametric instrumental variables with partial identification. Stanford University. Manuscript.
Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. John Wiley, New York.
Silvapulle, M.J., Sen, P.K., 2004. Constrained Statistical Inference: Inequality, Order and Shape Restrictions. Wiley-Interscience, New York.
Soares, G., 2006. Inference for partially identified models with inequality moment constraints. Yale University. Manuscript.
Tamer, E., 2003. Incomplete simultaneous discrete response model with multiple equilibria. The Review of Economic Studies 70, 147–165.
van der Vaart, A.W., 1998. Asymptotic Statistics. Cambridge University Press, Cambridge.
Wolak, F., 1987. An exact test for multiple inequality and equality constraints in the linear regression model. Journal of the American Statistical Association 82 (399), 782–793.
Wolak, F., 1991. The local nature of hypothesis tests involving inequality constraints in nonlinear models. Econometrica 59 (4), 981–995.
Zeitouni, O., Gutman, M., 1991. On universal hypothesis testing via large deviations. IEEE Transactions on Information Theory 37 (2), 285–290.