



THE UNIVERSITY OF
SYDNEY

Economics Working Paper Series

2019 - 07

Jackknife Empirical Likelihood for Inequality
Constraints on Regular Functionals

Ruxin Chen and Rami V. Tabri

April 2019

Jackknife Empirical Likelihood for Inequality Constraints on Regular Functionals

Ruxin Chen and Rami V. Tabri*

April 8, 2019

Abstract

Empirical likelihood is effective in many different practical situations involving moment equality and/or inequality restrictions. However, in applications with nonlinear functionals of the underlying distribution, it becomes computationally more difficult to implement. We propose the use of jackknife empirical likelihood (Jing et al., 2009) to circumvent the computational difficulties with nonlinear inequality constraints and establish the chi-bar-square distribution as the limiting null distribution of the resulting empirical likelihood-ratio statistic, where a finite number of inequalities on functionals that are *regular*, in the sense of Hoeffding (1948), defines the null hypothesis. The class of regular functionals includes many nonlinear functionals that arise in practice and has moments as a special case. To overcome the implementation challenges with this non-pivotal asymptotic null distribution, we propose an empirical likelihood bootstrap procedure that is valid with uniformity. Finally, we investigate the finite-sample properties of the bootstrap procedure using Monte Carlo simulations and find that the results are promising.

JEL Classification: C12;C14;C21

Keywords: Jackknife Empirical Likelihood; Bootstrap Test; Inequality Restrictions; U-statistics.

1 Introduction

The method of empirical likelihood, introduced by Owen (1988, 1990), is a nonparametric maximum likelihood procedure for conducting inference on population parameters which has many desirable statistical properties. In the context of point estimation in models with equality moment restrictions, it is semiparametrically efficient (Chamberlain, 1987) and exhibits higher-order improvements (Newey and Smith, 2004). Regarding statistical testing, the empirical likelihood-ratio (ELR) test admits Bartlett-correctable confidence intervals (DiCiccio et al., 1991), and is uniformly most powerful in a Generalized Neyman-Pearson sense for testing moment equality restrictions (Kitamura, 2001). Furthermore, the bootstrap version of the ELR test in such models, introduced by Brown and Newey (2002), exhibits asymptotic refinements in finite-samples (Beran, 1988). More recently, Canay (2010) extends the Generalized Neyman-Pearson optimality of the

*Corresponding author. School of Economics, The University of Sydney, Sydney, New South Wales 2006, Australia, Tel: +61 2 9351 3092, Fax: +61 2 9351 4341, Email: rami.tabri@sydney.edu.au.

ELR test to the setup involving moment inequality restrictions, establishes the asymptotic null distribution of the ELR statistic is chi-bar-square, and proposes a valid bootstrap ELR test.

On the computational side, empirical likelihood is straightforward to compute in moment equality and/or inequality models, because moments are linear functionals of the underlying distribution. However, it becomes significantly computationally expensive in other applications involving nonlinear statistical functionals of the underlying distribution, for example in conjunction with U-statistics, because the solution is defined by a number (dependent on the sample size) of nonlinear equality/inequality restrictions. The situation also worsens in large sample sizes and if one wishes to apply the bootstrap versions of the ELR tests mentioned above.

To illustrate the computational difficulties of empirical likelihood for nonlinear functionals with inequality constraints, we take one inequality restriction on a single U-statistic as an example. Given a random sample of size $n \geq 2$, X_1, \dots, X_n , from an unknown distribution, a U-statistic of degree $r \geq 2$ with a symmetric kernel function h is defined to be

$$\binom{n}{r}^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_r \leq n} h(X_{i_1}, \dots, X_{i_r}), \quad (1.1)$$

where $\theta = E_P[h(X_1, \dots, X_r)]$ is the parameter of interest. Denote by $F_{\mathbf{p}}(x) = \sum_{i=1}^n p_i 1[X_i \leq x]$, where $1[\cdot]$ is the indicator function. The naive application of empirical likelihood for testing the restriction $\theta \geq 0$ yields the ELR statistic

$$\min_{p_1, \dots, p_n} \left\{ 2 \sum_{i=1}^n \log \left(\frac{1/n}{p_i} \right) \mid p_i > 0, \sum_{i=1}^n p_i = 1, \hat{\theta}(F_{\mathbf{p}}) \geq 0 \right\}, \quad (1.2)$$

where

$$\hat{\theta}(F_{\mathbf{p}}) = \binom{n}{r}^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_r \leq n} n^r p_{i_1} \cdots p_{i_r} h(X_{i_1}, \dots, X_{i_r}).$$

In view of the functional form of $\hat{\theta}(F_{\mathbf{p}})$, computing the solution of (1.2) typically involves solving $(n+1)$ nonlinear equations, which becomes computationally more difficult as n increases.

One approach that can attenuate the computational burden due to the nonlinearity of the functionals is the *sequential linearization method* of Wood et al. (1996); it replaces the nonlinear functionals with their linear approximations at the null value. However, it is an ad hoc method and it remains to establish the asymptotic null distribution of the ELR statistic, even for U-statistics under equality and/or inequality constraints. The more promising approach in the case of nonlinear equality constraints is the method of *jackknife empirical likelihood* (JEL) of Jing et al. (2009). They propose to convert the nonlinear constraints into constraints on sample means based on jackknife pseudo-values, and establish a Wilks' theorem for one and two-sample U-statistics alike. Furthermore, their method is extremely simple to use in practice. On the other hand, the empirical likelihood literature has yet to consider the setup involving nonlinear inequality restrictions. It is reasonable to conjecture that, like in the setup with moment inequality constraints (Canay, 2010), the ELR statistic has a well-defined asymptotic null distribution that depends on the underlying distribution through

the number of binding constraints, rendering it as non-pivotal. Consequently, it may seem appealing to consider bootstrap calibration of the ELR statistic, as in Owen (1988), to develop a critical value. However, this approach is not valid because it uses the canonical bootstrap which is inconsistent in models defined by inequality constraints when the parameter is on the boundary of the parameter space (e.g., Andrews, 2000).

Our paper addresses this gap in the literature and validates our conjecture. The nonlinear constraints we consider are finite linear combinations of functionals that are *regular* in the sense of Hoeffding (1948). A regular functional can be represented as the expected value of a symmetric kernel function, and hence, gives rise to a U-statistic as in (1.1). This structure may be restrictive; however, it is a common specification which covers many important applications that arise in practice, such as order restrictions on measures of dispersion and association. In this paper, we extend the JEL method of Jing et al. (2009) to the setup involving multiple inequality constraints on regular functionals, and show

- (a) the asymptotic null distribution of the ELR statistic is chi-bar-square.
- (b) the empirical likelihood bootstrap of Canay (2010), based on the JEL modification, is valid.

As with JEL for equality constraints, the proposed JEL bootstrap procedure's most appealing feature is also its simplicity. It is simple enough to preserve the computational tricks that make empirical likelihood straightforward to use in practice. For other nonlinear functionals, the above results have to be checked on a case-by-case basis. An prominent example of a nonlinear functional that isn't regular is a finite-dimensional vector of Lorenz curve ordinates as in Beach and Davidson (1983) and Dardanoni and Forcina (1999).

The literature related to this paper includes numerous papers on inference with unconditional moment inequality models, where there is only a finite number of such inequalities and the parameter of interest is finite-dimensional; for example, Andrews and Soares (2010), Rosen (2008), Andrews and Guggenberger (2009, 2010), and Canay (2010), among others. By contrast, the parameter of interest in this paper is infinite-dimensional and the constraints are on regular functionals of the underlying distribution, which includes moment inequality restrictions as a special case.

The paper proceeds as follows. Section 2 introduces the testing problem of interest, examples of regular nonlinear functionals, and the JEL-based ELR statistic. Section 3 presents the main results of this paper described above. Section 4 investigates the performance of the proposed bootstrap ELR test via Monte Carlo simulations using a simple example, and Section 5 concludes. Proofs are deferred to the Appendix.

2 Setup and Examples

Let X_1, \dots, X_n be an i.i.d sample generated from an unknown distribution P with support on $\mathcal{X} \subset \mathbb{R}^d$. More generally, let $\mathbf{x} = [x_1, \dots, x_n]$ and $h(\mathbf{x})$ be a $k \times 1$ known function permutation symmetric components, $h_j(X_1, \dots, X_{r(j)})$, where $r(j) \leq n$ for each $j = 1, \dots, k$. The permutation symmetry of each h_j is only to simplify the exposition, as any given h_j could always be replaced by a symmetric one. The "parameter" is the vector $\beta(P) = \mathbf{A} E_P [h(X_1, \dots, X_n)]$, where \mathbf{A} is a given $q \times k$ matrix. The testing

problem of interest takes the following form:

$$H_0 : \beta(P) \geq \mathbf{0}_q \quad \text{Vs.} \quad H_1 : \text{not } H_0, \quad (2.1)$$

where $\mathbf{0}_q \in \mathbb{R}^q$ is the zero vector, and the inequality in H_0 is understood to be elementwise–yielding q inequality restrictions.

Next, we present examples that illustrate the functional $\beta(P)$ in the case of $d = 2$, so that $X_i = [X_{1i}, X_{2i}]$ for $i = 1, \dots, n$. We focus on this case for ease of exposition as it is straightforward to generalize the examples in their respective setups when $d > 2$. Throughout these examples, denote the cumulative distribution function of P by F , and the copula function yielding F by C .

Example 1. (Order Restriction on Variances) Interest centers on testing $\text{VAR}_P[X_{1i}] - \text{VAR}_P[X_{2i}] \geq 0$, which is a nonlinear inequality in P . This inequality in terms of $\beta(P)$ has $k = 2$, $\mathbf{A} = [1, -1]$, $r(1) = r(2) = 2$, and $h(\mathbf{x}) = \frac{1}{2}[(x_{11} - x_{12})^2, (x_{21} - x_{22})^2]^T$.

Example 2. (Order Restriction on Gini’s Mean Differences) Gini’s Mean Difference is a measure of dispersion and of income inequality. Denote the marginal CDFs of F by F_1 and F_2 . Interest may center on testing the inequality $\text{COV}_P(X_{1i}, F_1(X_{1i})) - \text{COV}_P(X_{2i}, F_2(X_{2i})) \geq 0$. This inequality in terms of $\beta(P)$ has $k = 2$, $\mathbf{A} = [1, -1]$, $r(1) = r(2) = 2$ and $h(\mathbf{x}) = [|x_{11} - x_{12}|, |x_{21} - x_{22}|]^T$.

Example 3. (Kendall’s τ Statistic) Kendall’s τ is a measure of nonlinear dependence between two populations that is invariant to any monotonically increasing transformations of X_{1i} and X_{2i} . Thus, this measure is defined as $\tau = 4 \int_{[0,1]^2} C(u) dC(u) - 1$, where $u = [u_1, u_2]$. As $\tau \in [-1, 1]$, interest may center on whether $\tau \geq 0$. In this case, $k = 1$, and therefore, the matrix \mathbf{A} does not play a role. The kernel function corresponding to $\tau + 1$ is $h(\mathbf{x}) = 2 \mathbb{1}[(x_{22} - x_{21})(x_{12} - x_{11}) > 0]$, where $r = 2$.

Example 4. (Spearman’s ρ Statistic) Like Kendall’s τ , Spearman’s ρ statistic is also a measure of nonlinear dependence between two random variables that is invariant to any monotonically increasing transformations of X_{1i} and X_{2i} . This measure is defined as $\rho = 12 \int_{[0,1]^2} u dC(u) - 3$. As $\rho \in [-1, 1]$, interest may center on whether the relationship between the two variables is weakly monotonically increasing, i.e., $\rho \geq 0$. In this case, $k = 1$, and therefore, the matrix \mathbf{A} does not play a role. The kernel function corresponding to ρ is $h(\mathbf{x}) = 12 \mathbb{1}[x_{21} < x_{23}, x_{11} < x_{12}] - 3$, where $r = 3$. The symmetrized version of h has 6 terms.

This paper proposes the ELR statistic and a bootstrap procedure based upon it as the basis for testing H_0 . The construction of the ELR statistic follows the JEL procedure of Jing et al. (2009), but uses inequality constraints defined by H_0 in (2.1). Denote by U_n the $k \times 1$ vector whose components are the U-statistics $U_{j,n}$ defined as

$$U_{j,n} = \binom{n}{r(j)}^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_{r(j)} \leq n} h_j(X_{i_1}, \dots, X_{i_{r(j)}}), \quad j = 1, \dots, k. \quad (2.2)$$

Define $G_n = \mathbf{A} U_n$, and by \hat{V}_i the i th jackknife pseudo-value $\hat{V}_i = n G_n - (n-1)G_{n-1}^{-i}$ for $i = 1, 2, \dots, n$, where G_{n-1}^{-i} is the statistic G_n except that it is computed on the sample of $n-1$ variables formed from the

original dataset by deleting the i th observation. It is easy to check that

$$G_n = n^{-1} \sum_{i=1}^n \hat{V}_i \quad \text{and} \quad E_P[\hat{V}_i] = \beta(P),$$

hold, where the expectation is understood to be with respect to the n -fold distribution induced by random sampling from P . For notational simplicity, we suppress the sample size in the expected value notation throughout this paper.

The ELR statistic is defined as

$$ELR_n = \min_{p_1, \dots, p_n} \left\{ 2 \sum_{i=1}^n \log \left(\frac{1/n}{p_i} \right) \mid p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{V}_i \geq \mathbf{0}_q \right\}. \quad (2.3)$$

A test of H_0 in (2.1) that uses the ELR statistic rejects H_0 for large values of ELR_n , and therefore, has the form $1[ELR_n > c]$, where c is a critical value. The next section presents the asymptotic null distribution of ELR_n and an empirical likelihood bootstrap procedure for the construction of the critical value, c .

3 Results

Before presenting the results, we introduce some additional notation. Represent the pseudo-value \hat{V}_i with the following partition: $\hat{V}_i = [\hat{V}_{i,b}, \hat{V}_{i,s}]^\top$ where $\hat{V}_{i,b}$ is a $b(P) \times 1$ vector with zero means, $E_P[\hat{V}_{i,b}] = \mathbf{0}_{b(P)}$, and $\hat{V}_{i,s}$ is a $s(P) \times 1$ vector with positive means, $E_P[\hat{V}_{i,s}] > \mathbf{0}_{s(P)}$. Furthermore, define the covariance matrix $\Sigma(P) = \mathbf{A} E_P[\Psi(X_1)\Psi(X_1)^\top] \mathbf{A}^\top$, where the j th component of $\Psi(x)$ is

$$\Psi_j(x) = r(j) \left\{ E_P[h_j(X_1, \dots, X_{r(j)-1}, x)] - E_P[h_j(X_1, \dots, X_{r(j)})] \right\} \quad \forall j = 1, \dots, q. \quad (3.1)$$

Denote by $\Sigma_{b(P)}(P)$ the submatrix of $\Sigma(P)$ obtained by deleting the rows and columns corresponding to $E_P[\hat{V}_{i,s}] > \mathbf{0}_{s(P)}$. Finally, denote by \mathcal{M} the space of probability measures on the Borel σ -field $(\chi, \mathcal{B}(\chi))$. The asymptotic behavior of ELR_n will follow from the next standard assumption.

Assumption 1. (i) $\{X_i : i \leq n\}$ is a random sample from some $P \in \mathcal{M}$, where $n \geq \max_{1 \leq j \leq k} r(j)$. (ii) $E_P[h_j^2(X_1, \dots, X_{r(j)})] < +\infty$ for all $j = 1, \dots, k$. (iii) $\Sigma_{jj}(P) > 0$ for all $j = 1, \dots, q$. (iv) The covariance matrix $\Sigma_{b(P)}(P)$ is positive definite.

In general, the pseudo-values are dependent random variables and are asymptotically independent under mild conditions (Shi, 1984). For each $P \in \mathcal{M}$ that satisfies Assumption 1, Theorem 7.1 of Hoeffding (1948) implies that $\sqrt{n}(G_n - \beta(P)) \xrightarrow{d} N(\mathbf{0}_q, \Sigma(P))$, holds. In particular, under the same assumption, $\sqrt{n}G_{n,b(P)} \xrightarrow{d} N(\mathbf{0}_{b(P)}, \Sigma_{b(P)}(P))$, holds for each $P \in \mathcal{M}$.

The following theorem describes the pointwise asymptotic behavior of the test statistic, ELR_n , under H_0 .

Theorem 1. Suppose that $P \in \mathcal{M}$ satisfies H_0 in (2.1) and Assumption 1, and let ELR_n be given by (2.3). Furthermore, let $Z \sim N(\mathbf{0}_{b(P)}, \Sigma_{b(P)}(P))$. Then $ELR_n \xrightarrow{d} \min_{t \in \mathbb{R}_+^{b(P)}} (Z - t)^\top \Sigma_{b(P)}^{-1}(P) (Z - t)$.

Proof. See Appendix A.1. □

The statistic $\min_{t \in \mathbb{R}_+^{b(P)}} (Z - t)^\top \Sigma_{b(P)}^{-1}(P) (Z - t)$ measures the distance of the normal random vector $Z \sim N(\mathbf{0}_{b(P)}, \Sigma_{b(P)}(P))$ from the non-negative orthant $\mathbb{R}_+^{b(P)}$, and is known as the quasi-likelihood-ratio (QLR) statistic. Theorem 1 shows that the ELR statistic converges in distribution to the QLR statistic under the null hypothesis. As the QLR statistic has the chi-bar-square distribution (see, e.g., Section 3.8 of Silvapulle and Sen, 2005), which we denote by $\bar{\chi}_{b(P)}^2(\Sigma_{b(P)}(P))$, Theorem 1 implies that ELR_n has this distribution under H_0 in large-samples. Consequently, Theorem 1 extends Theorem 3.1 of Canay (2010) to the setup involving inequality restrictions on regular nonlinear functionals.

The result of Theorem 1 implies that the chi-bar-square distribution can be used to compute asymptotically valid critical values, which requires consistent estimators of $\Sigma_{b(P)}^{-1}(P)$ and $b(P)$. This approach is computationally demanding when $b(P) > 4$ because there are no closed-form expressions for the CDF of $\bar{\chi}_{b(P)}^2(\Sigma_{b(P)}(P))$. To circumvent these difficulties, Rosen (2008) proposes an alternative asymptotically valid approach that is computationally attractive because it avoids the estimation of $\Sigma_{b(P)}^{-1}(P)$ and $b(P)$, and the evaluation of the chi-bar-square CDF, by using a known upper bound on $b(P)$; see Corollary 2 and equation (15) of his paper. His procedure is in general asymptotically conservative. Alternatively, it is appealing to consider the (standard) bootstrap calibrated version of the ELR test; however, that approach is not asymptotically valid because the number of binding inequalities, $b(P)$, has a significant discontinuous effect on the shape of the $\bar{\chi}_{b(P)}^2(\Sigma_{b(P)}(P))$ distribution. Consequently, we propose to extend the empirical likelihood bootstrap for moment inequality models introduced by Canay (2010) to the setup of this paper.

The empirical likelihood bootstrap procedure follows steps identical to those of Canay (2010) except that the ELR statistic is given by (2.3), and one resamples the pseudo-values. This modification is straightforward to implement in practice, whose asymptotic validity relies on the Law of the Iterated Logarithm for U-statistics (Serfling, 1971), because $\beta(P)$ is a linear combination of expected values of known kernel functions of different degrees. We now describe this bootstrap procedure. The modified empirical likelihood function is

$$\max_{p_1, \dots, p_n} \left\{ \sum_{i=1}^n \log(p_i) \mid p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{V}_i \geq \varrho_n \right\}, \quad (3.2)$$

where $\varrho_n \geq \mathbf{0}_q$ is a $q \times 1$ vector of random (or deterministic) variables that satisfies

$$P \left(\lim_{n \rightarrow +\infty} \varrho_{j,n} = 0 \text{ and } \liminf_{n \rightarrow +\infty} \varrho_{j,n} \left(\frac{n}{2r_{\beta_j}^2 \log \log n} \right)^{1/2} \geq \Sigma_{jj}^{1/2}(P) \right) = 1, \quad \forall j = 1, \dots, q, \quad (3.3)$$

where $r_{\beta_j} \leq \max_{\ell: \mathbf{A}_{j\ell} \neq 0} r(\ell)$ is the degree of the kernel function that gives rise to the functional $\beta_j(P)$. By the Law of The Iterated Logarithm for U-statistics

$$P \left(\limsup_{n \rightarrow +\infty} (G_{j,n} - \varrho_{j,n}) \leq 0 \right) = \begin{cases} 0 & \text{if } E_P[\hat{V}_{ji}] > 0, \\ 1, & \text{if } E_P[\hat{V}_{ji}] = 0, \end{cases} \quad (3.4)$$

holds for all $j = 1, \dots, q$, so that the sequence $\varrho_{j,n}$ provides a rule to determine whether the j th inequality restriction is binding or slack, under H_0 . Denote by $\bar{P}_n = (\bar{p}_1, \dots, \bar{p}_n)$ the modified empirical likelihood probabilities that solve (3.2), and define $\bar{\mu}_n = \sum_{i=1}^n \bar{p}_i \hat{V}_i$. Let $\{\hat{V}_i^* : i \leq n\}$ be i.i.d according to \bar{P}_n so that $E_{\bar{P}_n}[\hat{V}_i^*] = \bar{\mu}_n$. The modified bootstrap jackknife ELR statistic is

$$ELR_n^* = \min_{p_1, \dots, p_n} \left\{ 2 \sum_{i=1}^n \log \left(\frac{1/n}{p_i} \right) \mid p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{V}_i^* \geq \varrho_n \right\}. \quad (3.5)$$

Let N_n be the number of bootstrap replications, the $(1 - \alpha)$ -quantile of $\{ELR_{n,\ell}^* : \ell \leq N_n\}$ is

$$\bar{c}_{n,1-\alpha} = \inf \left\{ x \in \mathbb{R} : \frac{1}{N_n} \sum_{\ell=1}^{N_n} 1 [ELR_{n,\ell}^* \leq x] \geq 1 - \alpha \right\}, \quad (3.6)$$

so that one rejects H_0 when $ELR_n > \bar{c}_{n,1-\alpha}$.

Before presenting the asymptotic validity of this bootstrap procedure, we introduce the space of null distributions over which the test will be valid with uniformity. We define this space through conditions on the Hájek projection of $G_n - \beta(P)$, which is given by $\hat{G}_n = n^{-1} \sum_{i=1}^n (\mathbf{A}\Psi(X_i))$, where the j component of Ψ is defined in (3.1). To simplify exposition, let $\Upsilon(x)$ denote the non-centered version of $\mathbf{A}\Psi(x)$, which is a q -vector whose components are

$$\Upsilon_j(x) = \sum_{\ell=1}^k \mathbf{A}_{j,\ell} r(\ell) E_P [h_\ell(X_1, X_2, \dots, X_{r(\ell)-1}, x)] \quad j = 1, \dots, q. \quad (3.7)$$

The basis for characterising the model in terms of conditions on the Hájek projection is that $\sqrt{n}(G_n - \beta(P) - \hat{G}_n) \xrightarrow{P} 0$, pointwise in P , by Theorem 11.2 in van der Vaart (2000). The model introduces enough conditions to extend this large-sample result to hold with uniformity, so that the uniform asymptotic behavior of \hat{G}_n determines that of the bootstrap ELR test.

Definition 1. Denote by \mathcal{M}_0 the space of all probabilities $P \in \mathcal{M}$ such that

- (i) $\beta(P) \geq \mathbf{0}_q$.
- (ii) $\{X_i : i \leq n\}$ is a random sample from $P \in \mathcal{M}$, where $n \geq \max_{1 \leq j \leq k} r(j)$;
- (iii) $E_P [h_j^2(X_1, \dots, X_{r(j)})] \leq \bar{B}$ for some $0 < \bar{B} < +\infty$ for all $j = 1, 2, \dots, k$;
- (iv) $\Sigma_{jj}(P) \in [\underline{\epsilon}, \bar{\epsilon}]$ for some $\underline{\epsilon} > 0$ and $\bar{\epsilon} < +\infty$, for all $j = 1, 2, \dots, q$,
- (v) $\det(\Omega(P)) \geq \epsilon$, where $\Omega(P) = D^{-1/2} \Sigma(P) D^{-1/2}$, and $D = \text{Diag}(\Sigma(P))$ denotes the $q \times q$ diagonal matrix whose diagonal elements equal those of $\Sigma(P)$; and
- (vi) $E_P [|\Upsilon_j(X_1) \Sigma_{jj}^{-1/2}(P)|^{2+\delta}] \leq M$ for all $j = 1, \dots, q$, where $\delta > 0$ and $M < +\infty$ are constants, and $\Upsilon_j(\cdot)$ is given by (3.7).

The following result describes the asymptotic behavior of the modified bootstrap procedure.

Theorem 2. Suppose that Assumption 1 and $\beta(P) \geq \mathbf{0}_q$, hold. Let $\varrho_n \geq \mathbf{0}_q$ be the vector satisfying (3.3), $\bar{c}_{n,1-\alpha}$ be given by (3.6), and \mathcal{M}_0 as given by Definition 1. Then, for $0 < \alpha < 1/2$, the statements below follow.

1. The modified bootstrap ELR statistic defined in (3.5) satisfies $ELR_n^* \xrightarrow{d} \bar{\chi}_{b(P)}^2(\Sigma_{b(P)}(P))$, where the convergence is conditional on $\{\bar{P}_n : n \geq 1\}$ for almost every sample path.
2. $\limsup_{n \rightarrow +\infty} P[ELR_n > \bar{c}_{n,1-\alpha}] \leq 1 - \alpha$.
3. $\limsup_{n \rightarrow +\infty} \sup_{P \in \mathcal{M}_0} P[ELR_n > \bar{c}_{n,1-\alpha}] = 1 - \alpha$.

Proof. See Appendix A.2. □

The first two results of Theorem 2 establish the pointwise (in P) asymptotic validity of the modified bootstrap procedure, and the third result establishes its uniform asymptotic validity. The results of this theorem are counterparts of the results in Theorem 4.1 of Canay (2010), who focuses on confidence sets; the same principle applies because of the duality of testing and the construction of confidence sets by test statistic inversion. Thus, Theorem 2 extends Theorem 4.1 of Canay (2010) to the setup involving inequality restrictions on regular nonlinear functionals of the underlying distribution.

4 Monte Carlo Study

This section evaluates the finite-sample performance of the bootstrap test developed in the previous section using an order restrictions on variances. The setup has $d = 3$, so that $X_i = [X_{1i}, X_{2i}, X_{3i}]$ for each $i = 1, \dots, n$, with the null hypothesis in (2.1) as $H_0 : \text{Var}(X_{1i}) \geq \text{Var}(X_{2i}) \geq \text{Var}(X_{3i})$. The variances index the data-generating processes (DGPs) we consider which are of the form $\{X_i\}_{i=1}^n \text{ IID } N(\mathbf{0}_d, \Pi)$, where the off-diagonal elements of the matrix Π are such that $\Pi_{j\ell} = (1/2)^{|j-\ell|} (\Pi_{jj}\Pi_{\ell\ell})^{1/2}$. In terms of the notation from the previous sections, we have $\mathbf{A}^\top = [\mathbf{a}_1^\top, \mathbf{a}_2^\top]$, where $\mathbf{a}_1 = [1, -1, 0]$ and $\mathbf{a}_2 = [0, 1, -1]$, $h(\mathbf{x}) = \frac{1}{2}[(x_{11} - x_{12})^2, (x_{21} - x_{22})^2, (x_{31} - x_{32})^2]^\top$, and $r_{\beta_j} = 2$ for $j = 1, 2$.

Each simulation experiment depends on two parameters: the sample size $n \in \{100, 300, 500\}$, and the variances Π_{jj} , $j = 1, 2, 3$. Across all experiments, the number of bootstrap samples was set to $N_n = 499$, the test size to $\alpha = 0.05$, the number of Monte Carlo simulations to 10^4 , and $\varrho_{n,j} = \hat{\sigma}_j r_{\beta_j} \sqrt{(2 \log \log n)/n}$ for $j = 1, \dots, q$, where $\hat{\sigma}_j^2$ is the sample variance based on the pseudo-values $\{\hat{V}_{ji}\}_{i=1}^n$. This choice for ϱ_n is a sample version of the benchmark value that Canay (2010) suggests for moment inequality models (see equation (4.9) of his paper), but for the setup of regular functionals based on (3.3). We also compare the performance of the test with the asymptotic procedure of Rosen (2008). As the maximal number of binding inequalities under H_0 is 2, his fixed asymptotic critical value is approximately equal to 5.1384, which is the solution of the following equation in c : $1/2 \text{Prob}[\chi_2^2 \geq c] + 1/2 \text{Prob}[\chi_1^2 \geq c] = \alpha$. By contrast, the proposed test uses a data-dependent critical value.

We consider two specifications under the null: (i) $\Pi_{11} = \Pi_{22} = \Pi_{33} = 1$, and (ii) $\Pi_{11} = \Pi_{22} = 1$ and $\Pi_{33} = 1/2$. In (i), all the constraints in H_0 are binding, and in (ii) the first constraint in H_0 is binding

and the second one is slack. Table 1 reports the empirical rejection probabilities of the proposed and Rosen (2008) tests, and their confidence intervals based on the normal approximation. The proposed bootstrap test performs really well across all n in the configuration $\Pi_{11} = \Pi_{22} = \Pi_{33} = 1$ because the empirical rejection probabilities are very close to the size $\alpha = 0.05$, and improves with the sample size under the configuration $\Pi_{11} = \Pi_{22} = 1$ and $\Pi_{33} = 1/2$. Rosen’s asymptotic procedure on the other hand is too conservative as the empirical rejection probabilities it yields can be much less than α , across all n and DGPs, which is because his critical value is too large.

Table 1: Simulations Under H_0

Sample Size	$\Pi_{11} = \Pi_{22} = \Pi_{33} = 1$		$\Pi_{11} = \Pi_{22} = 1$ and $\Pi_{33} = 1/2$	
	Bootstrap	Asymptotic	Bootstrap	Asymptotic
100	0.0886 [0.0540,0.0632]	0.044 [0.0404,0.0484]	0.0187 [0.0160,0.0214]	0.0135 [0.0112,0.0158]
300	0.0521 [0.0477,0.0565]	0.0358 [0.0394,0.0322]	0.0410 [0.0371,0.0449]	0.0114 [0.0096,0.0135]
500	0.0549 [0.0504,0.0594]	0.0369 [0.0332,0.0406]	0.0519 [0.0476,0.0562]	0.012 [0.0099,0.0141]

We consider two specifications under the alternative: (i) $\Pi_{11} = \Pi_{22} = 1$ and $\Pi_{33} = 1.2$, and (ii) $\Pi_{11} = 2, \Pi_{22} = 1$ and $\Pi_{33} = 1.2$. In both configurations, the second inequality in H_0 is violated, and what differentiates them is whether the first inequality is binding (case (i)) or slack (case (ii)). Table 2 reports the empirical rejection probabilities of the tests. The proposed bootstrap test outperforms Rosen’s asymptotic procedure in both configurations and across all n . This finding isn’t surprising because the fixed-critical-value in Rosen’s procedure is too large, while the proposed test uses a data-dependent critical value which accounts for the number of binding constraints. For example, with the configuration $\Pi_{11} = 2, \Pi_{22} = 1$ and $\Pi_{33} = 1.2$, in large-samples the bootstrap procedure detects that $\Pi_{11} > \Pi_{22}$ holds with probability tending to one, and thus, delivers a critical value that is suitable for that case (i.e., this slack inequality restriction is omitted from the computation of the critical value).

Table 2: Simulations Under H_1

Sample Size	$\Pi_{11} = \Pi_{22} = 1$ and $\Pi_{33} = 1.2$		$\Pi_{11} = 2, \Pi_{22} = 1$ and $\Pi_{33} = 1.2$	
	Bootstrap	Asymptotic	Bootstrap	Asymptotic
100	0.25 [0.2415,0.2585]	0.2056 [0.1977,0.2135]	0.1594 [0.1522,0.1666]	0.1177 [0.1114,0.1240]
300	0.5253 [0.5155,0.5351]	0.4649 [0.4551,0.4747]	0.4888 [0.4790,0.4986]	0.3288 [0.3196,0.3380]
500	0.7216 [0.7128,0.7304]	0.6761 [0.6669,0.6853]	0.7360 [0.7274,0.7446]	0.5289 [0.5188,0.5384]

5 Conclusion

This paper extends the domain of applicability of the empirical likelihood bootstrap of Canay (2010) for testing moment inequality restrictions to a more general class of functionals on the underlying distribution – those which are regular in the sense of Hoeffding (1948). The class of regular functionals includes all moments as special cases, and includes many nonlinear functionals that arise in empirical practice, such as those that define measures of dispersion and association. The basis of our approach is the computationally attractive JEL method of Jing et al. (2009), which converts nonlinear constraints into constraints on sample means based on jackknife pseudo-values.

6 Acknowledgments

The authors acknowledge the Sydney Informatics Hub and the University of Sydney's high performance computing cluster Artemis for providing the high performance computing resources that have contributed to the research results reported within this paper.

References

- Andrews, D. W. and P. Guggenberger (2009, 6). Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. *Econometric Theory* 25, 669–709.
- Andrews, D. W. and P. Guggenberger (2010, 4). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory* 26, 426–468.
- 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. and G. Soares (2010). Inference for Parameters Defined by Moment Inequalities using Generalized Moment Selection. *Econometrica* 78(1), 119–157.
- Beach, C. M. and R. Davidson (1983). Distribution-free statistical inference with lorenz curves and income shares. *The Review of Economic Studies* 50(4), 723–735.
- Beran, R. (1988). Pivoting Test Statistics: A Bootstrap View of Asymptotic Refinements. *Journal of the American Statistical Association* 83, 1967–1986.
- Brown, B. W. and W. Newey (2002). Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference. *Journal of Business and Economic Statistics* 20(4), 507–517.
- Canay, I. A. (2010). EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity. *Journal of Econometrics* 156(2), 408–425.
- Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34(3), 305 – 334.

- Dardanoni, V. and A. Forcina (1999). Inference for lorenz curve orderings. *The Econometrics Journal* 2(1), 49–75.
- DiCiccio, T., P. Hall, and J. Romano (1991, 06). Empirical likelihood is bartlett-correctable. *Ann. Statist.* 19(2), 1053–1061.
- Hoeffding, W. (1948, 09). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* 19(3), 293–325.
- Jing, B.-Y., J. Yuan, and W. Zhou (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association* 104(487), 1224–1232.
- Kitamura, Y. (2001). Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* 69(6), 1661–1672.
- Newey, W. and R. Smith (2004). Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica* 72(1), 219–255.
- Owen, A. (1990, 03). Empirical likelihood ratio confidence regions. *Ann. Statist.* 18(1), 90–120.
- Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75(2), 237–249.
- Rosen, A. M. (2008). Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *Journal of Econometrics* 146(1), 107 – 117.
- Serfling, R. J. (1971). The law of the iterated logarithm for u-statistics and related von mises statistics. *The Annals of Mathematical Statistics* 42(5), 1794. (abstracts).
- Shi, X. (1984). The approximate Independence of Jackknife Pseudo—values and The Bootstrap Methods. *Journal of Wuhan Institute of Hydraulic and Electric Engineering* 2, 83–90.
- Silvapulle, M. J. and P. K. Sen (2005). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. Wiley.
- van der Vaart, D. W. (2000). *Asymptotic Statistics* (First ed.). Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- Wood, A. T. A., K.-A. Do, and B. M. Broom (1996). Sequential linearization of empirical likelihood constraints with application to u-statistics. *Journal of Computational and Graphical Statistics* 5(4), 365–385.

A Proofs: Theorems 1 and 2

A.1 Theorem 1

Proof. The proof follows steps similar to those of Theorem 3.1 of Canay (2010). To derive the asymptotic distribution we use the dual parametrization of the ELR statistic (2.3) as in Canay (2010):

$$ELR_n = \min_{\tau \geq \mathbf{0}_q} \max_{\lambda(\tau) \in \mathbb{R}^q} \left\{ 2 \sum_{i=1}^n \log(1 + \lambda(\tau)^\top \hat{V}_i) - 2\lambda(\tau)^\top n\tau \right\} = \min_{\tau \geq \mathbf{0}_q} \max_{\lambda(\tau) \in \mathbb{R}^q} R_n(\lambda, \tau). \quad (\text{A.1})$$

In this representation of the ELR statistic, the derivative with respect to τ is non-negative, i.e. $\frac{\partial}{\partial \tau} R_n(\lambda, \tau) \geq \mathbf{0}_q$, which imposes the non-positiveness of the Lagrange multipliers on the inequality constraints, i.e. $\lambda \leq \mathbf{0}_q$. Additionally, note that $\hat{\tau}_n = \sum_{i=1}^n \hat{p}_i \hat{V}_i$, where the \hat{p}_i are the unique minimizers of (2.3). Under Assumption 1, Lemma B.1 in the Appendix indicates that ELR_n exists as $n \rightarrow +\infty$ with probability approaching unity.

Now consider an element j of $\hat{V}_{i,s}$. For each such $j \exists N$ such that $\forall n \geq N$, $G_{j,n,s} > 0$ with probability one. This implies that $\forall n \geq N$, $\hat{\tau}_{j,n,s} > 0$ with probability one since,

$$\hat{\tau}_{j,n,s} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{V}_{ji,s}}{1 + \hat{\lambda}_n^\top \hat{V}_i} \right) \geq \frac{G_{j,n,s}}{1 + \hat{\lambda}_n^\top G_n}$$

by Jensen's inequality, $\hat{\lambda}_n \leq \mathbf{0}_q$ and $\hat{\lambda}_n^\top G_n \geq 0$. The latter follows from

$$0 \leq \max_{\lambda \leq \mathbf{0}_q} 2 \sum_{i=1}^n \log[1 + \lambda^\top \hat{V}_i] \leq 2 \log \left[1 + \hat{\lambda}_n^\top \sum_{i=1}^n \hat{V}_i \right].$$

Given that $\hat{\tau}_{j,n,s} \times \hat{\lambda}_{j,n,s} = 0$ for all n , it follows that $\hat{\lambda}_{j,n,s} = 0$ for n large enough with probability one, and therefore, $\hat{\lambda}_{n,s} = \mathbf{0}_{s(P)}$ for n large enough with probability one. On the other hand, Lemma B.3 shows that $\|\hat{\lambda}_{n,b}\|_{\mathbb{R}^{b(P)}} = O_P(n^{-1/2})$. hence, $\exists N$ such that $\forall n \geq N$ the following equalities hold with probability one

$$\begin{aligned} ELR_n &= 2 \sum_{i=1}^n \log(1 + \hat{\lambda}_{n,b}^\top \hat{V}_{i,b} + \hat{\lambda}_{n,s}^\top \hat{V}_{i,s}) = 2 \sum_{i=1}^n \log(1 + \hat{\lambda}_{n,b}^\top \hat{V}_{i,b}) \\ &= \max_{\lambda_b(\tau) \leq \mathbf{0}_{b(P)}} 2 \sum_{i=1}^n \log(1 + \lambda_b(\tau)^\top \hat{V}_{i,b}) \\ &= \min_{\tau_b \geq \mathbf{0}_{b(P)}} \max_{\lambda_b(\tau_b) \in \mathbb{R}^{b(P)}} R_n(\lambda_b, \tau_b). \end{aligned}$$

Now we make a similar expansion to that on pages 100-102 of Owen (1990) for the expression given

above. For an arbitrary sequence $\mathbf{0}_{b(P)} \leq \tau_{n,b} = O_P(n^{-1/2})$, the first order condition for $\hat{\lambda}_{n,b}$ is given by:

$$-\sum_{i=1}^n \frac{\hat{V}_{i,b}}{1 + \hat{\lambda}_{n,b}^\top \hat{V}_{i,b}} + n\tau_{n,b} = \mathbf{0}_{b(P)}. \quad (\text{A.2})$$

Let $\gamma_i \equiv \hat{\lambda}_{n,b}^\top \hat{V}_{i,b}$ and note that by Lemma B.3 and Corollary B.1 we have:

$$\max_{1 \leq i \leq n} |\gamma_i| = O_P(n^{-1/2}) O_P(n^{1/2}) = O_P(1).$$

Expanding (A.2):

$$\begin{aligned} \mathbf{0}_{b(P)} &= -\sum_{i=1}^n \hat{V}_{i,b} (1 - \gamma_i + \gamma_i^2 / (1 + \gamma_i)) + n\tau_{n,b} \\ &= -G_{n,b} + \Sigma_{n,b} \hat{\lambda}_{n,b} - \frac{1}{n} \sum_{i=1}^n \hat{V}_{i,b} \gamma_i^2 / (1 + \gamma_i) + \tau_{n,b} \\ &= -G_{n,b} + \Sigma_{n,b} \hat{\lambda}_{n,b} + \tau_{n,b} + r_{1,n} \end{aligned}$$

where $\Sigma_{n,b} \equiv n^{-1} \sum_{i=1}^n \hat{V}_{i,b} \hat{V}_{i,b}^\top$ and $r_{1,n} \equiv -\frac{1}{n} \sum_{i=1}^n \hat{V}_{i,b} \gamma_i^2 / (1 + \gamma_i)$. Then,

$$\hat{\lambda}_{n,b} = \Sigma_{n,b}^{-1} (G_{n,b} - \tau_{n,b}) - \Sigma_{n,b}^{-1} r_{1,n} \quad (\text{A.3})$$

where

$$\begin{aligned} \|r_{1,n}\|_{\mathbb{R}^{b(P)}} &\leq \frac{1}{n} \sum_{i=1}^n \|\hat{V}_{i,b}\|_{\mathbb{R}^{b(P)}}^3 \|\hat{\lambda}_{n,b}(\tau_{n,b})\|_{\mathbb{R}^{b(P)}}^2 |1 + \gamma_i|^{-1} \\ &= O_P(1) O_P(n^{-1}) O_P(1) = O_P(n^{-1}), \end{aligned} \quad (\text{A.4})$$

by Lemma B.3 and Corollary B.1. Next, use $\log(1 + \gamma_i) = \gamma_i - \gamma_i^2/2 + r_{2,i}$, where for some finite $C > 0$

$$P(|r_{2,i}| \leq C |\gamma_i|^3, 1 \leq i \leq n) \rightarrow 1 \quad (\text{A.5})$$

as $n \rightarrow +\infty$. Now we can approximate the likelihood ratio and then use (A.3).

$$\begin{aligned}
ELR_n &= \min_{\tau_b \geq \mathbf{0}_{b(P)}} 2 \sum_{i=1}^n \log(1 + \hat{\lambda}_{n,b}^\top(\tau_b) \hat{V}_{i,b}) - 2 \hat{\lambda}_{n,b}^\top(\tau_b) n \tau_b \\
&= \min_{\tau_b \geq \mathbf{0}_{b(P)}} \left\{ 2 \hat{\lambda}_{n,b}^\top(\tau_b) \sum_{i=1}^n \hat{V}_{i,b} - \hat{\lambda}_{n,b}^\top(\tau_b) \sum_{i=1}^n \hat{V}_{i,b} \hat{V}_{i,b}^\top \hat{\lambda}_{n,b} - 2 \hat{\lambda}_{n,b}^\top(\tau_b) n \tau_b + 2 \sum_{i=1}^n r_{2,i} \right\} \\
&= \min_{\tau_b \geq \mathbf{0}_{b(P)}} \left\{ n \hat{\lambda}_{n,b}^\top(\tau_b) (G_{n,b} - \tau_b) + n \hat{\lambda}_{n,b}^\top r_{1,n} + 2 \sum_{i=1}^n r_{2,i} \right\} \\
&= \min_{\tau_b \geq \mathbf{0}_{b(P)}} \left\{ n \hat{\lambda}_{n,b}^\top(\tau_b) (G_{n,b} - \tau_b) \Sigma_{n,b}^{-1} (G_{n,b} - \tau_{n,b}) - n r_{1,n}^\top \Sigma_{n,b}^{-1} r_{1,n} + 2 \sum_{i=1}^n r_{2,i} \right\} \\
&= T_n + o_P(1)
\end{aligned} \tag{A.6}$$

where

$$\begin{aligned}
T_n &\equiv \min_{\tau_b \geq \mathbf{0}_{b(P)}} \left\{ n (G_{n,b} - \tau_b)^\top \Sigma_{n,b}^{-1} (G_{n,b} - \tau_{n,b}) \right\}, \\
n r_{1,n}^\top \Sigma_{n,b}^{-1} r_{1,n} &= n O_P(n^{-1}) O_P(1) O_P(n^{-1}) = O_P(n^{-1}),
\end{aligned}$$

and

$$\begin{aligned}
\left| \sum_{i=1}^n r_{2,i} \right| &\leq \sum_{i=1}^n |\gamma_i|^3 = n \|\hat{\lambda}_{n,b}(\tau_{n,b})\|_{\mathbb{R}^{b(P)}}^3 \frac{1}{n} \sum_{i=1}^n \|\hat{V}_{i,b}\|_{\mathbb{R}^{b(P)}}^3 \\
&= n O_P(n^{-3/2}) o(n^{1/2}) = O_P(n^{-1/2}) o(n^{1/2}) = o_P(1).
\end{aligned}$$

Next, define $Z_n \equiv \sqrt{n} G_{n,b} \xrightarrow{d} Z \sim N(\mathbf{0}_{b(P)}, \Sigma_{b(P)}(P))$, so that

$$T_n = \min_{\tau_b \geq \mathbf{0}_{b(P)}} \left\{ (Z_n - \sqrt{n} \tau_b)^\top \Sigma_{n,b}^{-1} (Z_n - \sqrt{n} \tau_{n,b}) \right\} \xrightarrow{d} \min_{t \geq \mathbf{0}_{b(P)}} \left\{ (Z - t)^\top \Sigma_{b(P)}^{-1}(P) (Z - t) \right\}$$

where $t = \sqrt{n} \tau_b$ and $\Sigma_{n,b} \xrightarrow{P} \Sigma_{b(P)}(P)$ by Assumption 1. □

A.2 Theorem 2

Proof. The proof follows steps similar to those of Theorem 4.1 of Canay (2010).

Part 1. To prove this part of the theorem, consider the j th element of $\hat{V}_{i,s}$, $\hat{V}_{j,i,s}$. For this component,

$$\begin{aligned}
\tau_{j,n,s}^* &= \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{V}_{j,i,s}^* - \varrho_{j,n,s}}{1 + \hat{\lambda}_n^{*\top} (\hat{V}_i^* - \varrho_n)} \right) \geq \frac{G_{j,n,s}^* - \varrho_{j,n,s}}{1 + \hat{\lambda}_n^{*\top} (G_n^* - \varrho_n)} \\
&= \frac{G_{j,n,s}^* - \bar{\mu}_{j,n,s}}{1 + \hat{\lambda}_n^{*\top} (G_n^* - \varrho_n)} + \frac{\bar{\mu}_{j,n,s} - \varrho_{j,n,s}}{1 + \hat{\lambda}_n^{*\top} (G_n^* - \varrho_n)},
\end{aligned}$$

where the first term is $O_P(n^{-1/2})$ \bar{P}_n -a.e. and the second term is positive for n larger enough with probability one by Lemma B.4. Thus, $\tau_{j,n,s}^* > 0$ for n sufficiently large \bar{P}_n -a.e. Furthermore, since $\tau_{j,n,s}^* \times \hat{\lambda}_{j,n,s}^* = 0$ for all n we can conclude that $\hat{\lambda}_{j,n,s}^* = 0$ for n sufficiently large \bar{P}_n -a.e. for all j such that $E_P[\hat{V}_{j^*}^*] > 0$. Additionally, by Lemma B.3 it follows that $\|\hat{\lambda}_{n,b}^*(\tau_{n,b})\|_{\mathbb{R}^{b(P)}} = O_P(n^{-1/2})$ \bar{P}_n -a.e. Therefore, similar arguments as those in the proof of Theorem 1 yield

$$ELR_n^* = \min_{\tau_b \geq \mathbf{0}_{b(P)}} \max_{\lambda_b \in \mathbb{R}^{b(P)}} 2 \sum_{i=1}^n \log(1 + \lambda_b^\top (\hat{V}_{i,b}^* - \varrho_{n,b})) - 2n\lambda_b^\top \tau_b,$$

so that the corresponding first order condition is,

$$\sum_{i=1}^n \left[\frac{\hat{V}_{i,b}^* - \varrho_{n,b}}{1 + \gamma_i^*} \right] - n\tau_b = \mathbf{0}_{b(P)}$$

where $\gamma_i^* \equiv \hat{\lambda}_{n,b}^{*\top} (\hat{V}_{i,b}^* - \varrho_{n,b})$, and $\max_{1 \leq i \leq n} |\gamma_i^*| = o_P(1)$ \bar{P}_n -a.e. Therefore, using the same expansion as in (A.6), the statistic ELR_n^* is equivalent to a QLR statistic of the form

$$\begin{aligned} ELR_n^* &= \min_{\tau_b \geq \mathbf{0}_{b(P)}} \left\{ n(G_{n,b}^* - \varrho_{n,b} - \tau_b)^\top \tilde{\Sigma}_{n,b}^{*-1} (G_{n,b}^* - \varrho_{n,b} - \tau_b) \right\} + o_P(1) \quad \bar{P}_n \text{ a.e.} \\ &= T_n^* + o_P(1) \quad \bar{P}_n \text{ a.e.}, \end{aligned}$$

where $\tilde{\Sigma}_{n,b}^* = n^{-1} \sum_{i=1}^n (\hat{V}_{i,b}^* - \varrho_{n,b})(\hat{V}_{i,b}^* - \varrho_{n,b})^\top$ and $G_{n,b}^* = n^{-1} \sum_{i=1}^n \hat{V}_{i,b}^*$. Next, define the random vector Z_n^* and note that,

$$Z_n^* \equiv \sqrt{n}(G_{n,b}^* - \bar{\mu}_{n,b}) \xrightarrow{d} Z \sim N(\mathbf{0}_{b(P)}, \Sigma_{b(P)}(P)),$$

where $\bar{\mu}_{n,b} = \sum_{i=1}^n \bar{p}_i \hat{V}_{i,b}$ and the convergence is \bar{P}_n -a.e. Then letting $t = \sqrt{n}(\tau_b - (\bar{\mu}_{n,b} - \varrho_{n,b}))$

$$\begin{aligned} T_n^* &= \min_{\tau_b \geq \mathbf{0}_{b(P)}} \left\{ (Z_n^* - \sqrt{n}(\tau_b - (\bar{\mu}_{n,b} - \varrho_{n,b})))^\top \tilde{\Sigma}_{n,b}^{*-1} (Z_n^* - \sqrt{n}(\tau_b - (\bar{\mu}_{n,b} - \varrho_{n,b}))) \right\} \\ &= \min_{t \geq -\sqrt{n}(\bar{\mu}_{n,b} - \varrho_{n,b})} (Z_n^* - t)^\top \tilde{\Sigma}_{n,b}^{*-1} (Z_n^* - t). \end{aligned}$$

By Lemma B.4, $-\sqrt{n}(\bar{\mu}_{n,b} - \varrho_{n,b}) = \mathbf{0}_{b(P)}$ with probability 1 for n sufficiently large, meaning that for n large enough the following equality holds with probability one:

$$T_n^* = \min_{t \geq \mathbf{0}_{b(P)}} (Z_n^* - t)^\top \tilde{\Sigma}_{n,b}^{*-1} (Z_n^* - t).$$

Because $\tilde{\Sigma}_{n,b}^* \xrightarrow{P} \Sigma_{b(P)}(P)$ \bar{P}_n -a.e, it follows that $T_n^* \xrightarrow{d} T$ \bar{P}_n -a.e., where

$$T = \min_{t \geq \mathbf{0}_{b(P)}} (Z - t)^\top \Sigma_{b(P)}^{-1}(P) (Z - t).$$

This is exactly the same asymptotic distribution of ELR_n and the result follows.

Part 2. Noting that $ELR_n = 0$ with probability one for $\beta(P) > \mathbf{0}_q$, it follows that

$$\limsup_{n \rightarrow +\infty} P[ELR_n > c] = 0$$

for any $c \geq 0$; and in particular, for $\bar{c}_{n,1-\alpha}$ as defined by (3.6). Otherwise, if $\exists j \in \{1, 2, \dots, q\}$ such that $\beta_j(P) = 0$, then from the result in Part 1

$$\limsup_{n \rightarrow +\infty} P[ELR_n > \bar{c}_{n,1-\alpha}] = \alpha.$$

Part 3. This part uses notation from Appendix C. Let $\varsigma_n^* = n^{1/2} \bar{D}_n^{-1/2} \left(n^{-1} \sum_{i=1}^n \hat{V}_i^* - \bar{\mu}_n \right)$. The JEL bootstrap critical value $\bar{c}_{n,1-\alpha}$ is the $(1 - \alpha)$ quantile of $T_n^* = \min_{t \in \mathbb{R}_+^q} (\varsigma_n^* + \bar{\varphi}_n - t)^\top \bar{\Omega}_n^{*-1} (\varsigma_n^* + \bar{\varphi}_n - t)$, where $\bar{\varphi}_n$ and $\bar{\Omega}_n^*$ are defined in (C.3) and (C.4), respectively. Now let $\{\gamma_{n,\nu} = (\gamma_{n,\nu,1}, \text{vech}_*(\Omega_{n,\nu}), P_{n,\nu}) : n \geq \max_{1 \leq j \leq k} r(j)\}$ be a sequence of parameters in Γ that satisfies (i) $n^{1/2} \gamma_{n,\nu,1} \rightarrow \nu_1 \in \mathbb{R}_{+\infty}^q$, (ii) $\kappa_{n,j}^{-1} n^{1/2} \gamma_{n,\nu,1,j} \rightarrow \pi_{1,j} \in \mathbb{R}_{+\infty}$ for $j = 1, \dots, q$, and (iii) $\text{vech}_*(\Omega_{n,\nu}) \rightarrow \nu_2 \in \mathbb{R}_{\pm\infty}^{q(q-1)/2}$. Let $\nu = (\nu_1, \nu_2)$, $\pi = (\pi_1, \pi_2)$ and $\pi_2 = \nu_2$. We which to show that

- (a) $\bar{c}_{n,1-\alpha} \geq \bar{c}_n^*$ a.s. for all n for a sequence of random variables $\{\bar{c}_n^* : n \geq \max_{1 \leq j \leq k} r(j)\}$ that satisfies $\bar{c}_n^* \xrightarrow{P} c_{\pi,1-\alpha}$ under $\{\gamma_{n,\nu} : n \geq \max_{1 \leq j \leq k} r(j)\}$;
- (b) $\liminf_{n \rightarrow +\infty} P_{\gamma_{n,\nu}}(T_n \leq \bar{c}_{n,1-\alpha}) \geq 1 - \alpha$, where T_n is defined in (C.2);
- (c) for any subsequence $\{w_n : n \geq \max_{1 \leq j \leq k} r(j)\}$ of $\{n : n \geq \max_{1 \leq j \leq k} r(j)\}$, 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 .

Parts (a)-(c) follow by the same arguments in Canay (2010); hence, we omit them for brevity. Consequently, the conclusions of Lemmas 2 and 3 of Andrews and Soares (2010) hold for $\bar{c}_{n,1-\alpha}$ in place of their critical value, and Theorem 1 of Andrews and Soares (2010) holds for the JEL bootstrap. \square

B Auxiliary Technical Results

In this section we present auxiliary technical results that are used in the proofs of Theorems 1 and 2. The first three technical results we present are versions of Lemma A.1, Lemma A.4 and Corollary A.1 in Jing et al. (2009) that apply to a vector of kernel functions whose components can have different degrees greater than or equal to 2. Lemma A.4 and Corollary A.1 in Jing et al. (2009) only applies to a single kernel function of degree 2. Throughout, $\|\cdot\|_{\mathbb{R}^a}$ denotes the L_2 -norm on \mathbb{R}^a .

Lemma B.1. Suppose that $E_P [\|\mathbf{A}h(X)\|_{\mathbb{R}^q}^2] < +\infty$ and $\Sigma_{jj}(P) > 0$ for each $j = 1, 2, \dots, q$. Then, as $n \rightarrow +\infty$, we have $P \left[\max_{1 \leq i \leq n} \hat{V}_{ji} > 0, j = 1, 2, \dots, q \right] \rightarrow 1$.

Proof. By Fréchet's inequality,

$$P \left[\max_{1 \leq i \leq n} \hat{V}_{ji} > 0, j = 1, 2, \dots, q \right] \geq \max \left\{ 0, \sum_{j=1}^q P \left[\max_{1 \leq i \leq n} \hat{V}_{ji} > 0 \right] - (q-1) \right\}. \quad (\text{B.1})$$

Hence, to derive the result, it is sufficient show that $P \left[\max_{1 \leq i \leq n} \hat{V}_{ji} \leq 0 \right] \rightarrow 0$ as $n \rightarrow +\infty$, for each $j = 1, 2, \dots, q$, because it implies that the right side of the inequality (B.1) tends to unity with the sample size. By using steps identical to those of Lemma A.1 of Jing et al. (2009), we can show that each of these probabilities tend to zero. We omit the details for brevity. \square

Lemma B.2. Define

$$H_n = \max_{\ell_1 < \ell_2 < \dots < \ell_{\max_j r(j)}} \sqrt{\sum_{j=1}^k h_j^2 (X_{\ell_1}, X_{\ell_2}, \dots, X_{\ell_{r(j)}})}, \quad (\text{B.2})$$

and suppose that $E_P[\|h(X_1, X_2, \dots, X_n)\|_{\mathbb{R}^k}^2] < +\infty$, holds. Then, with probability one, $H_n = o(n^{1/2})$.

Proof. The condition $E_P[\|h(X_1, X_2, \dots, X_n)\|_{\mathbb{R}^k}^2] < +\infty$ implies that

$$\sum_{j=1}^k h_j^2 (X_1, X_2, \dots, X_{r(j)}) > n \quad \text{finitely often,}$$

which implies that $\sqrt{\sum_{j=1}^k h_j^2 (X_1, X_2, \dots, X_{r(j)})} > n^{1/2}$ finitely often. In consequence,

$$H_n = \max_{\ell_1 < \ell_2 < \dots < \ell_{\max_j r(j)}} \sqrt{\sum_{j=1}^k h_j^2 (X_{\ell_1}, X_{\ell_2}, \dots, X_{\ell_{r(j)}})} > n^{1/2} \quad \text{finitely often.}$$

Furthermore, by the same argument, $H_n > An^{1/2}$ finitely often for any $A > 0$. Therefore,

$$\limsup_{n \rightarrow +\infty} H_n n^{-1/2} \leq A \quad \text{with probability one.} \quad (\text{B.3})$$

The inequality B.3 holds simultaneously with probability 1 for any countable set of values A , and hence, $H_n = o(n^{1/2})$ with probability one. \square

Corollary B.1. Define

$$W_n = \max_{1 \leq i \leq n} \|\hat{V}_{i,b}\|_{L_1(\mathbb{R}^{b(P)})} = \max_{1 \leq i \leq n} \sum_{j=1}^{b(P)} |\hat{V}_{ji,b}|, \quad (\text{B.4})$$

and suppose that $E_P[\|h(X_1, X_2, \dots, X_n)\|_{\mathbb{R}^k}^2] < +\infty$. Then with probability one, $W_n = o(n^{1/2})$ and $n^{-1} \sum_{i=1}^n \|\hat{V}_{i,b}\|_{\mathbb{R}^{b(P)}}^3 = o(n^{1/2})$.

Proof. For any $1 \leq j \leq k$, it is easy to check that

$$U_{j,n} = \binom{n}{r(j)}^{-1} \sum_{1 \leq \ell_1 < \ell_2 < \dots < \ell_{i-1} < \ell_{i+1} < \dots < \ell_{r(j)}} h_j(X_{\ell_1}, \dots, X_{\ell_{i-1}}, X_i, X_{\ell_{i+1}}, X_{\ell_{r(j)}}) + \frac{n-r(j)}{n} U_{j,n-1}^{-i}.$$

Then for any $1 \leq i \leq n$

$$\begin{aligned} \|\hat{V}_{i,b}\|_{L_1(\mathbb{R}^{b(P)})} &= \|nG_{n,b} - (n-1)G_{n-1,b}^{-i}\|_{L_1(\mathbb{R}^{b(P)})} \\ &\leq \|\mathbf{A}\|_{\max} \|(nU_n - (n-1)U_{n-1}^{-i})\|_{L_1(\mathbb{R}^k)}, \end{aligned}$$

We now focus on the term $\|(nU_n - (n-1)U_{n-1}^{-i})\|_{L_1(\mathbb{R}^k)}$. For each j , $(nU_{j,n} - (n-1)U_{j,n-1}^{-i})$ equals

$$\frac{n}{\binom{n}{r(j)}} \sum_{1 \leq \ell_1 < \ell_2 < \dots < \ell_{i-1} < \ell_{i+1} < \dots < \ell_{r(j)}} h_j(X_{\ell_1}, \dots, X_{\ell_{i-1}}, X_i, X_{\ell_{i+1}}, X_{\ell_{r(j)}}) - (1-r(j))U_{j,n-1}^{-i}.$$

By the Triangular inequality, $|nU_{j,n} - (n-1)U_{j,n-1}^{-i}|$ is less than or equal to

$$\frac{n}{\binom{n}{r(j)}} \sum_{\ell_1 < \ell_2 < \dots < \ell_{i-1} < \ell_{i+1} < \dots < \ell_{r(j)}} |h_j(X_{\ell_1}, \dots, X_{\ell_{i-1}}, X_i, X_{\ell_{i+1}}, X_{\ell_{r(j)}})| + (1+r(j))|U_{j,n-1}^{-i}|.$$

Summing over j yields

$$\sum_{j=1}^k \frac{n}{\binom{n}{r(j)}} \sum_{\ell_1 < \ell_2 < \dots < \ell_{i-1} < \ell_{i+1} < \dots < \ell_{r(j)}} |h_j(X_{\ell_1}, \dots, X_{\ell_{i-1}}, X_i, X_{\ell_{i+1}}, X_{\ell_{r(j)}})| + \sum_{j=1}^k (1+r(j))|U_{j,n-1}^{-i}|,$$

which is less than or equal to

$$(2 + \max_j r(j)) \max_{\ell_1 < \ell_2 < \dots < \ell_{\max_j r(j)}} \sum_{j=1}^k |h_j(X_{\ell_1}, X_{\ell_2}, \dots, X_{\ell_{r(j)}})| \leq (2 + \max_j r(j)) H_n$$

where the last inequality follows from the Cauchy-Schwartz inequality. Therefore,

$$W_n \leq (2 + \max_j r(j)) H_n,$$

and by Lemma B.2, $W_n = o(n^{1/2})$ with probability 1.

From the above manipulation, we have

$$n^{-1} \sum_{i=1}^n \|\hat{V}_{i,b}\|_{\mathbb{R}^{b(P)}}^3 \leq \max_{1 \leq i \leq n} \|\hat{V}_{i,b}\|_{\mathbb{R}^{b(P)}} n^{-1} \sum_{i=1}^n \|\hat{V}_{i,b}\|_{\mathbb{R}^{b(P)}}^2 \leq W_n n^{-1} \sum_{i=1}^n \|\hat{V}_{i,b}\|_{\mathbb{R}^{b(P)}}^2.$$

Thus, to prove the result, we must show that $n^{-1} \sum_{i=1}^n \|\hat{V}_{i,b}\|_{\mathbb{R}^{b(P)}}^2$ converges to a positive constant with

probability one. To that end, note that

$$n^{-1} \sum_{i=1}^n \|\hat{V}_{i,b}\|_{\mathbb{R}^{b(P)}}^2 = n^{-1} \sum_{i=1}^n \|\hat{V}_{i,b} - G_{n,b} + G_{n,b}\|_{\mathbb{R}^{b(P)}}^2 = n^{-1} \sum_{i=1}^n \|\hat{V}_{i,b} - G_{n,b}\|_{\mathbb{R}^{b(P)}}^2 + \|G_{n,b}\|_{\mathbb{R}^{b(P)}}^2,$$

where we made use of the fact that $n^{-1} \sum_{i=1}^n \hat{V}_{i,b} = G_{n,b}$. Because G_b is a vector whose components are finite linear combinations of U-statistics, we can apply the Strong Law of Large numbers for U-statistics to it, to deduce that $\|G_{n,b}\|_{\mathbb{R}^{b(P)}}^2 = o(1)$ with probability one.

The remaining term can be bounded from above as follows

$$n^{-1} \sum_{i=1}^n \|\hat{V}_{i,b} - G_{n,b}\|_{\mathbb{R}^{b(P)}}^2 \leq n^{-1} \sum_{i=1}^n \|\hat{V}_i - G_n\|_{\mathbb{R}^q}^2 \leq \|\mathbf{A}\|_{\max}^2 n^{-1} \sum_{i=1}^n \|(nU_n - (n-1)U_{n-1}^{-i}) - U_n\|_{\mathbb{R}^q}^2,$$

where $n^{-1} \sum_{i=1}^n \|(nU_n - (n-1)U_{n-1}^{-i}) - U_n\|_{\mathbb{R}^q}^2$ is the jackknife estimate of $\sum_{j=1}^k E_P[\Psi_j^2(X_1)]$. Following arguments identical to those of the proof of Lemma A.3 in Jing et al. (2009), we can show that each j th component converges almost surely to $E_P[\Psi_j^2(X_1)]$; therefore,

$$n^{-1} \sum_{i=1}^n \|(nU_n - (n-1)U_{n-1}^{-i}) - U_n\|_{\mathbb{R}^q}^2 = \sum_{j=1}^k E_P[\Psi_j^2(X_1)] + o(1) \quad \text{a.s.},$$

so that $n^{-1} \sum_{i=1}^n \|\hat{V}_{i,b}\|_{\mathbb{R}^{b(P)}}^2$ is bounded from above by $\sum_{j=1}^k E_P[\Psi_j^2(X_1)]$ with probability one, and this concludes the proof as $W_n = o(n^{1/2})$ with probability one. \square

The next technical result is a version of Lemma B.2 in Canay (2010).

Lemma B.3. Consider the set of binding moments $E_P[\hat{V}_i] = \mathbf{0}_{b(P)}$, and define $\hat{\lambda}_{n,b}^*(\tau_{n,b})$ as,

$$\hat{\lambda}_{n,b}^*(\tau_{n,b}) = \arg \max_{\lambda_b \in \mathbb{R}^b} 2 \sum_{i=1}^n \log \left[1 + \lambda_b^\top (\hat{V}_{i,b}^* - \varrho_{n,b}) \right] - 2n \lambda_b^\top \tau_{n,b}, \quad (\text{B.5})$$

where $\hat{V}_{i,b}^* \sim \bar{P}_n$ as defined in (3.2), $\mathbf{0}_{b(P)} \leq \tau_{n,b} = O_P(n^{-1/2})$, and ϱ_n is the sequence in (3.3). Then, under Assumption 1, $\|\hat{\lambda}_{n,b}^*(\tau_{n,b})\|_{\mathbb{R}^{b(P)}} = O_P(n^{-1/2})$ conditional on $\{\bar{P}_n : n \geq 1\}$ for almost every sample path. In addition, letting $\hat{\lambda}_{n,b}(\tau_{n,b})$ be defined as in (B.5) but setting $\varrho_{n,b} = \mathbf{0}_{b(P)}$ and replacing \hat{V}_i^* with \hat{V}_i . Under the same assumption, $\|\hat{\lambda}_{n,b}(\tau_{n,b})\|_{\mathbb{R}^{b(P)}} = O_P(n^{-1/2})$.

Proof. The proof follows steps identical to those of Lemma B.2 in Canay (2010), except for an appropriate adjustment in notation and use of lemmas. The first order condition for $\hat{\lambda}_{n,b}^* \equiv \hat{\lambda}_{n,b}^*(\tau_{n,b})$ is,

$$\mathbf{0}_q = \frac{1}{n} \sum_{i=1}^n \frac{(\hat{V}_{i,b}^* - \varrho_{n,b})}{1 + \lambda_b^\top (\hat{V}_{i,b}^* - \varrho_{n,b})} - \tau_{n,b} \equiv g(\hat{\lambda}_{n,b}^*).$$

Let $\hat{\lambda}_{n,b}^* = c_n a_n$, where $c_n \geq 0$ and $\|a_n\|_{\mathbb{R}^{b(P)}} = 1$. Now,

$$\begin{aligned} 0 &= \|g(\hat{\lambda}_{n,b}^*)\|_{\mathbb{R}^{b(P)}} \geq |a_n^\top g(c_n a_n)| \\ &= \frac{1}{n} \left| a_n \left(\sum_{i=1}^n (\hat{V}_{i,b}^* - \varrho_{n,b}) - c_n \sum_{i=1}^n \frac{(\hat{V}_{i,b}^* - \varrho_{n,b}) a_n^\top (\hat{V}_{i,b}^* - \varrho_{n,b})}{1 + c_n a_n^\top (\hat{V}_{i,b}^* - \varrho_{n,b})} - \tau_{n,b} \right) \right| \\ &\geq \frac{c_n}{n} \sum_{i=1}^n \frac{a_n^\top (\hat{V}_{i,b}^* - \varrho_{n,b}) (\hat{V}_{i,b}^* - \varrho_{n,b})^\top a_n}{1 + c_n a_n^\top (\hat{V}_{i,b}^* - \varrho_{n,b})} - \frac{1}{n} \left| \sum_{j=1}^{b(P)} e_j^\top \sum_{i=1}^n (\hat{V}_{i,b}^* - \varrho_{n,b}) - \tau_{n,b} \right|, \end{aligned}$$

where e_j is the unit vector in the j th coordinate direction. Now let $A_n = \max_i a_n^\top (\hat{V}_{i,b}^* - \varrho_{n,b})$, and note that $A_n = o(n^{-1/2})$ by similar arguments to those in Corollary B.1. Note that $\tilde{\Sigma}_{n,b}^* = n^{-1} \sum_{i=1}^n (\hat{V}_{i,b}^* - \varrho_{n,b}) (\hat{V}_{i,b}^* - \varrho_{n,b})^\top \xrightarrow{P} \Sigma_{b(P)}$, \bar{P}_n -a.e. By Assumption 1, $\sigma_{(1)} > 0$, where $\sigma_{(1)}$ is the smallest eigenvalue of $\Sigma_{b(P)}$ so that $a^\top \Sigma_{b(P)} a_n \geq \sigma_{(1)} + o_P(1)$ and

$$\begin{aligned} 0 &\geq \frac{c_n a_n^\top \tilde{\Sigma}_{n,b}^* a_n}{1 + c_n A_n} - \frac{1}{n} \left| \sum_{j=1}^{b(P)} e_j^\top \sum_{i=1}^n (\hat{V}_{i,b}^* - \varrho_{n,b}) - \tau_{n,b} \right| \\ &\geq \frac{c_n (\sigma_{(1)} + o_P(1))}{1 + c_n A_n} - \frac{1}{n} \left| \sum_{j=1}^{b(P)} e_j^\top \sum_{i=1}^n (\hat{V}_{i,b}^* - \varrho_{n,b}) - \tau_{n,b} \right|, \end{aligned}$$

so that

$$\frac{c_n (\sigma_{(1)} + o_P(1))}{1 + c_n A_n} \leq \frac{1}{n} \left| \sum_{j=1}^{b(P)} e_j^\top \sum_{i=1}^n (\hat{V}_{i,b}^* - \bar{\mu}_{n,b}) \right| + \left| \sum_{j=1}^{b(P)} e_j^\top (\bar{\mu}_{n,b} - \varrho_{n,b}) \right| + |\tau_{n,b}| \leq O_P(n^{-1/2}),$$

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

$$c_n = \|\hat{\lambda}_{n,b}^*(\tau_{n,b})\|_{\mathbb{R}^{b(P)}} \leq \frac{O_P(n^{-1/2})}{\sigma_{(1)} + o_P(1)}.$$

Finally, note that setting $\varrho_{n,b} = \mathbf{0}_{b(P)}$ and replacing \hat{V}_i^* with \hat{V}_i , we have $\|\hat{\lambda}_{n,b}(\tau_{n,b})\|_{\mathbb{R}^{b(P)}} = O_P(n^{-1/2})$ following the same steps as above without re-centering by $\bar{\mu}_{n,b}$ in the last part. \square

The next technical result is a version of Lemma B.4 in Canay (2010).

Lemma B.4. Let $\bar{\mu}_n = \sum_{i=1}^n \bar{p}_i \hat{V}_i$ denote mean of the modified empirical likelihood problem, and ϱ_n be the sequence in (3.3). Under Assumption 1 the following two statements hold.

1. $\bar{\mu}_{n,s} > \varrho_{n,s}$ for n large enough with probability one.
2. $\bar{\mu}_{n,b} = \varrho_{n,b}$ for n large enough with probability one.

Proof. The proof follows steps identical to those of Lemma B.4 in Canay (2010), except for an appropriate adjustment in notation and the use of the Law of the Iterated Logarithm for U-statistics instead of its counterpart for sample means.

Part 1. Recall that $\bar{\mu}_{j,n,s}$ comes from problem (3.2) and note that from

$$0 \leq \max_{\lambda \leq \mathbf{0}_q} 2 \sum_{i=1}^n \log[(1 + \lambda^\top (\hat{V}_i - \varrho_n))] \leq 2 \log \left[1 + \hat{\lambda}_n^\top \sum_{i=1}^n (\hat{V}_i - \varrho_n) \right],$$

it follows that $\hat{\lambda}^\top \sum_{i=1}^n (\hat{V}_i - \varrho_n) \geq 0$. Now, from $G_n = n^{-1} \sum_{i=1}^n \hat{V}_i$, the FOC of $\lambda_{j,n,s}$ and Jensen's inequality

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{V}_{ji,s} - \varrho_{j,n,s}}{1 + \hat{\lambda}_n^\top \sum_{i=1}^n (\hat{V}_i - \varrho_n)} \right) \geq \frac{G_{j,n,s} - \varrho_{j,n,s}}{1 + \hat{\lambda}_n^\top (G_n - \varrho_n)}. \quad (\text{B.6})$$

Because $\hat{\lambda}^\top (G_n - \varrho_n) \geq 0$ for all n and $G_{j,n,s} - \varrho_{j,n,s} > 0$ for n large enough with probability one by (3.4) we have,

$$P \left(\liminf_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{V}_{ji,s} - \varrho_{j,n,s}}{1 + \hat{\lambda}_n^\top (\hat{V}_i - \varrho_n)} \right) > 0 \right) = 1$$

meaning that $\bar{\mu}_{j,n,s} \equiv n^{-1} \sum_{i=1}^n \left(\frac{\hat{V}_{ji,s}}{1 + \hat{\lambda}_n^\top (\hat{V}_i - \varrho_n)} \right) > \varrho_{j,n,s}$ with probability 1 for n sufficiently large.

Part 2. Set $\lambda = \mathbf{0}_q$ and consider the FOC for $\lambda_{j,b}$,

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{V}_{ji,b} - \varrho_{j,n,b}}{1 + \lambda^\top (\hat{V}_i - \varrho_n)} \right) \Big|_{\lambda=\mathbf{0}_q} = G_{j,n,b} - \varrho_{j,n,b}. \quad (\text{B.7})$$

By (3.4) we know that $G_{j,n,b} - \varrho_{j,n,b} \leq 0$ for n large with probability 1. If $G_{j,n,b} - \varrho_{j,n,b} = 0$, we are done since $\bar{\mu}_{j,n,b} = G_{j,n,b} = \varrho_{j,n,b}$ and $\lambda_{j,b} = 0$ is optimal. If $G_{j,n,b} - \varrho_{j,n,b} < 0$, then the optimal value of $\lambda_{j,b}$ has to decrease (so it will be negative) by continuity of the objective function in $\lambda_{j,b}$. Since the optimal solution has to satisfy $\lambda_{j,b}(\bar{\mu}_{j,n,b} - \varrho_{j,n,b}) = 0$, it follows that $\bar{\mu}_{j,n,b} = \varrho_{j,n,b}$. \square

C Auxiliary Notation and Technical Results

This appendix develops an alternative parametrization of the model \mathcal{M}_0 described in Definition 1, presents two technical results in the next section. We are essentially following Appendix C of Canay (2010) but in the context of U-statistics. We stick to his notation whenever possible. Let $(\gamma_1, \gamma_2, \gamma_3) \in \Gamma$ with components

$$\gamma_{1,j} = \Sigma_{jj}^{-1/2}(P) E_P[\Upsilon_j(X_1)] \quad \text{for } j = 1, \dots, q, \quad \gamma_2 = \text{vech}_*(\Omega(P)), \quad \gamma_3 = P, \quad (\text{C.1})$$

where Υ_j is defined in (3.7), and vech_* gives the elements of a matrix below the main diagonal. The parameter space for γ is $\Gamma = \{\gamma = (\gamma_1, \gamma_2, \gamma_3) : \text{for some } P \in \mathcal{M}_0\}$. Under any sequence $\{\gamma_{n,\nu} =$

$(\gamma_{n,\nu,1}, \text{vech}_*(\Omega_{n,\nu}), P_{n,\nu}) : n \geq \max_{1 \leq j \leq k} r(j)$ of parameters in Γ that satisfies $n^{1/2}\gamma_{n,\nu,1} \rightarrow \nu_1 \in \mathbb{R}_{+, \infty}^q$ and $\text{vech}_*(\Omega_{n,\nu}) \rightarrow \nu_2 \in \mathbb{R}_{\pm\infty}^{q(q-1)/2}$, Lemma C.1 and Corollary C.1 (in the ensuing section) imply that $ELR_n - T_n = o_P(1)$, where

$$T_n \equiv \min_{\tau_b \geq \mathbf{0}_{b(P_{n,\nu})}} \left\{ n(\widehat{G}_{n,b} - \tau_b)^\top \widehat{\Sigma}_{n,b}^{-1} (\widehat{G}_{n,b} - \tau_b) \right\} \quad \text{and} \quad (C.2)$$

$$\widehat{\Sigma}_n \equiv n^{-1} \sum_{i=1}^n \left(\Upsilon(X_i) - \widehat{G}_n \right) \left(\Upsilon(X_i) - \widehat{G}_n \right)^\top.$$

This result is an extension of (A.6); hence, ELR_n satisfies Assumption B0 in Andrews and Guggenberger (2009). Consequently, we can focus on T_n in proving Part 3 of Theorem 2.

Let $\bar{\Sigma}_n = \sum_{i=1}^n \bar{p}_i \left(\hat{V}_i - \bar{\mu}_n \right) \left(\hat{V}_i - \bar{\mu}_n \right)^\top$, $\bar{D}_n = \text{Diag}(\bar{\Sigma}_n)$, and $\varrho_{n,j} = n^{-1/2} \kappa_{n,j} \times [\bar{D}_n]_{jj}^{1/2}$ where $\kappa_{n,j} = \sqrt{2 r_{\beta_j}^2 \log \log n}$ for $j = 1, \dots, q$. Define the GMS function $\bar{\varphi}_n$ as the q -vector

$$\bar{\varphi}_n = n^{-1/2} \bar{D}_n^{-1/2} (\bar{\mu}_n - \varrho_n). \quad (C.3)$$

One can use arguments similar to those in Lemma B.4 to show that $\bar{\varphi}_n$ is indeed a GMS function that satisfies Assumptions GMS1, GMS2, and GMS6 in Andrews and Soares (2010).

Next we define the bootstrap correlation matrix, $\bar{\Omega}_n^*$. It is given by

$$\bar{\Omega}_n^* = \bar{D}_n^{-1/2} \bar{\Sigma}_n^* \bar{D}_n^{-1/2}, \quad \text{where} \quad \bar{\Sigma}_n^* = n^{-1} \sum_{i=1}^n \bar{p}_i \left(\hat{V}_i^* - \bar{\mu}_n \right) \left(\hat{V}_i^* - \bar{\mu}_n \right)^\top, \quad (C.4)$$

and $\hat{V}_i^* \sim \bar{P}_n$ denotes bootstrap data.

C.1 Technical Results

Lemma C.1. Under any sequence $\{\gamma_{n,\nu} = (\gamma_{n,\nu,1}, \text{vech}_*(\Omega_{n,\nu}), P_{n,\nu}) : n \geq \max_{1 \leq j \leq k} r(j)\}$ of parameters in Γ that satisfies $n^{1/2}\gamma_{n,\nu,1} \rightarrow \nu_1 \in \mathbb{R}_{+, \infty}^q$ and $\text{vech}_*(\Omega_{n,\nu}) \rightarrow \nu_2 \in \mathbb{R}_{\pm\infty}^{q(q-1)/2}$,

$$\lim_{n \rightarrow +\infty} \frac{\text{VAR}_{P_{n,\nu}}(G_{n,j})}{\text{VAR}_{P_{n,\nu}}(\widehat{G}_{n,j})} = 1 \quad \forall j = 1, \dots, q.$$

Proof. Let $j \in \{1, \dots, q\}$. Firstly, note that

$$\text{VAR}_{P_{n,\nu}}(\widehat{G}_{n,j}) = n^{-1} \Sigma_{jj}(P_{n,\nu}) = n^{-1} \sum_{\ell=1}^k \mathbf{A}_{j,\ell}^2 E_{P_{n,\nu}} [\Psi_\ell^2(X_1)] \quad \text{and}$$

$$\text{VAR}_{P_{n,\nu}}(G_{n,j}) = \sum_{\ell=1}^k \mathbf{A}_{j,\ell}^2 \text{VAR}_{P_{n,\nu}}(U_{n,\ell}). \quad (C.5)$$

Then, using the manipulations from Section 5 of Hoeffding (1948), for each ℓ ,

$$\text{VAR}_{P_{n,\nu}}(U_{n,\ell}) = n^{-1} E_{P_{n,\nu}} [\Psi_\ell^2(X_1)] + \Delta_{n,\ell},$$

where

$$\begin{aligned} \Delta_{n,\ell} &= \sum_{c=2}^{r(\ell)} \frac{r(\ell)!^2}{c!(r(\ell)-c)!^2} \frac{(n-r(\ell))(n-r(\ell)-1)\cdots(n-2r(\ell)+c+1)}{n(n-1)\cdots(n-r+1)} \eta_c, \\ \eta_c &= E_{P_{n,\nu}} [h_{\ell,c}^2(X_1, \dots, X_c)] - (E_{P_{n,\nu}} [h_\ell(X_1, \dots, X_{r(\ell)})])^2, \\ h_{\ell,c}(x_1, \dots, x_c) &= E_{P_{n,\nu}} [h_\ell(x_1, x_2, \dots, x_c, X_{c+1}, \dots, X_{r(\ell)})]. \end{aligned}$$

Therefore,

$$\frac{\text{VAR}_{P_{n,\nu}}(G_{n,j})}{\text{VAR}_{P_{n,\nu}}(\widehat{G}_{n,j})} = 1 + \frac{n \sum_{\ell=1}^k \mathbf{A}_{j,\ell}^2 \Delta_{n,\ell}}{\Sigma_{jj}(P_{n,\nu})},$$

where

$$\left| \frac{n \sum_{\ell=1}^k \mathbf{A}_{j,\ell}^2 \Delta_{n,\ell}}{\Sigma_{jj}(P_{n,\nu})} \right| \leq \frac{O(n^{-1})\overline{B}}{\underline{\epsilon}} \rightarrow 0$$

as $n \rightarrow +\infty$, because

$$\frac{(n-r(\ell))(n-r(\ell)-1)\cdots(n-2r(\ell)+c+1)}{n(n-1)\cdots(n-r+1)} = O(n^{-2}) \quad \forall \ell = 1, \dots, k,$$

and for each n , $\eta_c \leq \overline{B}$ for $c = 2, \dots, r(\ell)$ and $\Sigma_{jj}(P_{n,\nu}) \geq \underline{\epsilon}$. As $j \in \{1, \dots, q\}$ was arbitrary, the above manipulations hold for all such j , which concludes the proof. \square

Corollary C.1. $\sqrt{n} \left(G_n - \beta(P_{n,\nu}) - \widehat{G}_n \right) \xrightarrow{P} \mathbf{0}_q$, along $\{\gamma_{n,\nu} : n \geq \max_{1 \leq j \leq k} r(j)\}$ satisfying the conditions of Lemma C.1.

Proof. The proof follows steps similar to those of the proof of Theorem 11.2 in van der Vaart (2000). We prove convergence in second mean, which is stronger. The expectation of the difference is zero. For each $j \in \{1, \dots, q\}$, its variance is equal to

$$n \left(\text{VAR}_{P_{n,\nu}}(G_{n,j}) + \text{VAR}_{P_{n,\nu}}(\widehat{G}_{n,j}) - 2\text{COV}_{P_{n,\nu}}(G_{n,j} - \beta_j(P_{n,\nu}), \widehat{G}_{n,j}) \right). \quad (\text{C.6})$$

Because $\widehat{G}_{n,j}$ is the Hájek projection of $G_{n,j} - \beta_j(P_{n,\nu})$, $G_{n,j} - \beta_j(P_{n,\nu}) - \widehat{G}_{n,j}$ and $\widehat{G}_{n,j}$ are orthogonal, and therefore,

$$E_{P_{n,\nu}} \left[(G_{n,j} - \beta_j(P_{n,\nu})) \widehat{G}_{n,j} \right] = E_{P_{n,\nu}} \left[\widehat{G}_{n,j}^2 \right]$$

holds for each n , and is equal to $\text{VAR}_{P_{n,\nu}}(\widehat{G}_{n,j})$. Consequently, the expression (C.6) simplifies to

$$n \left(\text{VAR}_{P_{n,\nu}}(G_{n,j}) + \text{VAR}_{P_{n,\nu}}(\widehat{G}_{n,j}) - 2\text{VAR}_{P_{n,\nu}}(\widehat{G}_{n,j}) \right) = n \left(\text{VAR}_{P_{n,\nu}}(G_{n,j}) - \text{VAR}_{P_{n,\nu}}(\widehat{G}_{n,j}) \right),$$

and hence, equals

$$n \text{VAR}_{P_{n,\nu}}(\widehat{G}_{n,j}) \left(\frac{\text{VAR}_{P_{n,\nu}}(G_{n,j})}{\text{VAR}_{P_{n,\nu}}(\widehat{G}_{n,j})} - 1 \right).$$

This expression tends to zero by Lemma C.1, because $n \text{VAR}_{P_{n,\nu}}(\widehat{G}_{n,j}) = \Sigma_{jj}(P_{n,\nu}) \in [\underline{\epsilon}, \bar{\epsilon}]$ for each n . □