Economics Working Paper Series

2015 - 15

An Improved Bootstrap Test for Restricted Stochastic Dominance

Thomas M. Loki and Rami V. Tabri

May 10 2016

An Improved Bootstrap Test for Restricted Stochastic Dominance

Thomas M. Loki[*] and Rami V. Tabri[†]

May 10, 2016

Abstract

This paper proposes a method of testing for restricted stochastic dominance between two income distributions based on the bootstrap test of Linton et al. (2010) (LSW). The proposed testing procedure reformulates the LSW bootstrap test statistics using an estimator of the contact set based on the method of constrained empirical likelihood that imposes the restrictions of the null hypothesis. The testing procedure of this paper is uniformly asymptotically valid, and less conservative than the one LSW propose. Furthermore, it is consistent and has the same asymptotic local power properties as the LSW test. We report simulation results that show the proposed test is noticeably less conservative than the test of LSW and improves its power.

JEL Classification: C12 (Hypothesis Testing); C14 (Semiparametric and Nonparametric Methods)

Keywords: Empirical Likelihood; Constrained Estimation; Restricted Stochastic Dominance; Bootstrap Test.

[*]School of Economics, The University of Sydney, Sydney, New South Wales 2006, Australia.

[†]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.

1 Introduction

Stochastic dominance orderings of income distributions are fundamental in poverty and income studies. Since the 1980's, these orderings were used to establish whether poverty or social welfare is greater in one income distribution than in another for general classes of poverty indices and for ranges of possible poverty lines (e.g. Atkinson, 1987 and Foster and Shorrocks, 1988).

In practice, population income distributions are in general not observable, and so comparisons must be based on statistical tests that make use of distributions estimated from samples. Bootstrap tests that posit a null of *unrestricted* stochastic dominance of a given order appeared over the last two decades (e.g. Barrett and Donald, 2003, Horváth et al., 2006, and Linton et al., 2010), and all of them apply to testing for *restricted* stochastic dominance orderings[1]. Linton et al. (2010) (LSW, hereafter) proposed a bootstrap testing procedure that is asymptotically similar over a large set of distributions in the boundary of the null hypothesis. They allow for the case of dependent populations and show that their test is asymptotically valid uniformly over the distributions in the null hypothesis under certain regularity conditions. To obtain these properties of their test, LSW introduce consistent estimation of the "contact set", which is the subset of the domain on which the dominance functions coincide. This set is of great importance because it enters the asymptotic null distributions of conventional test statistics (e.g. one-sided Cramér-von-Mises and Kolmogorov-Smirnov test statistics). LSW construct bootstrap test statistics that adapt to the contact set using a consistent contact set estimator, whose use is similar to the generalized moment selection procedure in Andrews and Soares (2010). Although their test has significantly advanced the inference literature on stochastic dominance orderings, simulation-based evidence suggests that it is conser-

[1]Stochastic dominance orderings can either be unrestricted or restricted, as to whether the comparison of the distributions is carried out over the entire range of the union of their supports or only over some predesignated restricted range of it. The saliency of this distinction in poverty and income studies arises for a normative reason, being that unrestricted stochastic dominance orderings do not impose sufficient limits on the ranges over which certain ethical principles must be obeyed. Said differently, these orderings discriminate between the living standards of everyone below a survival poverty line when in fact this should not matter because everyone under that threshold should certainly be deemed to be in very difficult circumstances. In consequence, these orderings do not give equal ethical weight to all those who are below a survival poverty line. By contrast, the rankings of income distributions based on the restricted stochastic dominance orderings eliminates the concern with the precise living standards of the most deprived. See Bourguignon and Fields (1997) for more on this point.

vative on configurations in the boundary of the null hypothesis outside of the least favorable case. The present paper builds on the contribution of LSW by introducing a bootstrap test for restricted stochastic dominance that is asymptotically less conservative than their test on these configurations in the null hypothesis.

The bootstrap test this paper proposes retains the LSW testing procedure but replaces their contact set estimator with one based on a constrained estimator of the dominance functions. The contact set estimator this paper introduces estimates the dominance functions using the method of constrained empirical likelihood, where the constraints represent the restrictions of the null hypothesis. By contrast, the LSW contact set estimator uses the sample analogue estimator of the dominance functions. Therefore, our empirical likelihood estimation procedure defines a "constrained" estimator of the contact set. In effect, the test we propose is a modification of the LSW test that reformulates the LSW bootstrap test statistics in a data-dependent way that incorporates the statistical information contained in the constraints of the null hypothesis. We show that the test we propose is uniformly asymptotically valid.

The bootstrap test this paper proposes has a couple of advantages. First, it is asymptotically less conservative than the LSW test on configurations in the boundary of the null hypothesis outside of the least favorable case. This property is a consequence of the proposed contact set estimator being asymptotically a subset of its LSW counterpart under such population configurations. For population configurations in the least favorable case, the two contact set estimators are asymptotically equal. Accordingly, the LSW bootstrap test statistic is asymptotically no smaller than its modified counterpart, under the null. Second, we provide sufficient conditions under which the proposed test is asymptotically more powerful than the LSW test against fixed alternatives, and vice-versa. The proposed test is shown to be consistent against all alternatives, which is also a property that the LSW test shares. Furthermore, we adopt the framework of LSW to compare the asymptotic local power properties of the tests, who consider Pitman local alternative sequences. Interestingly, we find that this criterion does not discriminate between the tests, regardless of the Pitman local alternative directions.

These desirable properties, however, come at the expense of introducing an additional tuning parameter, and the use of a slightly more restrictive parameter space than the one LSW employ. The tuning parameter is a grid that approximates the predesignated range in the domain of the dominance functions. The grid indexes the points in this domain on which we wish to impose the restrictions of the null in the constrained empirical likelihood estimation problem. This estimation problem is easy to implement in practice using standard computing packages. The parameter space in this paper is a subset of the one that LSW employ because two conditions in addition to the ones LSW impose, define it. The first condition is sufficient for the existence and uniqueness of the solution in the constrained empirical likelihood estimation problem, and the existence of Lagrange multipliers via Karush-Kuhn-Tucker conditions. The second condition imposes a zero lower bound on the supports of the income distributions, which is not restrictive as incomes are non-negative.

To explore the finite-sample properties of the tests, we report Monte Carlo simulation results that compare the tests. The simulations use the experimental designs in Section 5 of LSW, which are for fixed data-generating processes. The simulation results show the proposed test has better Type I error properties. The data-generating processes under the alternative are in the directions of dominance and non-dominance, which allows us to gauge the differences in the powers of the tests that our theoretical findings suggest. Overall, the simulation results suggest that the proposed test weakly dominates the LSW test, and that there can be substantial differences in their performance.

Tests based on the procedure Barrett and Donald (2003) introduce also apply to testing for restricted stochastic dominance. They proposed a consistent bootstrap test, for the case of independent populations, where the bootstrap critical value is computed using a bootstrap data-generating process in the least favorable case of the null hypothesis. Similar works in this area include, for example, Horváth et al. (2006). In general, these types of testing procedures are too conservative, and have asymptotically exact size equal to the nominal level when the dominance functions are equal almost everywhere. Tests for restricted stochastic dominance are not new. Davidson and Duclos (2013) and Davidson (2009) propose asymptotic and bootstrap tests that posit instead a null of non-dominance. By contrast, our paper and the literature discussed earlier, have non-dominance as one of the configurations under the alternative. Therefore, these two approaches are not directly comparable, but they certainly do complement each other.

The rest of this paper is organized as follows. Section 2 defines the null hypothesis of restricted stochastic dominance, introduces the test statistic LSW utilize, their definition of a test having asymptotically exact size, and their bootstrap testing procedure. Section 3 introduces the the proposed contact set estimator and its asymptotic properties under the null. Section 4 presents the asymptotic properties of the proposed testing procedure. Section 5 presents a further discussion on the differences between the LSW test and the one this paper proposes. Section 6 reports the findings of preliminary Monte Carlo simulation experiments. Finally, Section 7 concludes and Section 8 collates the acknowledgements of the individuals and institutions who provided help during the research.

2 Setup

This section introduces the testing problem this paper focuses on, the test statistic and its pointwise-asymptotic null distribution. Furthermore, it presents the LSW contact set estimator and bootstrap testing procedure.

2.1 The Null Hypothesis

Consider two populations, A and B, with respective distribution functions P_A and P_B, and suppose that there is a joint distribution function, P, with marginal distributions P_A and P_B. It is important to account for the statistical dependence between the two populations in many applications, such as the comparison of distributions over time, or before and after an economic policy or event. Distribution B is said to dominate distribution A, stochastically at order $s \in \mathbb{Z}_+$ and over the range $[\underline{t}, \overline{t}] \subset \text{supp}\,(P_A) \cup \text{supp}\,(P_B)$, if

$$E_P\left[\frac{\left(t - X^B\right)^{s-1}}{(s-1)!} 1\left[X^B \leq t\right] - \frac{\left(t - X^A\right)^{s-1}}{(s-1)!} 1\left[X^A \leq t\right]\right] \leq 0 \quad \forall t \in [\underline{t}, \overline{t}], \tag{2.1}$$

where $\mathbf{X} = [X^A, X^B]$ is a random vector whose distribution is P, and supp (P_K) is the support of $P_K,\ K = A, B$. The unrestricted stochastic dominance orderings are defined as above, but with the equality: $[\underline{t}, \overline{t}] =$ supp $(P_A) \cup$ supp (P_B).

Let P_0 denote the "true" distribution of $\mathbf{X}$. Given $s \in \mathbb{Z}_+$ and $[\underline{t}, \overline{t}]$, we wish to test that P_0 satisfies the moment inequalities (2.1). For ease of exposition, let $\left\{\mathbf{x} \mapsto g(\mathbf{x}; t), t \in [\underline{t}, \overline{t}]\right\}$ denote the set of moment functions in (2.1). Implicit in this notation for the moment functions is the order of stochastic dominance, which is fixed by the null hypothesis. Let $\mathcal{P}$ denote the set of all potential continuous distributions of $\mathbf{X}$ that satisfies the following assumption.

Assumption 1. (i) $P\left[\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X}; t) < 0\right]$; (ii) supp $(P) \subseteq \mathbb{R}^2_+$; (iii) $\{\mathbf{X}_i\}_{i=1}^n$ is a random sample, and (iv) for some $\delta > 0$, $\sup_{P\in\mathcal{P}} E_P\left[|X^K|^{2((s-1)\vee 1)+\delta}\right] < +\infty$ for $K = A, B$.

An important consequence of Condition (ii) is that for each $s \in \mathbb{Z}_+$, the set of moment functions $\left\{\mathbf{x} \mapsto g(\mathbf{x}; t), t \in [\underline{t}, \overline{t}]\right\}$ is uniformly bounded. This property of $\left\{\mathbf{x} \mapsto g(\mathbf{x}; t), t \in [\underline{t}, \overline{t}]\right\}$ along with the fact that it is suitably measurable and Vapnik-Chervonenkis, drives the uniform asymptotic results in the present work.

The null hypothesis is

$$H_0: \ P_0 \in \mathcal{P}_0, \tag{2.2}$$

where

$$\mathcal{P}_0 = \left\{P \in \mathcal{P} : E_P\left[g(\mathbf{X}; t)\right] \leq 0 \quad \forall t \in [\underline{t}, \overline{t}]\right\}. \tag{2.3}$$

The alternative hypothesis is understood to be $H_1: \ P_0 \in \mathcal{P} - \mathcal{P}_0$.

Remark 2.1. LSW allows for X^A and X^B to depend on unknown parameters that are finite or infinite dimensional. This accommodation, albeit useful in certain applications, is not applicable to the comparison of income distributions. Conditional stochastic dominance orderings would be the appropriate approach to accommodate the effect of covariates on the comparison of income distributions, which is beyond the scope of this paper.

2.2 Test Statistic and Asymptotic Theory

The testing procedure this paper proposes has asymptotically exact size in the sense of Definition 1 of LSW, which we repeat here for convenience.

Definition 1. (i) A test φ_α with a nominal level α is said to have an *asymptotically exact size* if there exists a nonempty subset $\mathcal{P}_0' \subset \mathcal{P}_0$ such that:

$$\limsup_{n\to+\infty} \sup_{P\in\mathcal{P}_0} E_P\varphi_\alpha \leq \alpha, \quad \text{and} \tag{2.4}$$

$$\limsup_{n\to+\infty} \sup_{P\in\mathcal{P}_0'} |E_P\varphi_\alpha - \alpha| = 0. \tag{2.5}$$

(ii) When a test φ_α satisfies (2.5), we say that the test is *asymptotically similar* on $\mathcal{P}_0'$.

LSW use a Cramér-von-Mises type test statistic in a bootstrap testing procedure for H_0. In the setting of this paper it is given by

$$\hat{T}_n = n \int_{\underline{t}}^{\overline{t}} \max \left\{ E_{\hat{P}_n} \left[g\left(\mathbf{X}; t\right)\right], 0 \right\}^2 \, dt, \tag{2.6}$$

where $\hat{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{X}_i}$ is the empirical measure based on the random sample $\{\mathbf{X}_i\}_{i=1}^n$, and $E_{\hat{P}_n}$ denotes the expectation under $\hat{P}_n$. The asymptotic null distribution of $\hat{T}_n$ depends on the form of the contact set

$$\Delta\left(P\right) = \left\{ t \in [\underline{t}, \overline{t}] : E_P\left[g\left(\mathbf{X}; t\right)\right] = 0 \right\}, \tag{2.7}$$

for $P \in \mathcal{P}_0$, and follows from the Donsker property of the class of moment functions $\mathcal{F}$ with respect to $\mathcal{P}_0$ for each $s \in \mathbb{Z}_+$ and $[\underline{t}, \overline{t}]$. Let $\mathcal{P}_{00} = \left\{ P \in \mathcal{P}_0 : \int_{\Delta(P)} dt > 0 \right\}$. Then the pointwise asymptotic null distribution of $\hat{T}_n$ is given by

$$\hat{T}_n \xrightarrow{d} \begin{cases} \int_{\Delta(P)} \max\{\nu(t),0\}^2 \, dt, & \text{if } P \in \mathcal{P}_{00}, \\ 0, & \text{if } P \in \mathcal{P}_0 - \mathcal{P}_{00}, \end{cases} \tag{2.8}$$

where $\nu(\cdot)$ is a zero-mean Gaussian process on $[\underline{t}, \overline{t}]$ with a covariance kernel given by

$$C(t_1, t_2) = \text{Cov}_P\left(g\left(\mathbf{X}; t_1\right), g\left(\mathbf{X}; t_2\right)\right). \tag{2.9}$$

The limiting null distribution of $\hat{T}_n$ exhibits a discontinuity in the underlying probability P that generates the data. The consequence of this large sample behavior of the test statistic is that it invalidates the use of the canonical bootstrap for testing H_0 (e.g. see Andrews, 2000). For this reason, LSW propose a bootstrap testing procedure that uses a contact set estimator. LSW propose an estimator of $\Delta\left(P_0\right)$ based on the sample analogue estimator of the moments $E_{P_0}\left[g\left(\mathbf{X};\cdot\right)\right]$. Specifically, they estimate $\Delta\left(P_0\right)$ using

$$\hat{\Delta}_n = \left\{t \in [\underline{t}, \overline{t}] : \left|E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right| < r_n\right\}, \tag{2.10}$$

where $\{r_n\}_{n\geq 1}$ is a suitably chosen null sequence of positive (possibly random) numbers that satisfies $\sqrt{n} r_n \to +\infty$ as $n \to +\infty$.

The LSW bootstrap test procedure follows these steps:

1. Using the data, compute $\hat{T}_n$ and $\hat{P}_n$.

2. Generate B bootstrap samples each of size n, $\left\{\mathbf{X}^\star_{i,l}\right\}_{i=1}^n$ for $l = 1, \ldots, B$, using resampling with replacement from $\hat{P}_n$. That is, draw $\mathbf{X}^\star_{i,l}$ randomly with replacement from $\{\mathbf{X}_i\}_{i=1}^n$ according to $\hat{P}_n$ for $i = 1, \ldots, n$ and $l = 1, \ldots, B$.

3. For each bootstrap sample, compute the bootstrap test statistic as follows:

$$\hat{T}^{\star}_{n,l} = \begin{cases} \int_{\underline{t}}^{\overline{t}} \left(\max\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g\left(\mathbf{X}^{\star}_{i,l}; t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right], 0 \right\} \right)^2 dt, & \text{if } \int_{\hat{\Delta}_n} dt = 0, \\ \int_{\hat{\Delta}_n} \left(\max\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g\left(\mathbf{X}^{\star}_{i,l}; t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right], 0 \right\} \right)^2 dt, & \text{if } \int_{\hat{\Delta}_n} dt > 0, \end{cases} \tag{2.11}$$

where $\hat{\Delta}_n$ is defined in (2.10).

4. Compute the approximate bootstrap p-value $\hat{\Upsilon}_B = \frac{1}{B}\sum_{l=1}^{B} 1\left[\hat{T}^{\star}_{n,l} \geq \hat{T}_n\right]$.

5. Reject H_0 if $\hat{\Upsilon}_B \leq \alpha$, where $\alpha \in (0, 1/2)$ is a given nominal level.

LSW pay attention to the control of asymptotic rejection probabilities uniform in $P \in \mathcal{P}$. For this reason, they introduce a regularity condition on the asymptotic Gaussian process ν, which is given by Definition 2 of their paper. In the context of the present work, this condition is the following.

Definition 2. A Gaussian process ν is *regular* on $A \subset [\underline{t}, \overline{t}]$ if for any $\alpha \in (0, 1/2]$, there exists $\overline{\epsilon} > 0$ depending only on α such that

$$P\left[\int_A \max\{\nu(t), 0\}^2 \, dt < \overline{\epsilon}\right] < 1 - \alpha \tag{2.12}$$

and for any $c > 0$,

$$\limsup_{\eta \downarrow 0} \sup_{P \in \mathcal{P}_0} P\left[\left|\int_A \max\{\nu(t), 0\}^2 \, dt - c\right| \leq \eta\right] = 0. \tag{2.13}$$

See pages 190 and 191 of LSW for a discussion of this regularity condition.

The test procedure this paper proposes follows the steps of the LSW bootstrap test procedure, but with $\hat{\Delta}_n$ replaced by a different set estimator of $\Delta(P_0)$ when computing the bootstrap test statistics in the third step above. The LSW contact set estimator is based on the empirical measure $\hat{P}_n$. As the empirical measure is in fact the unrestricted empirical likelihood estimator of P_0, this paper proposes to replace it with the constrained empirical likelihood estimator of P_0 that imposes the moment inequality restrictions of $\mathcal{P}_0$, in estimation of the contact set. The next section introduces this procedure.

3 Contact Set Estimation

This section introduces the proposed contact set estimator and its large-sample properties under H_0. Let $\left\{\mathcal{T}_{N(n)}\right\}_{n\geq 1}$ be a given sequence of subsets of $[\underline{t},\overline{t}]$ with $|\mathcal{T}_{N(n)}| = N(n)$ $\forall n$ that converges to $[\underline{t},\overline{t}]$ in the Hausdorff metric as $n \to +\infty$. The proposed contact set estimator replaces $E_{\hat{P}_n}\left[g\left(\mathbf{X};\cdot\right)\right]$ with $E_{\acute{P}_n}\left[g\left(\mathbf{X};\cdot\right)\right]$ in the definition of $\hat{\Delta}_n$, where $\acute{P}_n = \sum_{i=1}^{n}\acute{p}_i\delta_{\mathbf{X}_i}$ with the probabilities $\acute{p}_1,\ldots,\acute{p}_n$ defined as the solution of the following optimization problem:

$$\max_{p_1,\ldots,p_n}\sum_{i=1}^{n}\log p_i \quad \text{subject to} \quad p_i \geq 0\ i=1,\ldots,n, \quad \sum_{i=1}^{n}p_i = 1, \quad \text{and}$$
$$\sum_{i=1}^{n}p_i g\left(\mathbf{X}_i;t\right) \leq 0 \quad \forall t \in \mathcal{T}_{N(n)}. \tag{3.1}$$

The estimator $\acute{P}_n$ is the approximate constrained empirical likelihood estimator of P_0, and we denote the contact set estimator based on it by $\acute{\Delta}_n$. The estimator $\hat{P}_n$ solves the above optimization problem, but without imposing the constraints (3.1); therefore, $E_{\hat{P}_n}\left[g\left(\mathbf{X};\cdot\right)\right]$ does not necessarily satisfy the restrictions of the null hypothesis. By contrast, from (3.1), the definition of $\acute{P}_n$ implies $E_{\acute{P}_n}\left[g\left(\mathbf{X};\cdot\right)\right]$ approximately satisfies the constraints (2.1) but with the approximation disappearing asymptotically.

Next, we characterize the set of probabilities in $\mathcal{P}_0$ under which $\acute{\Delta}_n$ is a uniformly consistent estimator of the contact set. Noting that the moment functions for each $s \in \mathbb{Z}_+$ are of the form

$$g(\mathbf{x};t) = h(x^B;t) - h(x^A;t) \tag{3.2}$$

where is $h(\cdot;t)$ weakly monotonic in its first argument for a given $t \in [\underline{t},\overline{t}]$, implies that they satisfy the following property.

Property 1. The class of functions $\mathcal{F} = \left\{\mathbf{x} \mapsto g(\mathbf{x};t),\ t \in [\underline{t}, \overline{t}]\right\}$ satisfies the following property. For each $\mathbf{x} \in \mathbb{R}^2_+$ either $g(\mathbf{x};t) \leq 0\, \forall t \in [\underline{t}, \overline{t}]$ or $g(\mathbf{x};t) \geq 0\, \forall t \in [\underline{t}, \overline{t}]$.

This property states that the sign of the moment functions g is determined by the configuration in its data dimension independently of t. For $P \in \mathcal{P}$, an important consequence of Condition (i) of Assumption 1 and Property 1 that we exploit is that the covariance kernel (2.9) satisfies

$$C(t,t') = E_P\left[g(\mathbf{X};t)\, g(\mathbf{X};t')\right] \geq 0 \quad \forall (t,t') \in \Delta(P) \times \Delta(P). \tag{3.3}$$

Definition 3. For each $[c_1, c_2] \in (0,1] \times (0,+\infty)$, let $\mathcal{P}_0(c_1,c_2)$ be the collection of probabilities in $\mathcal{P}_0$ under which (i) $P\left[\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X};t) < 0\right] \geq c_1$, and (ii) $\inf\limits_{t,t' \in \Delta(P)} E_P\left[g(\mathbf{X};t)\, g(\mathbf{X};t')\right] \geq c_2$.

Condition (i) of Definition 3 further restricts $\mathcal{P}_0$ by ruling out distributions that become arbitrarily close to distributions that place zero probability on the event $\left\{\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X};t) < 0\right\}$. It begets the uniform asymptotic existence of the probabilities $\acute{p}_1, \ldots, \acute{p}_n$ and Lagrange multipliers that solve Karush-Kuhn-Tucker conditions. Condition (ii) also restricts $\mathcal{P}_0$; it excludes distributions whose covariance kernel (3.3) is arbitrarily close to zero. This condition is useful for showing that the norm of the Lagrange multipliers converges to zero in probability, uniformly over $\mathcal{P}_0(c_1,c_2)$. For brevity, we relegate the formal statements and proofs of these technical intermediate results to Appendix B.

We have the following result.

Theorem 1. For each $[c_1, c_2] \in (0,1] \times (0,+\infty)$,

1. $\lim_{n\to+\infty} \sup_{P\in\mathcal{P}_0(c_1,c_2)} P\left[E_{\acute{P}_n}\left[g(\mathbf{X};t)\right] \leq E_{\hat{P}_n}\left[g(\mathbf{X};t)\right]\ \forall t \in [\underline{t},\overline{t}]\right] = 1$.
2. $\lim_{n\to+\infty} \sup_{P\in\mathcal{P}_0(c_1,c_2)} P\left[\Delta(P) \subset \acute{\Delta}_n \subset \hat{\Delta}_n\right] = 1$.
3. $\forall \epsilon > 0$, $\lim_{n\to+\infty} \sup_{P\in\mathcal{P}_0(c_1,c_2)} P\left[\hat{\Delta}_n \subset \left\{t \in [\underline{t},\overline{t}] : \left|E_P\left[g(\mathbf{X};t)\right]\right| \leq (1+\epsilon) r_n\right\}\right] = 1$.

Proof. See Appendix A.1. □

Remark 3.1. Since $\{(1+\epsilon)r_n\}_{n\geq 1}$ is a null sequence for each $\epsilon > 0$, the set

$$\left\{t \in [\underline{t}, \overline{t}] : |E_P [g(\mathbf{X}; t)]| \leq (1+\epsilon)r_n\right\},$$

is an $(1+\epsilon)r_n$-enlargement of the contact set that shrinks to the contact set as the sample size tends to infinity. Therefore, Parts 2 and 3 of Theorem 1 imply that the two contact set estimators of $\Delta(P_0)$ are equal in the limit, with uniformity. These results of this theorem drive the uniform asymptotic equivalence of the testing procedures under the null, which the next section presents.

Remark 3.2. Part 2 of Theorem 1 implies the LSW bootstrap test statistic, described in (2.11), weakly dominates its modified counterpart stochastically at the first-order, conditional on the sample $\{\mathbf{X}_i\}_{i=1}^n$, and under probabilities P in $\mathcal{P}_0(c_1, c_2)$ when n is large enough. In consequence, the proposed test is asymptotically less conservative than the LSW test. The next section presents the asymptotic properties of the proposed test.

Remark 3.3. The asymptotic property $P\left[\hat{\Delta}_n \subset \acute{\Delta}_n\right] \to 1$ as $n \to +\infty$, does *not* hold under probabilities P in the boundary of $\mathcal{P}_0(c_1, c_2)$ outside of the least favorable case[2]. That's because

$$\lim_{n\to+\infty} P\left[E_{\acute{P}_n}[g(\mathbf{X}; t)] < E_{\hat{P}_n}[g(\mathbf{X}; t)] \ \forall t \in [\underline{t}, \overline{t}]\right] > 0 \quad \text{and} \tag{3.4}$$

$$\lim_{n\to+\infty} P\left[E_{\hat{P}_n}[g(\mathbf{X}, t')] < -r_n\right] = 1, \quad \text{where} \quad E_P[g(\mathbf{X}; t')] < 0, \tag{3.5}$$

hold for such probabilities in $\mathcal{P}_0(c_1, c_2)$. The limit (3.5) and the continuity of P imply $\exists t_0 \in [\underline{t}, \overline{t}]$ such that $E_{\hat{P}_n}[g(\mathbf{X}; t_0)] = -r_n$ with probability approaching unity, as the sample size increases. Using this result, the limit (3.4) implies $\exists t_{00} \in \hat{\Delta}_n$ in a neighborhood of t_0 such that $t_{00} \notin \acute{\Delta}_n$ with positive probability as the sample size increases. Finally note that the large-sample property (3.4) follows from Lemma C.1 and because $\lim_{n\to+\infty} P\left[\acute{P}_n \neq \hat{P}_n\right] > 0$ holds for such probabilities, where

$$\acute{P}_n \neq \hat{P}_n \iff \sup_{t\in\mathcal{T}_{N(n)}} E_{\hat{P}_n}[g(\mathbf{X}, t)] > 0.$$

[2]Note that P is in the boundary of $\mathcal{P}_0(c_1, c_2)$ if $\exists t' \in [\underline{t}, \overline{t}]$ such that $E_P[g(\mathbf{X}; t')] < 0$.

4 Asymptotic Size and Power Properties

This section introduces the asymptotic size and power properties of the proposed test. We also characterize the set of probabilities under H_0 for which the proposed test has asymptotically exact size.

4.1 Asymptotic Size Properties

LSW impose regularity on r-enlargements of the contact sets,

$$B(r) = \left\{ t \in [\underline{t}, \bar{t}] : \left| E_P \left[g\left(\mathbf{X}; t\right) \right] \right| \leq r \right\},$$

to characterize the set of probabilities on which their bootstrap test has asymptotically exact size. This is done by introducing a regularity condition on the asymptotic Gaussian process in (2.8), given by Definition 2. The set of probabilities under which the proposed bootstrap test has asymptotically exact size is given by the following.

Definition 4. (i) For each $\epsilon > 0$ and $[c_1, c_2] \in (0, 1] \times (0, +\infty)$, let $\mathcal{P}_0(\epsilon, c_1, c_2)$ be the collection of probabilities in $\mathcal{P}_0(c_1, c_2)$ under which ν in (2.8) is regular on B_n for each $n \geq 1$, where

$$B_n = \begin{cases} B\left((1-\epsilon) r_n\right), & \text{if } \int_{B((1+\epsilon) r_n)} dt > 0, \text{ and} \\ [\underline{t}, \bar{t}], & \text{if } \int_{B((1+\epsilon) r_n)} dt = 0; \end{cases}$$

(ii) Given $\xi_n \to 0$, let $\mathcal{P}_{00}(\epsilon, c_1, c_2, \{\xi_n\})$ be the collection of probabilities in $\mathcal{P}_0(\epsilon, c_1, c_2)$ under which for each $n > 1/\epsilon$, ν in (2.8) is *regular* on $B\left(n^{-1/2}\xi_n\right)$,

$$\int_{B((1-\epsilon) r_n)} dt > 0 \quad \text{and} \quad \int_{B((1+\epsilon) r_n) - B\left(n^{-1/2}\xi_n\right)} dt \leq \xi_n. \tag{4.1}$$

Let $\left\{ \acute{T}^{\star}_{n,l} \right\}_{l=1}^{B}$ denote the bootstrap test statistics computed as above but with $\hat{\Delta}_n$ replaced by $\acute{\Delta}_n$, and let $\mathcal{A}_n$ denote the Borel sigma-algebra generated by the random sample $\{\mathbf{X}_i\}_{i=1}^{n}$. Also, let $\acute{\Upsilon}_B = \frac{1}{B}\sum_{l=1}^{B} 1\left[\acute{T}^{\star}_{n,l} \geq \hat{T}_n\right]$.

Theorem 2. Given $[c_1, c_2] \in (0,1] \times (0,+\infty)$ and $\epsilon > 0$, suppose that $P_0 \in \mathcal{P}_0(\epsilon, c_1, c_2)$. Then $\hat{\Upsilon}_B - \acute{\Upsilon}_B \xrightarrow{P} 0$ conditional on $\mathcal{A}_n$ uniformly in $\mathcal{P}_0(\epsilon, c_1, c_2)$.

Proof. See Appendix A.2. □

Theorem 2 establishes the asymptotic equivalence of the bootstrap test statistics $\hat{\Upsilon}_B$ and $\acute{\Upsilon}_B$, uniformly over $\mathcal{P}_0(\epsilon, c_1, c_2)$. Since for each $\epsilon > 0$ and $[c_1, c_2] \in (0,1] \times (0,+\infty)$ the LSW test has *asymptotically exact size*, in the sense of Definition 1, uniformly over a superset of $\mathcal{P}_0(\epsilon, c_1, c_2)$, it also has this property over $\mathcal{P}_0(\epsilon, c_1, c_2)$. Consequently, Theorem 2 implies that the proposed test inherits the uniform asymptotic properties of the LSW test for probabilities in $\mathcal{P}_0(\epsilon, c_1, c_2)$. By applying Theorem 2 of LSW in the setup of our paper, these properties are

1. for each $\epsilon > 0$ and $[c_1, c_2] \in (0,1] \times (0,+\infty)$,

$$\limsup_{n\to+\infty} \sup_{P\in\mathcal{P}_0(\epsilon,c_1,c_2)} P\left[\hat{\Upsilon}_\infty \leq \alpha\right] \leq \alpha, \quad and \tag{4.2}$$

2. for each decreasing sequence $\xi_n \to 0$, $\epsilon > 0$ and $[c_1, c_2] \in (0,1] \times (0,+\infty)$,

$$\limsup_{n\to+\infty} \sup_{P\in\mathcal{P}_{00}(\epsilon,c_1,c_2,\{\xi_n\})} \left|P\left[\hat{\Upsilon}_\infty \leq \alpha\right] - \alpha\right| = 0. \tag{4.3}$$

Therefore, Theorem 2 implies that the above statements in (4.2) and (4.3) hold with $\hat{\Upsilon}_\infty$ replaced by $\acute{\Upsilon}_\infty$.

4.2 Asymptotic Power Properties

This section introduces the power properties of the proposed modification of the LSW bootstrap test. The first result concerns consistency of the test.

Theorem 3. Suppose that we are under a fixed alternative $P \in \mathcal{P} - \mathcal{P}_0$ such that

$$\int_{\underline{t}}^{\overline{t}} \max\left\{E_P\left[g\left(\mathbf{X};t\right)\right],0\right\}^2\,dt > 0. \tag{4.4}$$

Then, $\lim_{n\to+\infty} P\left[\acute{\Upsilon}_B \leq \alpha\right] = 1$.

Proof. See Appendix A.3. □

Therefore, the proposed test is consistent against all alternatives. This property is also shared by the LSW test. Recall that $P \in \mathcal{P}-\mathcal{P}_0$ also satisfies $P\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right) < 0\right] > 0$ by Assumption 1. In fact, the LSW test is also consistent when $P\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right) < 0\right] = 0$ holds under the alternative hypothesis. Under $P\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right) < 0\right] = 0$, the asymptotic existence and uniqueness of the probabilities $\acute{p}_1,\ldots,\acute{p}_n$ is no longer guaranteed. In consequence, the proposed contact set estimator $\acute{\Delta}_n$ does not exist with positive probability in large sample sizes under the alternative hypothesis. As already mentioned, we defer a detailed discussion of this condition to Section 5.

The next result develops sufficient conditions on $P \in \mathcal{P} - \mathcal{P}_0$ that imply the superiority of one testing procedure over the other in terms of asymptotic power. First, we introduce the following notation: $\hat{\Delta}_n^c = [\underline{t},\overline{t}] - \hat{\Delta}_n$, $\acute{\Delta}_n^c = [\underline{t},\overline{t}] - \acute{\Delta}_n$, and $\hat{\sigma}^2(t) = E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]^2 - \left(E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]\right)^2$.

Theorem 4. Suppose that the conditions of Theorem 3 hold. The following statements hold.

1. If $\lim_{n\to+\infty} P\left[\inf_{t\in\hat{\Delta}_n\cap\acute{\Delta}_n^c} \hat{\sigma}^2(t) \geq \sup_{t\in\acute{\Delta}_n\cap\hat{\Delta}_n^c} \hat{\sigma}^2(t)\right] = 1$, then $P\left[\acute{\Upsilon}_B \leq \alpha\right] \geq P\left[\hat{\Upsilon}_B \leq \alpha\right]$ for large enough n.
2. If $\lim_{n\to+\infty} P\left[\inf_{t\in\acute{\Delta}_n\cap\hat{\Delta}_n^c} \hat{\sigma}^2(t) \geq \sup_{t\in\hat{\Delta}_n\cap\acute{\Delta}_n^c} \hat{\sigma}^2(t)\right] = 1$, then $P\left[\hat{\Upsilon}_B \leq \alpha\right] \geq P\left[\acute{\Upsilon}_B \leq \alpha\right]$ for large enough n.

Proof. See Appendix A.4. □

The significance of the sample variance function $\hat{\sigma}^2(\cdot)$ in the statement of Theorem 4 is that under the bootstrap measure, it is equal to the variance functions of the bootstrap empirical processes $\left\{\mathbb{G}_n(t), t\in\acute{\Delta}_n\right\}$ and $\left\{\mathbb{G}_n(t), t\in\hat{\Delta}_n\right\}$, where $\mathbb{G}_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[g\left(\mathbf{X}_{i,l}^{\star};t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]\right]$.

Because these bootstrap empirical process converges weakly to a zero-mean Gaussian process, conditional on $\mathcal{A}_n$, the limit condition in parts 1 and 2 of Theorem 4 imply for large n that

$$\inf_{t\in\hat{\Delta}_n} \mathbb{G}_n(t) \geq \sup_{t\in\acute{\Delta}_n} \mathbb{G}_n(t) \quad \text{and} \quad \inf_{t\in\acute{\Delta}_n} \mathbb{G}_n(t) \geq \sup_{t\in\hat{\Delta}_n} \mathbb{G}_n(t), \tag{4.5}$$

respectively, with probability approaching unity conditional on $\mathcal{A}_n$. The results of Theorem 4 follow immediately from the conditions in (4.5).

Next, we focus on the asymptotic local power property of the test. Following LSW, we consider a sequence of probabilities $\{P_n\}_{n\geq 1} \subset \mathcal{P} - \mathcal{P}_0$ such that

$$E_{P_n}\left[g\left(\mathbf{X};t\right)\right] = H(t) + \delta(t)/\sqrt{n} \quad \text{and} \tag{4.6}$$

$$\sigma^2_{P_n}(t) = E_{P_n}\left[g^2\left(\mathbf{X};t\right)\right] - \left(E_{P_n}\left[g\left(\mathbf{X};t\right)\right]\right)^2, \tag{4.7}$$

where the functions $H(t)$ and $\delta(t)$ satisfy the following conditions.

Assumption 2. (i) $\int_C dt > 0$, where $C = \left\{t \in [\underline{t}, \overline{t}] : H(t) = 0\right\}$. (ii) $\sup_{t\in[\underline{t},\overline{t}]} H(t) \leq 0$. (iii) $\int_C \max\{\delta(t), 0\}^2\, dt > 0$. (iv) $\inf_{t\in[\underline{t},\overline{t}],n\in\mathbb{N}} \sigma^2_{P_n}(t) > 0$. (v) $P_n\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right) < 0\right] > 0$ for each n.

Except for parts (iv) and (v), Assumption 2 is identical to Assumption 5 of LSW. Therefore, the sequence $\{P_n\}_{n\geq 1}$ represents local alternatives that converge to the boundary points $\mathcal{P}_{00}$ at the $\sqrt{n}$ rate in the direction $\delta(t)$. Part (iv) ensures the valid use of the Weak Law of Large Numbers and the Central Limit Theorem for triangular arrays of row-wise IID random variables. Part (v) implies the asymptotic existence and uniqueness of $\acute{P}_n$, and that it is characterizable in terms of Lagrange multipliers using Karush-Kuhn-Tucker conditions.

The bootstrap test procedure this paper proposes has the same asymptotic local power as the LSW bootstrap test.

Theorem 5. Under the local alternatives $\{P_n\}_{n\geq 1} \subset \mathcal{P} - \mathcal{P}_0$ satisfying the conditions in (4.6), (4.7), and Assumption 2,

$$\lim_{n\to+\infty} P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] = \lim_{n\to+\infty} P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right].$$

Proof. See Appendix A.5. □

A remarkable point concerning the result of Theorem 5 is that it holds for all of the Pitman local alternative directions. Thus, this asymptotic criterion fails to discriminate between the tests. The reason is that the contact set estimators $\acute{\Delta}_n$ and $\hat{\Delta}_n$ are asymptotically equal under the sequences of local alternatives, and they have positive Lebesgue measure[3]. This behavior of the contact set estimators forces the bootstrap test statistics of the tests to be asymptotically equal.

5 Further Discussion

This section discusses the differences between the LSW test and the one this paper proposes. Section 5.1 discusses the relationship between the tests' parameter spaces. Section 5.2 discusses the relationship between the sets of probabilities on which the tests have asymptotically correct size, and asymptotic similarity on the boundary of the null hypothesis. Another difference between the tests is a procedural one. The test this paper proposes follows the LSW bootstrap testing procedure, but replaces their contact set estimator with one that employs a constrained estimator of the moments in (2.1). For a given sample size, this constrained estimator approximately imposes the restrictions of the null hypothesis (2.1) by imposing them on a grid $\mathcal{T}_{N(n)}$ as in (3.1), with the approximation disappearing asymptotically because $\left\{\mathcal{T}_{N(n)}\right\}_{n\geq 1}$ converges to $[\underline{t}, \overline{t}]$. Section 5.3 presents some intuition behind the choice of the sequence of grids $\left\{\mathcal{T}_{N(n)}\right\}_{n\geq 1}$ under the null and alternative hypotheses. Finally, Section 5.4 presents a modification of the constraints (3.1) that can increase the numerical accuracy of the solution in the constrained empirical likelihood optimization problem.

5.1 The Parameter Space $\mathcal{P}$

Recall that we denote the parameter space this paper employs by $\mathcal{P}$, which is the set of all the potential continuous distributions of $\mathbf{X}$ that satisfy the conditions of Assumption 1. In the con-

[3]See Lemmas C.4 and C.5.

text of the present work, the parameter space LSW employ is the set of all potential continuous distributions of $\mathbf{X}$ that satisfy Conditions (iii) and (iv) of Assumption 1. Therefore, $\mathcal{P}$ is a subset of the parameter space that LSW employ. As incomes are non-negative, imposing Condition (ii) of Assumption 1 in the definition of the parameter space, is only natural. It is Condition (i) of this assumption that begets the results of the present work. Specifically, Condition (i) is that it is sufficient for the asymptotic existence and uniqueness of the constrained empirical likelihood estimator $\acute{P}_n$ and the asymptotic existence of the Lagrange multipliers that solve Karush-Kuhn-Tucker conditions; See Appendix B for the technical details under the null hypothesis.

Recall that Condition (i) of Assumption 1 is $P\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right)<0\right]>0$. Given $[\underline{t},\overline{t}]$ and $s\in\mathbb{Z}_+$, the event $\left\{\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right)<0\right\}$ is given by

$$\left\{\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right)<0\right\}=\begin{cases}\left\{X^A<\underline{t},X^B>\overline{t}\right\}, & \text{if } s=1, \text{ and}\\ \left\{X^A<\underline{t},X^B>\overline{t}\right\}\cup\left\{X^A<\underline{t},X^A<X^B\leq\overline{t}\right\}, & \text{if } s\geq 2.\end{cases} \tag{5.1}$$

As this paper and LSW focus on continuous distributions, in the case $s=1$, Condition (i) excludes probabilities from the LSW parameter space that satisfy

$$P\left[X^A<\underline{t},X^B>\overline{t}\right]=0. \tag{5.2}$$

Condition (5.2) does not hold for probabilities such that supp$(P)=\mathbb{R}^2_+$ because $[\underline{t},\overline{t}]$ is subset of $\mathbb{R}_+$. For this reason, the parameter space $\mathcal{P}$ includes probabilities such that supp$(P)=\mathbb{R}^2_+$. However, condition (5.2) holds for compactly supported probabilities that satisfy

$$\underline{t}\leq\inf\left\{\text{supp}\left(P_A\right)\right\} \quad \text{and} \quad \overline{t}\geq\sup\left\{\text{supp}\left(P_B\right)\right\}. \tag{5.3}$$

Therefore, $\mathcal{P}$ excludes such probabilities. Similarly, in the case $s\geq 2$, Condition (i) does not exclude probabilities supported on $\mathbb{R}^2_+$. Furthermore, it does not exclude compactly supported probabilities that satisfy (5.3), provided that they also satisfy $P\left[X^A < \underline{t}, X^A < X^B \leq \overline{t}\right] > 0.$

A natural question to raise at this point is whether it is possible to extend the results in the text to the case where the parameter space $\mathcal{P}$ includes probabilities such that $P\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right) < 0\right] = 0$. The answer is yes, provided that the feasible set in the constrained empirical likelihood problem i.e. the random set

$$\left\{p_i, i = 1\ldots n : p_i > 0\ \forall i,\ \sum_{i=1}^{n} p_i = 1,\ \text{and}\ \sum_{i=1}^{n} p_i g\left(\mathbf{X}_i;t\right) \leq 0\ \forall t \in [\underline{t},\overline{t}]\right\}, \tag{5.4}$$

is asymptotically non-empty with probability tending to unity under P. For this asymptotic condition to hold, P must also satisfy the stronger support condition

$$P\left[g\left(\mathbf{X};t\right) = 0 \quad \forall t \in [\underline{t},\overline{t}]\right] = 1. \tag{5.5}$$

The reason is that $P\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right) < 0\right] = 0$ implies

$$P\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right) \geq 0\right] = 1 \iff P\left[g\left(\mathbf{X};t\right) \geq 0 \quad \forall t \in [\underline{t},\overline{t}]\right] = 1, \tag{5.6}$$

where the equivalence follows from Property 1 of the moment functions. These probabilities are degenerate in the sense that they place unit mass on the zero function in the set of realizations $\{t \mapsto g(\mathbf{x},t), \mathbf{x} \in \text{supp}(P)\}$.

5.2 Asymptotically Correct Size and Asymptotic Similarity

The sets of probabilities on which the proposed test has asymptotically correct size are defined in Part (i) of Definition 4. And the sets of probabilities on which the proposed test is asymptotically similar on the boundary of the null hypothesis are defined in Part (ii) of Definition 4. These sets are subsets of their LSW counterparts because two conditions, in addition to the ones LSW impose, define them. For each $[c_1, c_2] \in (0,1] \times (0,+\infty)$, these conditions define the set of probabilities $\mathcal{P}_0(c_1,c_2)$, which is given by Definition 3.

5.3 Choice of The Grid Sequence $\left\{\mathcal{T}_{N(n)}\right\}_{n\geq 1}$

The test procedure this paper proposes introduces a fine tuning parameter. This parameter is the sequence of grids $\left\{\mathcal{T}_{N(n)}\right\}_{n\geq 1}$, which converges to the interval $[\underline{t},\overline{t}]$ in the Hausdorff metric. The results in the previous sections do not depend on any specific choice of this parameter or on its rate of convergence. All that is required for the results to hold is that $\left\{\mathcal{T}_{N(n)}\right\}_{n\geq 1}$ converges to $[\underline{t},\overline{t}]$, which is because these results are asymptotic. However, the choice of $\left\{\mathcal{T}_{N(n)}\right\}_{n\geq 1}$ can significantly impact the finite-sample performance of the test this paper proposes. Thus, the question of the optimal choice of $\left\{\mathcal{T}_{N(n)}\right\}_{n\geq 1}$ is an important one, but it is beyond the scope of the present work. Instead, we present some intuition behind its choice from the perspectives of the null and alternative hypothesis, which illustrates a trade-off between test size and power. We also provide simulation-based evidence of this trade-off in Section 6.

Under H_0 and a large enough sample size, part 1 of Theorem 1 establishes that the estimator $\left\{E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right], t\in[\underline{t},\overline{t}]\right\}$ is biased downwards, uniformly over $\mathcal{P}_0(c_1,c_2)$. Generally, decreasing the Hausdorff distance between $[\underline{t},\overline{t}]$ and $\mathcal{T}_{N(n)}$ weakly exacerbates this bias, uniformly in $\mathcal{P}_0(c_1,c_2)$. In consequence, the contact set estimator $\acute{\Delta}_n$ based on a fine grid is a subset of its counterpart that uses a coarser grid, uniformly in $\mathcal{P}_0(c_1,c_2)$. Thus, an increase in the test's rejection probability can be achieved under the null by using a very fine grid, which can improve its finite-sample type 1 error rate under probability configurations outside of the least favorable case, whose contact sets have small (but positive) Lebesgue measure. Section 6 illustrates this behavior of the test using a Monte Carlo experiment.

By contrast, in the directions under H_1 such that $\inf_{t\in[\underline{t},\overline{t}]} E_P\left[g\left(\mathbf{X};t\right)\right] > 0$, the asymptotic power of the test decreases as the Hausdorff distance between $[\underline{t},\overline{t}]$ and $\mathcal{T}_{N(n)}$ shrinks to zero. We briefly explain why this is the case. Because this class of directions under the alternative satisfies

$$\lim_{n\to+\infty} P\left[E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right] > r_n \quad \forall t\in[\underline{t},\overline{t}]\right] = 1, \tag{5.7}$$

the LSW contact set estimator $\hat{\Delta}_n$ is asymptotically equal to the empty set, which implies that the bootstrap test statistics (2.11) in the LSW testing procedure are asymptotically equal to

$$\int_{\underline{t}}^{\overline{t}} \left(\max \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g\left(\mathbf{X}_{i,l}^{\star}; t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right], 0 \right\} \right)^2 dt. \tag{5.8}$$

Hence, the bootstrap test statistic (5.8) stochastically dominates the one this paper proposes at the first-order conditional on $\mathcal{A}_n$, because $\acute{\Delta}_n \subset [\underline{t}, \overline{t}]$ holds with probability approaching unity as n tends to infinity. Furthermore, the limit (5.7) implies that $\acute{\Delta}_n$ equals $[\underline{t}, \overline{t}]$ in the limit as the Hausdorff distance between $[\underline{t}, \overline{t}]$ and $\mathcal{T}_{N(n)}$ tends to zero. Thus, the asymptotic power of the proposed test must decrease as this Hausdorff distance tends to zero, because the two test statistics are equal in the limit.

5.4 Constraints in the Empirical Likelihood Estimation Problem

For $P \in \mathcal{P}$, the solution of the constrained empirical likelihood estimation procedure, introduced in Section 3, satisfies the following property

$$E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \leq E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \quad \forall t \in [\underline{t}, \overline{t}], \tag{5.9}$$

for large enough n with probability tending to unity. Therefore, the inequalities (5.9) can be used to increase the numerical accuracy and speed of the constrained empirical likelihood estimation procedure by replacing the constraints (3.1) with the following:

$$\sum_{i=1}^{n} p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t \in \mathcal{T}_{N(n)} \cap \left\{ t \in [\underline{t}, \overline{t}] : E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \geq 0 \right\} \quad \text{and} \tag{5.10}$$

$$\sum_{i=1}^{n} p_i g\left(\mathbf{X}_i; t\right) \leq E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \quad \forall t \in \mathcal{T}_{N(n)} \cap \left\{ t \in [\underline{t}, \overline{t}] : E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] < 0 \right\}, \tag{5.11}$$

This replacement shrinks the domain of the probabilities $p_1, \ldots, p_n$ over which the optimization routine searches for the solution $\acute{p}_1, \ldots, \acute{p}_n$.

6 Monte Carlo Experiments

This section reports the results of Monte Carlo experiments that compares the performance of the LSW test with the one this paper proposes. The experimental setup is the same as the one in Section 5 of LSW who focus on testing for first-order stochastic dominance. They construct data generating-processes using continuous uniform random variables on the unit interval. Because the first-order stochastic dominance ordering is invariant under positive monotonic transformations of the income variable, it is without loss of generality that we use their setup to study the behavior of the tests for income distributions. We find the test this paper proposes is noticeably less conservative for probabilities in the boundary of the null hypothesis outside of the least favourable case, and has higher power against directions in the alternative of dominance and non-dominance.

In each simulation experiment, the nominal level was fixed at 5%, $r_n(t) = \hat{\sigma}_t \sqrt{\frac{\log n}{n}}$, where $\hat{\sigma}_t^2 = E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]^2 - \left(E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]\right)^2$ and $t \in [\underline{t}, \overline{t}]$. This choice for the sequence $\{r_n\}_{n\geq 1}$ is known as the BIC choice. An alternative choice, which LSW use, is the one based on the Law of the Iterated Logarithm; it sets $r_n = an^{-1/2}\log\log n$, which is a constant function of $t \in [\underline{t}, \overline{t}]$, where a is a given constant. Presently, there isn't a theoretical reason to prefer one choice over the other. Instead, the moment inequality inference literature has relied on simulation-based evidence in proposing a choice for r_n. Andrews and Soares (2010) suggest the BIC choice for r_n, and we follow their lead. Unlike the Law of the Iterated Logarithm choice, the BIC choice does not depend on any additional fine-tuning parameters, which makes it more practical for practitioners. Furthermore, the BIC choice sets r_n as a function of $t \in [\underline{t}, \overline{t}]$ through $\hat{\sigma}_t$.

We set $\underline{t} = 0.05$ and $\overline{t} = 0.95$, and constructed the grid as follows:

$$\mathcal{T}_{N(n)} = \left\{\underline{t} = t_1 < t_2 < \cdots < t_{N(n)} = \overline{t}\right\}, \text{ where } t_{i+1} = t_i + \frac{\left(\overline{t} - \underline{t}\right)}{\lfloor \gamma\sqrt{n} \rfloor}, \tag{6.1}$$

for $i = 1, \ldots, N(n) - 1$, where $N(n) = \lfloor \gamma\sqrt{n} \rfloor + 1$ and $\gamma \in \{0.25, 1, 2\}$. Finally, the number of Monte Carlo replications was set to be 10000 in each simulation experiment, and the number of bootstrap replications was 199.

6.1 Simulation Under H_0

We compare the type I error rate properties of the our test and LSW test. LSW use the following generating process under the null. Let U_1 and U_2 be $U(0,1)$ random variables. Then define $X^B = U_1$ and $X^A = c_0^{-1}(U_2 - a_0)1\,[0 < U_2 \leq x_0] + U_2 1\,[x_0 < U_2 < 1]$, where $c_0 = (x_0 - a_0)/x_0 \in (0,1)$ and $x_0 \in (0,1)$. In this setup, the inequalities (2.1) hold for each $s \in \mathbb{Z}_+$, and we examine the case $s = 1$. The cumulative distribution function (CDF) of X^A has a "kink" at $X^A = x_0$ and the slope of the CDF changes from c_0 to 1 at the kink point x_0. See Figure 2 in LSW for a graphical representation of these CDFs.

In the simulations, we took $x_0 \in \{0, 0.1, 0.2, \ldots, 0.9\}$ and $c_0 \in \{0.2, 0.4, 0.6, 0.8\}$. The sample sizes we considered are $n = 500, 1000$. The case $x_0 = 0$ corresponds to the least favorable case as the CDFs of X^A and X^B are equal to the CDF of U_1. For a given $c_0 > 0$, the contact set gets smaller as x_0 increases; therefore, the data-generating process (DGP) moves away from the least favorable case toward the interior of the null. For each of these DGPs, the two CDFs coincide on a set of positive Lebesgue measure. Therefore, Theorem 2 of LSW establishes that their bootstrap test has an asymptotic size exactly equal to the nominal level under these DGPs. And Theorem 2 of this paper implies that the same result holds for the test this paper proposes.

The results are reported in Figure 1, which present the empirical rejection frequencies along with their pointwise 95% confidence intervals. For each value of c_0 we considered, the discrepancy between the performances of our method and the LSW test is not much for x_0 close to zero i.e. the least favorable case. However, as x_0 increases i.e. the contact set get smaller, the rejection probabilities under our test are statistically closer to the 5% nominal level than the ones based on the LSW test. These results suggest the bias of the LSW test is larger than the one this paper proposes. Furthermore, the simulations suggest that finer grids can attenuate the bias of the test this paper proposes on DGPs in the boundary of the null hypothesis that are outside of the least favorable case.

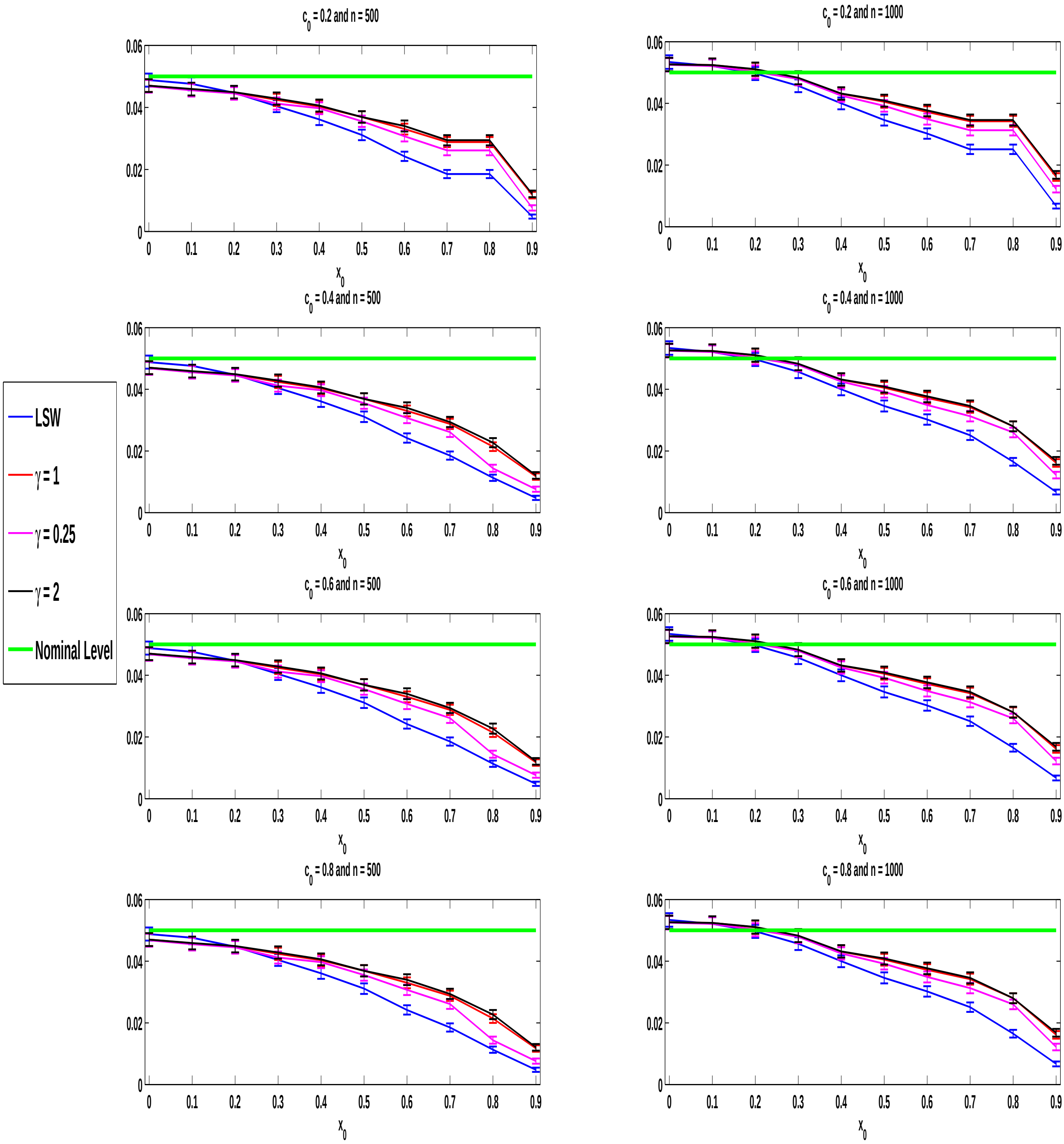

Figure 1: The empirical rejection probabilities under the null.

6.2 Simulation Under H_1 : Directions of Dominance

Let us now focus on the power properties of the two methods against directions of dominance under the alternative hypothesis. We consider DGPs as in Section 6.1 except that now X^A and X^B have exchanged roles in the numerical experiment; that is, $X^A = U_1$ and $X^B = c_0^{-1}(U_2 - a_0)1\left[0 < U_2 \leq x_0\right] + U_2 1\left[x_0 < U_2 < 1\right]$. This construction yields DGPs in the alternative where X^A dominates X^B at the first-order over the range $[0.05, 0.95]$. Furthermore, we considered the same values for c_0 as in the previous section, and the values for $x_0 \in \{0, 0.05, 0.1, 0.15, 0.2, 0.25\}$. For a given value of x_0, the DGP moves closer to the null model as c_0 increases. Therefore, we can expect that it will be harder for the tests to detect alternatives where c_0 is large and x_0 is small, when the sample size isn't large. Additionally, we can expect the tests to more easily detect alternatives where c_0 is small and x_0 large.

The simulation results are reported in Figure 2, which present the empirical rejection frequencies along with their pointwise 95% confidence intervals. Overall, these results indicate that the performance of the proposed test is no worse than that of the LSW test. For each sample size, c_0, and x_0 such that $x_0 < 0.15$ and $x_0 > 0.2$, the tests behave similarly. However, For values of $x_0 \in \{0.15, 0.2\}$, the proposed test statistically has strictly larger empirical power than that of the LSW test, and this difference in empirical power can also be quite large. For example, at $x_0 = 0.15$, $c_0 = 0.8$, and $n = 1000$ there is a difference of approximately 55% in empirical power for all of the selected values of γ.

At $n = 500$, the results indicate that the proposed test's performance is sensitive to the choice of γ. For example, at $c_0 = 0.4$, the simulation results indicate that the test with $\gamma = 2$ outperforms the rest of the tests, and at $c_0 = 0.6$, the proposed test with $\gamma = 1$ is the best. Furthermore, these differences in empirical power can be quite large: at $x_0 = 0.15$, the test with the highest power is statistically more powerful than the next best test by approximately 40%. However, this sensitivity vanishes when $n = 1000$. These results suggest that the choice of the grid $\mathcal{T}_{N(n)}$ is important in moderate sample sizes.

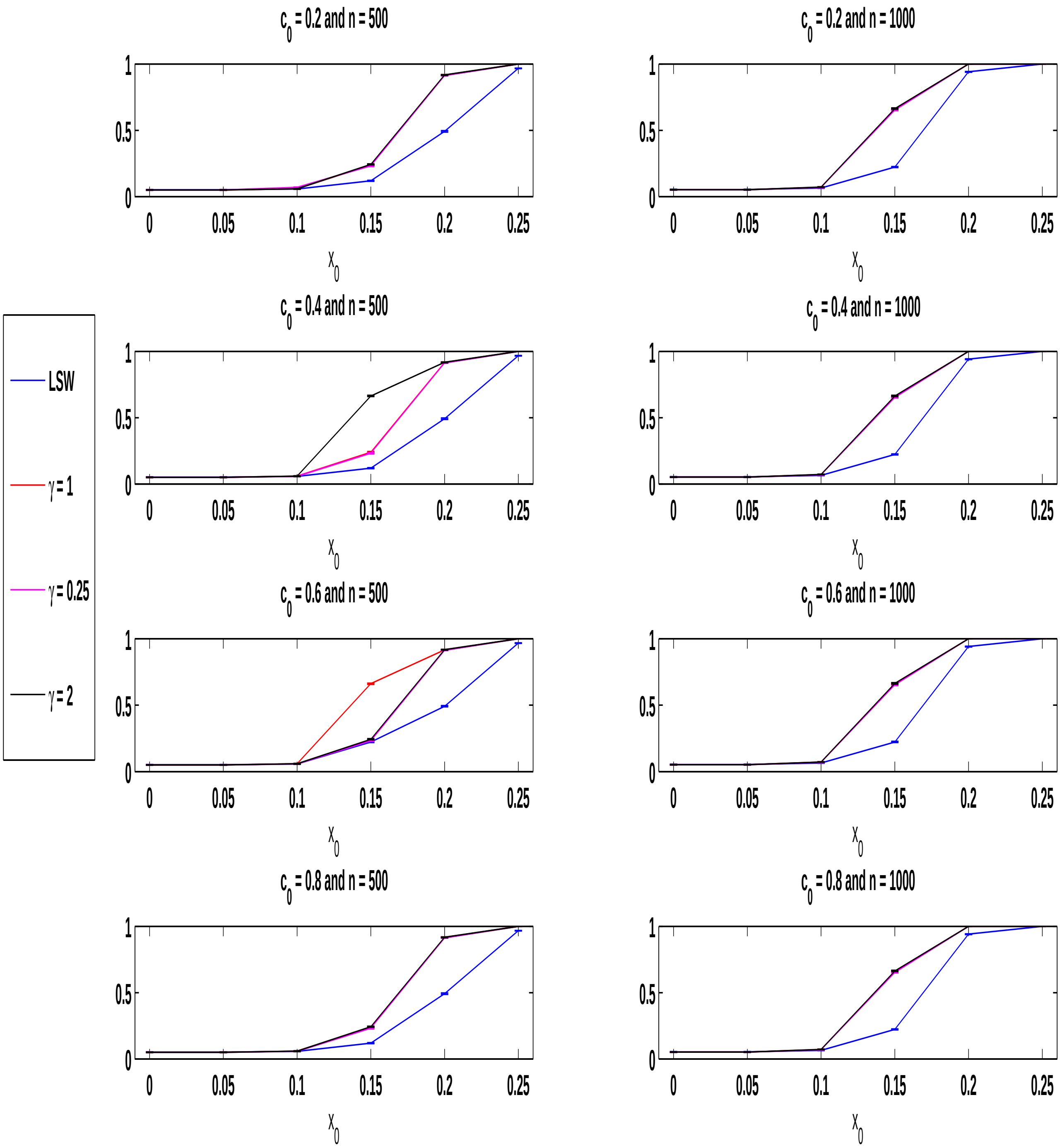

Figure 2: The empirical rejection probabilities under the alternative: directions of dominance.

6.3 Simulation Under H_1 : Directions of Non-Dominance

Let us now focus on the power properties of the two methods against directions of non-dominance. Directions of non-dominance in the alternative hypothesis have stochastic dominance conditions with some positive elements and some elements that are negative. Consider the following configuration of DGPs from LSW. Set $X^A \sim U[0,1]$. Then define

$$X^B = (U - a_0 b_1)\, 1\,[a_0 b_1 \leq U \leq x_0] + (U + a_0 b_2)\, 1\,[x_0 < U \leq 1 - a_0 b_2] \tag{6.2}$$

for $a_0 \in (0,1)$, where $U \sim U[0,1]$. As a_0 becomes closer to zero, the distribution of X^B becomes closer to the uniform distribution. The scale a_0 plays the role of the "distance" P_0 is from H_0. When a_0 is large, P_0 is farther from H_0, and when $a_0 = 0$, X^A and X^B have the same distribution which means P_0 belongs to the model of the null hypothesis under the least favorable configuration. For a graphical depiction of the CDFs of X^A and X^B, see Figure 4 in LSW. We set $(b_1, b_2, x_0) = (0.1, 0.5, 0.15)$ and $a_0 \in \{0, 0.05, 0.1, 0.15, 0.2, \ldots, 0.75\}$. The configurations for which $a_0 \neq 0$ correspond to alternative DGPs for which there are some non-violated inequalities for the case of $s = 1$ in the moments (2.1). We considered the following sample sizes $n = 256, 512, 1024, 2048$, and set X^A and the uniform random variable in the definition of X^B to be negatively correlated, with a correlation coefficient of -0.5.

The simulation results are reported in Figures 3, which present the empirical rejection frequencies along with their pointwise 95% confidence intervals. For each sample size and for a_0 sufficiently large, there is no difference between the two tests, which is expected since both tests are consistent. For moderate values of a_0, our test has statistically higher power than the LSW test, and the power differences can be large. For $n = 2048$, our test dominates the LSW test, and quite significantly when $a_0 = 0.1$ with a difference of approximately 34%. These simulation results suggest that the proposed test can better detect DGPs in H_1 that are "close" to H_0, when the sample size is large enough. However, a theoretical result is required for a more concrete conclusion. Overall, the simulation results show that our method performs better than the LSW test.

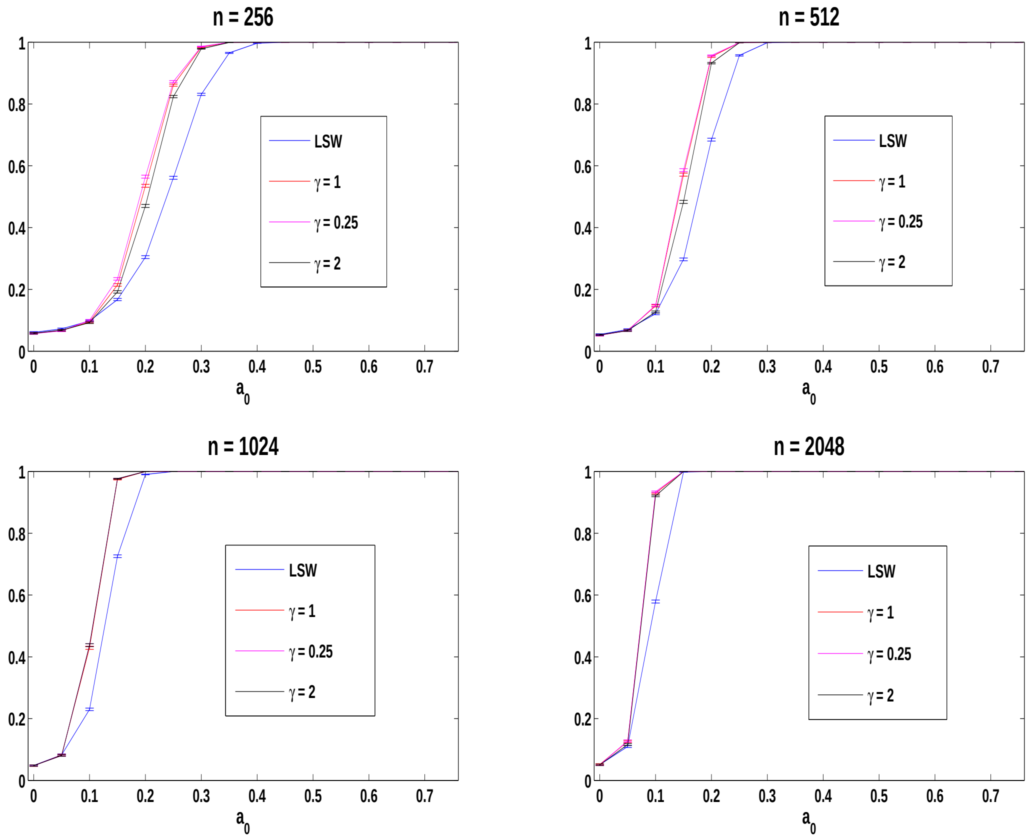

Figure 3: The empirical rejection probabilities under the alternative: directions of non-dominance.

7 Conclusion

This paper proposes a new method of testing for restricted stochastic dominance which is based on the test of LSW. The proposed testing procedure replaces the contact set estimator LSW use with the one based on the method of empirical likelihood. This method of estimation for the contact set incorporates the statistical information from imposing the restrictions of the null hypothesis in the estimation of the contact set, and alters the finite-sample distribution of the bootstrap test statistics in a data-dependent way. The proposed and LSW tests are uniformly asymptotically equivalent, and the former is asymptotically less conservative than it on the boundary of the null outside of the least favorable configuration. Under the alternative, the proposed test is consistent and has the same asymptotic local power properties of the LSW test. These properties of the proposed test come at the expense of working with a parameter space that is slightly more restrictive than the one LSW utilize, which is required for justifying the existence of the proposed contact set estimator. In comparison to the LSW test, the simulation study demonstrates that our test is noticeably less conservative on the boundary of the null outside of the least favorable configuration, and has overall improved power. These benefits, however, arise at the expense of introducing a fine tuning parameter, whose selection can be important in moderate sample sizes.

8 Acknowledgments

Rami V. Tabri thanks Russell Toth, Peter Exterkate, Mervyn Silvapulle, and Yoon-Jae Whang for helpful comments and discussions.

References

D. W. K Andrews. Inconcsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space. *Econometrica*, 68(2):399–405, 2000.

D. W. K Andrews and G. Soares. Inference for Parameters Defined by Moment Inequalities using Generalized Moment Selection. *Econometrica*, 78(1):119–157, 2010.

A. Atkinson. On the Measurement of Poverty. *Econometrica*, 55:749–764, 1987.

G. F. Barrett and S. G. Donald. Consistent Tests for Stochastic Dominance. *Econometrica*, 71(1): 71–104, 2003.

P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, third edition, 1995.

François Bourguignon and Gary Fields. Discontinuous losses from poverty, generalized pa measures, and optimal transfers to the poor. *Journal of Public Economics*, 63(2):155 – 175, 1997. ISSN 0047-2727.

R. Davidson. Testing for Restricted Stochastic Dominance: Some Further Analysis. *Review of Economic Analysis*, 1:34–59, 2009.

R. Davidson and J-Y Duclos. Testing for Restricted Stochastic Dominance. *Econometric Reviews*, 32(1):84–125, 2013.

J. Foster and A. Shorrocks. Poverty Orderings. *Econometrica*, 56(1):173–177, 1988.

L. Horváth, P. Kokoszka, and R. Zitikis. Testing for Stochastic Dominance using the Weighted McFadden-type Statistic. *Journal of Econometrics*, 133:191–205, 2006.

O. Linton, K. Song, and Y-J. Whang. An Improved Bootstrap Test for Stochastic Dominance. *Journal Of Econometrics*, 154:186–202, 2010.

This Appendix is not to be published. It will be made available on the web.

Appendix

to

An Improved Bootstrap Test for Restricted Stochastic Dominance

Thomas M. Lok and Rami V. Tabri

School of Economics, The University of Sydney,

Sydney, New South Wales, 2006, Australia. Email: rami.tabri@sydney.edu.au.

A Proofs of Results

A.1 Theorem 1

Proof. **Part 1**. The proof proceeds by the direct method. Lemma B.4 implies that the $\acute{p}_i$ can be characterized in terms of Lagrange multipliers as in (B.17). Without loss of generality, let

$$\Delta(\acute{P}_n) = \{t_1, t_2, \ldots, t_{\omega_n}\}. \tag{A.1}$$

Therefore, the probabilities (B.17) can be expressed as

$$\acute{p}_i = \frac{1}{n}\left(1 + \sum_{j=1}^{\omega_n} \acute{\mu}_j g(X_i; t_j)\right)^{-1}. \tag{A.2}$$

Therefore,

$$E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] = -\sum_{i=1}^{n} \acute{p}_i \sum_{t' \in \Delta(\acute{P}_n)} \acute{\mu}_{t'} g\left(\mathbf{X}_i; t'\right) g\left(\mathbf{X}_i; t\right), \tag{A.3}$$

where $\acute{\mu}_{t'} \geq 0 \; \forall t' \in \Delta(\acute{P}_n)$. Finally, given t', Property 1 implies the desired result because

$$g\left(\mathbf{x}; t'\right) g\left(\mathbf{x}; t\right) \geq 0 \quad \forall (t, \mathbf{x}) \in [\underline{t}, \overline{t}] \times \mathcal{X}. \tag{A.4}$$

This concludes the proof.

Part 2. First, we prove the probability of the event $\left\{\acute{\Delta}_n \subset \hat{\Delta}_n\right\}$ (in $\mathcal{A}_n$) converges to unity in probability uniformly over $\mathcal{P}_0(c_1, c_2)$, as the sample size tends to infinity. The proof follows the direct method and makes use of the result in part 1 of the this theorem.

Let $t \in \acute{\Delta}_n$, then $\left\{-r_n < E_{\acute{P}_n}\left[g\left(\mathbf{X}, t\right)\right]\right\}$ occurs. Consequently, part 1 of the theorem implies the probability of the event $\left\{-r_n < E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right]\right\}$, converges to unity in probability, uniformly over $\mathcal{P}_0(c_1, c_2)$, as the sample size tends to infinity. Now we show the probability of the event $\left\{r_n > E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right]\right\}$, for each $t \in \acute{\Delta}_n$, tends to unity with uniformity.

Noting that for $t \in \acute{\Delta}_n$,

$$E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] = E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] - E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right] + E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right] \tag{A.5}$$

$$< E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] - E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right] + r_n \tag{A.6}$$

$$= O_P(n^{-1/2}) + r_n \quad \text{uniformly over} \quad \mathcal{P}_0(c_1,c_2), \tag{A.7}$$

where (A.7) follows by Proposition B.1. Next we show that the probability of the event

$$E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] \in [r_n, O_P(n^{-1/2}) + r_n)$$

is uniformly asymptotically negligible.

Consider the following probability $P\left[E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] \in [r_n, O_P(n^{-1/2}) + r_n)\right]$, which is equal to

$$P\left[\sqrt{n}\left(E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] - E_P\left[g\left(\mathbf{X},t\right)\right]\right) + \sqrt{n}E_P\left[g\left(\mathbf{X},t\right)\right] \in [\sqrt{n}r_n, O_P(1) + \sqrt{n}r_n)\right]. \tag{A.8}$$

For $t \in \Delta(P)$, this probability is equal to $P\left[\sqrt{n}E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] \in [\sqrt{n}r_n, O_P(1) + \sqrt{n}r_n)\right]$, and the Uniform Central Limit Theorem establishes that $\sqrt{n}E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] = O_p(1)$, uniformly over $\mathcal{P}_0(c_1,c_2)$. Because $\sqrt{n}r_n \to +\infty$, it follows that

$$\sup_{P \in \mathcal{P}_0(c_1,c_2)} P\left[\sqrt{n}E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] \in [\sqrt{n}r_n, O_P(1) + \sqrt{n}r_n)\right] \to 0. \tag{A.9}$$

Therefore, if $t \in \Delta(P)$, then $\sup_{P \in \mathcal{P}_0(c_1,c_2)} P\left[E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] < r_n\right] \to 1$.

Now we focus on the last case under H_0, which is when $t \notin \Delta(P)$. In this case, $E_P\left[g\left(\mathbf{X},t\right)\right] < 0$ and we have that

$$\sqrt{n}E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] = \sqrt{n}\left(E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] - E_P\left[g\left(\mathbf{X},t\right)\right]\right) + \sqrt{n}E_P\left[g\left(\mathbf{X},t\right)\right] \tag{A.10}$$

$$O_P(1) + \sqrt{n}E_P\left[g\left(\mathbf{X},t\right)\right] \quad \text{uniformly over} \quad \mathcal{P}_0(c_1,c_2). \tag{A.11}$$

Since $\sqrt{n}E_P\left[g\left(\mathbf{X},t\right)\right]$ diverges to $-\infty$, $\sqrt{n}E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right]$ also diverges to $-\infty$, but uniformly over $\mathcal{P}_0(c_1,c_2)$. Combining this result with the fact that $\sqrt{n}r_n \to +\infty$, implies that the probability (A.8) tends to zero with uniformity. Therefore, $\sup_{P\in\mathcal{P}_0(c_1,c_2)} P\left[E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] < r_n\right] \to 1$, which concludes the proof of this part of the theorem.

Now we turn our focus to the event $\left\{\Delta\left(P\right) \subset \acute{\Delta}_n\right\}$. Let $t \in \Delta\left(P\right)$ and consider the event

$$\left\{-r_n < E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right] < r_n\right\} = \left\{-r_n < E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right] - E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] + E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] < r_n\right\} \tag{A.12}$$

$$= \left\{-r_n < O_P(n^{-1/2}) + E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] < r_n\right\} \tag{A.13}$$

$$= \left\{-\sqrt{n}r_n < O_P(1) + E_{\hat{P}_n} + \sqrt{n}E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] < \sqrt{n}r_n\right\}, \tag{A.14}$$

uniformly over $\mathcal{P}_0(c_1,c_2)$ by Proposition B.1. As $t \in \Delta\left(P\right)$, we have that $\sqrt{n}E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] = O_P(1)$ uniformly over $\mathcal{P}_0(c_1,c_2)$, by the Uniform Central Limit Theorem. Therefore, the event (A.14) is equal to

$$\left\{-\sqrt{n}r_n \leq O_P(1) \leq \sqrt{n}r_n\right\}, \tag{A.15}$$

whose probability tends to unity uniformly over $\mathcal{P}_0(c_1,c_2)$, because $\sqrt{n}r_n \to +\infty$. This concludes part 2.

Part 3. The proof follows identical steps to those in the second part of the proof of Claim 1, in LSW on page 200. □

A.2 Theorem 2

Proof. The proof proceeds by the direct method. Let

$$\gamma_n^{\star}\left(t\right) = \left(\max\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[g\left(\mathbf{X}_{i,l}^{\star};t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]\right], 0\right\}\right)^2, \tag{A.16}$$

then consider the following,

$$\left|\hat{T}^{\star}_{n,l}-\acute{T}^{\star}_{n,l}\right| = \begin{cases} \int_{[\underline{t},\overline{t}]-\acute{\Delta}_n} \gamma^{\star}_n(t)\ dt & \text{if } \int_{\acute{\Delta}_n} dt > 0, \int_{\hat{\Delta}_n} dt = 0 \\ \int_{[\underline{t},\overline{t}]-\hat{\Delta}_n} \gamma^{\star}_n(t)\ dt & \text{if } \int_{\acute{\Delta}_n} dt = 0, \int_{\hat{\Delta}_n} dt > 0 \\ \int_{\acute{\Delta}_n \ominus \hat{\Delta}_n} \gamma^{\star}_n(t)\ dt & \text{if } \int_{\acute{\Delta}_n} dt > 0, \int_{\hat{\Delta}_n} dt > 0 \\ 0 & \text{if } \int_{\acute{\Delta}_n} dt = 0, \int_{\hat{\Delta}_n} dt = 0, \end{cases} \tag{A.17}$$

where $\ominus$ denotes the symmetric difference operator on sets. We have

$$\left|\hat{T}^{\star}_{n,l}-\acute{T}^{\star}_{n,l}\right| \leq \begin{cases} \left(\sup_{t\in[\underline{t},\overline{t}]} \gamma^{\star}_n(t)\right) \int_{[\underline{t},\overline{t}]-\acute{\Delta}_n} dt & \text{if } \int_{\acute{\Delta}_n} dt > 0, \int_{\hat{\Delta}_n} dt = 0 \\ \left(\sup_{t\in[\underline{t},\overline{t}]} \gamma^{\star}_n(t)\right) \int_{[\underline{t},\overline{t}]-\hat{\Delta}_n} dt & \text{if } \int_{\acute{\Delta}_n} dt = 0, \int_{\hat{\Delta}_n} dt > 0 \\ \left(\sup_{t\in[\underline{t},\overline{t}]} \gamma^{\star}_n(t)\right) \int_{\acute{\Delta}_n \ominus \hat{\Delta}_n} dt & \text{if } \int_{\acute{\Delta}_n} dt > 0, \int_{\hat{\Delta}_n} dt > 0, \\ 0 & \text{if } \int_{\acute{\Delta}_n} dt = 0, \int_{\hat{\Delta}_n} dt = 0. \end{cases} \tag{A.18}$$

To prove the result we need to prove that $\left(\sup_{t\in[\underline{t},\overline{t}]} \gamma^{\star}_n(t)\right)$ is $O_P(1)$ conditional on $\mathcal{A}_n$, uniformly in $\mathcal{P}_0(\epsilon, c_1, c_2)$. and then apply Theorem 1 to the integrals in (A.18). Since the set of moment functions

$$\mathcal{F} = \left\{\mathbf{x} \mapsto g(\mathbf{x}, t), t \in [\underline{t}, \overline{t}]\right\}$$

is uniform Donsker with respect to $\mathcal{P}_0(\epsilon, c_1, c_2)$, Lemma A.2 of LSW implies that it is also bootstrap uniform Donsker. Therefore, applying Lemma A.1 (uniform continuous mapping theorem) of LSW to $\left(\sup_{t\in[\underline{t},\overline{t}]} \gamma^{\star}_n(t)\right)$ yields the desired result.

Theorem 1 implies that $\hat{\Delta}_n$ converges to $\Delta(P)$ in probability, uniformly in $\mathcal{P}_0(c_1, c_2)$. Then, Theorem 1 implies $\acute{\Delta}_n$ converges to $\Delta(P)$ in probability, uniformly in $\mathcal{P}_0(\epsilon, c_1, c_2)$, as $\mathcal{P}_0(\epsilon, c_1, c_2) \subset \mathcal{P}_0(c_1, c_2)$. Firstly, suppose that $\int_{\Delta(P)} dt = 0$. Then for large n, the bootstrap statistics $\hat{T}^{\star}_{n,l}$ and $\acute{T}^{\star}_{n,l}$ will be equal in the limit as n tends to infinity with probability tending to 1, uniformly in $\mathcal{P}_0(\epsilon, c_1, c_2)$, which yields the desired result.

Now suppose that $\int_{\Delta(P)} dt > 0$. Then, for large n, we must have $\acute{\Delta}_n \neq \emptyset, \hat{\Delta}_n \neq \emptyset$ with probability tending to one, uniformly in $\mathcal{P}_0(\epsilon, c_1, c_2)$. Applying Theorem 1 to this case in (A.18) implies $\acute{\Delta}_n \ominus \hat{\Delta}_n$ converges in probability to the empty set, uniformly in $\mathcal{P}_0(\epsilon, c_1, c_2)$. Therefore,

$$\left(\sup_{t\in[\underline{t},\overline{t}]} \gamma_n^\star(t)\right) \int_{\acute{\Delta}_n \ominus \hat{\Delta}_n} dt \xrightarrow{P} 0 \tag{A.19}$$

conditional on $\mathcal{A}_n$ uniformly in $\mathcal{P}_0(\epsilon, c_1, c_2)$. Therefore, $\hat{T}^\star_{n,l} - \acute{T}^\star_{n,l} \xrightarrow{P} 0$ conditional on $\mathcal{A}_n$ uniformly in $\mathcal{P}_0(\epsilon, c_1, c_2)$.

Now Consider the following

$$\begin{aligned}
\left|\hat{\Upsilon}_B - \acute{\Upsilon}_B\right| &= \left|\frac{1}{B}\sum_{l=1}^{B} 1\left[\hat{T}^\star_{n,l} \geq \hat{T}_n\right] - \frac{1}{B}\sum_{l=1}^{B} 1\left[\acute{T}^\star_{n,l} \geq \hat{T}_n\right]\right| \\
&= \left|\frac{1}{B}\sum_{l=1}^{B}\left(1\left[\hat{T}^\star_{n,l} \geq \hat{T}_n\right] - 1\left[\acute{T}^\star_{n,l} \geq \hat{T}_n\right]\right)\right| \\
&\leq \frac{1}{B}\sum_{l=1}^{B}\left|1\left[\hat{T}^\star_{n,l} \geq \hat{T}_n\right] - 1\left[\acute{T}^\star_{n,l} \geq \hat{T}_n\right]\right| \\
&= \frac{1}{B}\sum_{l=1}^{B} 1\left[\hat{T}^\star_{n,l} \leq \hat{T}_n \leq \acute{T}^\star_{n,l} \quad \text{Xor} \quad \acute{T}^\star_{n,l} \leq \hat{T}_n \leq \hat{T}^\star_{n,l}\right] \\
&= 1 - \frac{1}{B}\sum_{l=1}^{B} 1\left[\hat{T}^\star_{n,l} \leq \hat{T}_n \leq \acute{T}^\star_{n,l} \quad \text{and} \quad \acute{T}^\star_{n,l} \leq \hat{T}_n \leq \hat{T}^\star_{n,l}\right]
\end{aligned} \tag{A.20}$$

where Xor is the exclusive "or" operator.

The fact that the bootstrap test statistics are uniformly asymptotically equal implies

$$\frac{1}{B}\sum_{l=1}^{B} 1\left[\hat{T}^\star_{n,l} \leq \hat{T}_n \leq \acute{T}^\star_{n,l} \quad \text{and} \quad \acute{T}^\star_{n,l} \leq \hat{T}_n \leq \hat{T}^\star_{n,l}\right] \xrightarrow{P} 1 \tag{A.21}$$

conditional on $\mathcal{A}_n$, uniformly in $\mathcal{P}_0(\epsilon, c_1, c_2)$. Therefore, the right side of (A.20) converges to zero in probability conditional on $\mathcal{A}_n$, uniformly in $\mathcal{P}_0(\epsilon, c_1, c_2)$. This yields the desired result, and concludes the proof. □

A.3 Theorem 3

Proof. The proof proceeds by the direct method. Because Theorem 3 of LSW shows the test statistic $\hat{T}_n$, given by (2.6), diverges to infinity under the alternative, to prove the desired result we only need to show that the bootstrap test statistic is $O_P(1)$ (conditional on $\mathcal{A}_n$) under H_1. These two conditions imply that the bootstrap p-value $\acute{\Upsilon}_B$ converges to zero in probability under H_1.

For $P_0 \in \mathcal{P}$, Lemma C.2 shows the existence and uniqueness of the solution from the constrained empirical likelihood problem, $\acute{\mathbf{p}}$, to be an event with probability converging to unity. Hence, the contact set $\acute{\Delta}_n$ exists with probability converging to unity. Since the bootstrap test statistic $\acute{T}_n^{\star}$ is bounded above by

$$\int_{\underline{t}}^{\bar{t}} \left(\max \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g\left(\mathbf{X}_i^{\star}; t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right) \right] \right], 0 \right\} \right)^2 dt, \tag{A.22}$$

which converges in distribution (conditional on $\mathcal{A}_n$) to the distribution of $\int_{\underline{t}}^{\bar{t}} \left(\max\left\{\nu(t), 0\right\}\right)^2 dt$, it follows that $\acute{T}_n^{\star} = O_P(1)$ conditional on $\mathcal{A}_n$. This concludes the proof.

□

A.4 Theorem 4

Proof. **Part 1**. The proof proceeds by the direct method. We first show that the condition

$$\lim_{n \to +\infty} P\left[\inf_{t \in \hat{\Delta}_n \cap \acute{\Delta}_n^c} \hat{\sigma}^2(t) \geq \sup_{t \in \acute{\Delta}_n \cap \hat{\Delta}_n^c} \hat{\sigma}^2(t) \right] = 1$$

implies that $\acute{T}_n^{\star} \leq \hat{T}_n^{\star}$, for large enough n with probability approaching unity conditional on $\mathcal{A}_n$. Conditional on $\mathcal{A}_n$, the bootstrap central limit theorem implies that the bootstrap empirical processes $\left\{\mathbb{G}_n(t), t \in \acute{\Delta}_n\right\}$ and $\left\{\mathbb{G}_n(t), t \in \hat{\Delta}_n\right\}$, where $\mathbb{G}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g\left(\mathbf{X}_{i,l}^{\star}; t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right) \right] \right]$, converge weakly to zero-mean Gaussian processes. The only difference between these processes is their index sets. The sample variance function $\hat{\sigma}^2(\cdot)$ is equal to the variance functions of the these processes, but over the respective domains. In general, we have that

$$\hat{\sigma}^2(t) = E^\star\left[\mathbb{G}_n^2(t)\right] \quad \forall t \in [\underline{t}, \overline{t}] \tag{A.23}$$

where $E^\star$ denotes expectation with respect to the bootstrap measure. The condition

$$\lim_{n\to+\infty} P\left[\inf_{t\in\hat{\Delta}_n\cap\acute{\Delta}_n^c} \hat{\sigma}^2(t) \geq \sup_{t\in\acute{\Delta}_n\cap\hat{\Delta}_n^c} \hat{\sigma}^2(t)\right] = 1$$

implies that

$$\inf_{t\in\hat{\Delta}_n} E^\star\left[\mathbb{G}_n^2(t)\right] \geq \sup_{t\in\acute{\Delta}_n} E^\star\left[\mathbb{G}_n^2(t)\right] \tag{A.24}$$

holds for large n with probability approaching unity. Since the processes are asymptotically Gaussian conditional on $\mathcal{A}_n$, for large n the inequality (A.24) implies $\inf_{t\in\hat{\Delta}_n} \mathbb{G}_n(t) \geq \sup_{t\in\acute{\Delta}_n} \mathbb{G}_n(t)$ with probability approaching unity conditional on $\mathcal{A}_n$. As the bootstrap test statistics are a (deterministic) continuous functional of the bootstrap empirical process, we must have $\hat{T}_n^\star \geq \acute{T}_n^\star$ for large enough n with probability approaching unity conditional on $\mathcal{A}_n$, which implies the desired result.

Part 2. Arguments similar to those in part 1 can be used to prove the desired result; therefore, we omit the proof for brevity.

□

A.5 Theorem 5

Proof. We prove this theorem by showing that $\hat{T}_n^\star - \acute{T}_n^\star = o_{P_n}(1)$ conditional on $\mathcal{A}_n$, under the local alternatives. The Lemma C.3 implies the existence of $\acute{\mathbf{p}}$ and the Lagrange multipliers for large enough n with probability approaching unity, under the local alternatives. Furthermore, Lemma C.4 implies that $\int_{\hat{\Delta}_n} dt > 0$ and $\int_{\acute{\Delta}_n} dt > 0$ for large enough n with probability approaching unity, under the local alternatives. Therefore, for large n,

$$\hat{T}_n^\star - \acute{T}_n^\star = \int_{\hat{\Delta}_n} \gamma_n^\star(t)\, dt - \int_{\acute{\Delta}_n} \gamma_n^\star(t)\, dt \quad \text{where} \tag{A.25}$$

$$\gamma_n^\star(t) = \left(\max\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g\left(\mathbf{X}_i^\star; t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right], 0 \right\}\right)^2 . \tag{A.26}$$

Now since $\sup_{t\in[\underline{t},\overline{t}]} \gamma_n^\star(t) = O_{P_n}(1)$ conditional on $\mathcal{A}_n$, Lemma C.5 implies that the contact set estimators are asymptotically equal with probability approaching unity under the local alternatives, which yields

$$\hat{T}_n^\star - \acute{T}_n^\star = o_{P_n}(1) \quad \text{conditional on} \quad \mathcal{A}_n. \tag{A.27}$$

Therefore, it follows that $\lim_{n\to+\infty} P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] = \lim_{n\to+\infty} P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right]$. □

B Intermediate Technical Results for Theorems 1 and 2

This section presents intermediate technical results that are used in the proofs of Theorems 1 and 2. Subsection B.1 presents a technical result that is an immediate consequence of Part (i) of Definition 3. Section B.2 uses this technical result to prove the existence and uniqueness of the solution from the constrained empirical likelihood problem, under the null. In turn, Section B.3 uses the results of the previous sections to develop the large-sample properties of the Lagrange multipliers under H_0, which make use of Part (ii) of Definition 3. Finally, Section B.4 uses all of the previous results to establish the uniform consistency of the proposed empirical likelihood estimator of the moments, under H_0.

B.1 A Consequence of Part (i) of Definition 3

The result of this subsection concerns the large-sample behavior of the likelihood of observing the event $\{I_n^- \neq \emptyset\}$ under H_0, where

$$I_n^- = \left\{i \in \{1, \ldots, n\} : g\left(\mathbf{X}_i; t\right) < 0 \; \forall t \in [\underline{t}, \overline{t}]\right\}. \tag{B.1}$$

Lemma B.1. Given $[c_1, c_2] \in (0,1] \times (0, +\infty)$, let $P_0 \in \mathcal{P}_0(c_1, c_2)$. Then $\sup_{P \in \mathcal{P}_0(c_1,c_2)} P\left[I_n^- \neq \emptyset\right] \to 1$.

Proof. The proof proceeds by the direct method. We show that the probability of the complement of $\{I_n^- \neq \emptyset\}$ converges to zero uniformly in $\mathcal{P}_0(c_1, c_2)$. This complement of this event is

$$\left\{I_n^- = \emptyset\right\} = \left\{\text{for each i } \exists t \in [\underline{t}, \overline{t}]; g\left(\mathbf{X}_i; t\right) \geq 0\right\}.$$

By the bivariate random sampling assumption on $\{\mathbf{X}_i\}_{i=1}^n$, we have that

$$\sup_{P \in \mathcal{P}_0(c_1,c_2)} P\left[I_n^- = \emptyset\right] = \sup_{P \in \mathcal{P}_0(c_1,c_2)} \left(P\left[\sup_{t \in [\underline{t}, \overline{t}]} g\left(\mathbf{X}_1; t\right) \geq 0\right]\right)^n \tag{B.2}$$

$$= \sup_{P \in \mathcal{P}_0(c_1,c_2)} \left(1 - P\left[\sup_{t \in [\underline{t}, \overline{t}]} g\left(\mathbf{X}_1; t\right) < 0\right]\right)^n \tag{B.3}$$

$$\leq (1 - c_1)^n \to 0 \tag{B.4}$$

as $n \to +\infty$, since $c_1 \in (0, 1]$. $\square$

B.2 Existence and Uniqueness of $\acute{\mathbf{p}}$ under H_0

Let $\mathcal{H}_n = \{p_i, i = 1, \ldots, n; \sum_{i=1}^n p_i = 1, p_i \geq 0, \forall i = 1, \ldots, n\}$, and denote the interior of this set by $\mathcal{H}_n^\circ$. Additionally, let $\mathcal{H}_n^0(\mathbf{X}) = \left\{\mathbf{p} \in \mathcal{H}_n : \sum_{i=1}^n p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t \in \mathcal{T}_{N(n)}\right\}$.

Lemma B.2. On the event $\{\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\}$, the random set

$$\arg\max \left\{ \sum_{i=1}^n \log(p_i) \,;\, p_i \geq 0,\, \sum_{i=1}^n p_i = 1,\, \sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq 0 \quad \forall t \in \mathcal{T}_{N(n)} \right\}$$

is nonempty and a singleton.

Proof. The proof proceeds by verifying the conditions of Weierstrass' Theorem. The objective function is strictly concave in the probabilities. The constraint set, $\mathcal{H}_n^0(\mathbf{X})$, is certainly bounded. It is the finite intersection of closed half-planes (which are convex), and since convexity and closedness are preserved under a finite number of intersections, it is closed and convex. Thus, we are done whenever $\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset$. □

Lemma B.3. Given $[c_1, c_2] \in (0, 1] \times (0, +\infty)$, suppose that $P_0 \in \mathcal{P}_0(c_1, c_2)$. Then

$$\sup_{P \in \mathcal{P}_0(c_1, c_2)} P\left[\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right] \to 1 \quad \text{as} \quad n \to +\infty.$$

Proof. The proof proceeds by the direct method.

For large n and uniformly in $\mathcal{P}_0(c_1, c_2)$, Lemma B.1 implies that the event

$$\exists i \in \{1, 2, \ldots, n\} \;\; g(\mathbf{X}_i, t) < 0 \quad \forall t \in \mathcal{T}_{N(n)} \tag{B.5}$$

occurs with probability approaching 1, since $\mathcal{T}_{N(n)} \subset [\underline{t}, \overline{t}]$ for each n. Therefore, without loss of generality, suppose that this event holds only for $i = 1$. Then it follows that

$$\forall i \in \{2, 3, \ldots, n\} : \sup_{t \in \mathcal{T}_{N(n)}} g(\mathbf{X}_i, t) \geq 0. \tag{B.6}$$

A probability vector $\mathbf{p}$ in $\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ$ must satisfy $p_i > 0 \; \forall i = 1, \ldots, n$, $\sum_{i=1}^n p_i = 1$ and

$$\sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq 0 \quad \forall t \in \mathcal{T}_{N(n)} \iff 1 > p_1 \geq \sum_{i=2}^n p_i \left(\frac{g(\mathbf{X}_i; t)}{-g(\mathbf{X}_1; t)} \right) \quad \forall t \in \mathcal{T}_{N(n)} \tag{B.7}$$

Therefore, a sufficient condition for the inequalities in (B.7) is

$$\sup_{t\in\mathcal{T}_{N(n)}} \sum_{i=2}^{n} p_i \left(\frac{g(\mathbf{X}_i;t)}{-g(\mathbf{X}_1;t)} \right) \le \max_{2\le i\le n} \left(\sup_{t\in\mathcal{T}_{N(n)}} \frac{g(\mathbf{X}_i;t)}{-g(\mathbf{X}_1;t)} \right) \sum_{i=2}^{n} p_i \tag{B.8}$$

$$\le \max_{2\le i\le n} \left(\frac{\sup_{t\in\mathcal{T}_{N(n)}} g(\mathbf{X}_i;t)}{\inf_{t\in\mathcal{T}_{N(n)}} -g(\mathbf{X}_1;t)} \right) \sum_{i=2}^{n} p_i < 1. \tag{B.9}$$

It should be noted that $\inf_{t\in\mathcal{T}_{N(n)}} -g(\mathbf{X}_1;t) > 0$ follows directly from (B.5). On the event

$$\max_{2\le i\le n} \left(\frac{\sup_{t\in\mathcal{T}_{N(n)}} g(\mathbf{X}_i;t)}{\inf_{t\in\mathcal{T}_{N(n)}} -g(\mathbf{X}_1;t)} \right) \in [0,1],$$

any positive probability vector satisfies the inequalities (B.7). Otherwise, on the event

$$\max_{2\le i\le n} \left(\frac{\sup_{t\in\mathcal{T}_{N(n)}} g(\mathbf{X}_i;t)}{\inf_{t\in\mathcal{T}_{N(n)}} -g(\mathbf{X}_1;t)} \right) > 1,$$

the inequality (B.8) is equivalent to

$$\sum_{i=2}^{n} p_i < \frac{1}{\max_{2\le i\le n} \left(\frac{\sup_{t\in\mathcal{T}_{N(n)}} g(\mathbf{X}_i;t)}{\inf_{t\in\mathcal{T}_{N(n)}} -g(\mathbf{X}_1;t)} \right)} \iff 1 - \frac{1}{\max_{2\le i\le n} \left(\frac{\sup_{t\in\mathcal{T}_{N(n)}} g(\mathbf{X}_i;t)}{\inf_{t\in\mathcal{T}_{N(n)}} -g(\mathbf{X}_1;t)} \right)} < p_1. \tag{B.10}$$

Thus, for any p_1 such that

$$1 - \left(\max_{2\le i\le n} \left(\frac{\sup_{t\in\mathcal{T}_{N(n)}} g(\mathbf{X}_i;t)}{\inf_{t\in\mathcal{T}_{N(n)}} -g(\mathbf{X}_1;t)} \right) \right)^{-1} < p_1 < 1,$$

there is a set of possible choices for $p_2, p_3, \ldots, p_n$ such that $p_i > 0\ \forall i = 1, \ldots, n$, and $\sum_{i=1}^{n} p_i = 1$. This concludes the proof. □

B.3 Properties of Lagrange Multipliers under H_0

In this subsection we present the properties of the Lagrange multipliers arising in the constrained empirical likelihood problem introduced in Section 3. This optimization problem has the following Lagrangian function:

$$\mathcal{L} = \sum_{i=1}^{n} \log(p_i) + \lambda \left(1 - \sum_{i=1}^{n} p_i\right) - n \sum_{t \in \mathcal{T}_{N(n)}} \mu_t \sum_{i=1}^{n} p_i g(\mathbf{X}_i; t), \tag{B.11}$$

where $\lambda \in \mathbb{R}$ is the multiplier on the equality constraint $\sum_{i=1}^{n} p_i = 1$, and $\mu_t \geq 0$ for $t \in \mathcal{T}_{N(n)}$ are the multipliers on the inequality constraints. The Karush-Kuhn-Tucker (KKT) conditions are

$$\frac{1}{p_i} = \lambda + n \sum_{t \in \mathcal{T}_{N(n)}} \mu_t g(\mathbf{X}_i; t) \quad i = 1, 2, \ldots, n \tag{B.12}$$

$$\sum_{i=1}^{n} p_i g(\mathbf{X}_i; t) \leq 0 \quad \forall t \in \mathcal{T}_{N(n)}, \quad \sum_{i=1}^{n} p_i = 1 \tag{B.13}$$

$$\mu_t \sum_{i=1}^{n} p_i g(\mathbf{X}_i; t) = 0 \quad \forall t \in \mathcal{T}_{N(n)}. \tag{B.14}$$

In classical optimization theory, the existence of $\acute{\lambda}$ and $\acute{\mu}_t$ for $t \in \mathcal{T}_{N(n)}$ that solve KKT conditions along with $\acute{\mathbf{p}}$ results from a constraint qualification. This paper uses the Mangasarian-Fromovitz constraint qualification. In the setting of this paper, the Mangasarian-Fromovitz constraint qualification is the following event

$$\mathcal{S}_n = \left\{ \exists d \in \mathbb{R}^n : \sum_{i=1}^{n} d_i = 0 \quad \text{and} \quad \sum_{i=1}^{n} d_i g(\mathbf{X}_i; t) < 0 \; \forall t \in \Delta\left(\acute{P}_n\right) \right\}, \tag{B.15}$$

where $\Delta\left(\acute{P}_n\right) = \left\{ t \in \mathcal{T}_{N(n)} : \sum_{i=1}^{n} \acute{p}_i g(\mathbf{X}_i; t) = 0 \right\}$. We have the following result.

Lemma B.4 (Existence)**.** Given $[c_1, c_2] \in (0, 1] \times (0, +\infty)$, suppose that $P_0 \in \mathcal{P}_0(c_1, c_2)$. Then

$$\sup_{P \in \mathcal{P}_0(c_1, c_2)} P[\mathcal{S}_n] \to 1 \quad \text{as} \quad n \to +\infty.$$

Proof. The proof proceeds by using the direct method. For large enough n and uniformly in $\mathcal{P}_0(c_1, c_2)$ Lemmas B.2 and B.3 imply that $\acute{\mathbf{p}}$ exists and is unique. To prove the desired result, we will show the probability of the event $\mathcal{S}_n$ converges to one, uniformly in $\mathcal{P}_0(c_1, c_2)$.

Noting that the moment functions satisfy Property 1, consider the following construction for the $d \in \mathbb{R}^n : \sum_{i=1}^n d_i = 0$, and the sign restrictions

$$\text{sign}\,(d_i) = \begin{cases} < 0, & \text{if } g\left(\mathbf{X}_i, t\right) \geq 0\ \forall t \in \mathcal{T}_{N(n)} \\ > 0, & \text{if } g\left(\mathbf{X}_i, t\right) < 0\ \forall t \in \mathcal{T}_{N(n)}, \end{cases} \tag{B.16}$$

Lemma B.1 implies the occurrence of the event $\left\{\exists i : g\left(\mathbf{X}_i, t\right) < 0\ \forall t \in \mathcal{T}_{N(n)}\right\}$ with probability approaching one, uniformly in $\mathcal{P}_0(c_1, c_2)$. Therefore, the above construction is asymptotically feasible. Such vectors d trivially satisfy the conditions of the Mangasarian-Fromovitz constraint qualification. This concludes the proof since the above implies that the probability of the event S_n converges to one uniformly in $\mathcal{P}_0(c_1, c_2)$. □

In fact, using the KKT conditions, one can easily show that $\acute{\lambda} = n$,

$$\acute{p}_i = \frac{1}{n}\left(\frac{1}{1 + \sum_{t \in \Delta(\acute{P}_n)} \acute{\mu}_t g\left(\mathbf{X}_i; t\right)}\right) \quad i = 1, 2, \ldots, n, \tag{B.17}$$

where $\left\{\acute{\mu}_t = 0, \forall t \in \mathcal{T}_{N(n)} - \Delta(\acute{P}_n)\right\}$ and $\left\{\acute{\mu}_t \geq 0, \forall t \in \Delta(\acute{P}_n)\right\}$. The Mangasarian-Fromovitz constraint qualification implies that there exists a compact set of multipliers on the binding constraints that satisfy the KKT conditions. We denote this set of multipliers by

$$\Lambda_n\left(\acute{P}_n\right) = \left\{\acute{\mu}_t\ t \in \Delta(\acute{P}_n) \text{ that satisfy } (B.12) -\ (B.14)\right\}. \tag{B.18}$$

Next, we focus on the large sample properties of the multipliers in $\Lambda_n\left(\acute{P}_n\right)$, under H_0. Let

$w \in \mathbb{Z}_+ \cup \{+\infty\}$, and define the Banach spaces, as indexed by w,

$$l_w^1 = \left\{ a = (a_1, a_2, \ldots, a_w) \in \mathbb{R}^w : \sum_{j=1}^{w} |a_j| < +\infty \right\}, \quad \text{(B.19)}$$

normed by $\|a\|_{l_w^1} = \sum_{j=1}^{w} |a_j|$.

Lemma B.5 (Asymptotic Bound for Lagrange Multipliers)**.** Given $[c_1, c_2] \in (0, 1] \times (0, +\infty)$, suppose that $P_0 \in \mathcal{P}_0(c_1, c_2)$. Then

(i) $\lim_{n\to+\infty} \sup_{P\in\mathcal{P}_0(c)} P\left[\Delta(\acute{P}_n) \subset \Delta(P_0)\right] = 1$.

(ii) Let $\omega_n = \left|\Delta(\acute{P}_n)\right|$. Denote the vector of Lagrange multipliers on the constraints binding constraints by $\acute{\boldsymbol{\mu}}$ and the $l_{\omega_n}^1$ norm of the vector $\acute{\boldsymbol{\mu}}$ by $||\acute{\boldsymbol{\mu}}||_{l_{\omega_n}^1}$. Then $\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l_{\omega_n}^1} = o_P(1)$ uniformly in $\mathcal{P}_0(c_1, c_2)$ at the $\sqrt{n}$-rate.

Proof.

(i) We show this result using proof by contrapositive, that is, we show that for large n,

$$t \notin \Delta(P_0) \implies t \notin \Delta(\acute{P}_n)$$

Consider any $t \in [\underline{t}, \overline{t}]$. For large enough n, Property 1 and the non-negativity of the Lagrange multipliers implies that

$$\sum_{i=1}^{n} p_i' g(X_i;t) \leq \frac{1}{n}\sum_{i=1}^{n} g(X_i;t) = \frac{1}{n}\sum_{i=1}^{n} g(X_i;t) - E_{P_0}\left[g(X;t)\right] + E_{P_0}\left[g(X;t)\right] \quad \text{(B.20)}$$

Now, consider $t \notin \Delta(P_0)$. As $P_0 \in \mathcal{P}_0(c_1, c_2)$, this implies that $E_{P_0}\left[g(X;t)\right] < 0$. By the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i;t) - E_{P_0}\left[g(X;t)\right] = O_P(n^{-1/2})$$

uniformly in $\mathcal{P}_0(c_1, c_2)$. Thus, for sufficiently large n, equation (B.20) simplifies to

$$\sum_{i=1}^{n} \acute{p}_i g(X_i; t) < 0$$

This shows that $t \notin \Delta(\acute{P}_n)$.

(ii) Recall that the cardinality of the set $\Delta(\acute{P}_n)$ is $\omega_n \leq N(n)$. Without loss of generality, let

$$\Delta(\acute{P}_n) = \{t_1, t_2, \ldots, t_{\omega_n}\}\,. \tag{B.21}$$

Therefore, the probabilities (B.17) can be expressed as

$$\acute{p}_i = \frac{1}{n}\left(1 + \sum_{j=1}^{\omega_n} \acute{\mu}_j g(X_i; t_j)\right)^{-1} \tag{B.22}$$

For any choice of $t_j \in \Delta(\acute{P}_n)$, we have

$$\sum_{i=1}^{n} \acute{p}_i g(X_i; t_j) = \frac{1}{n}\sum_{i=1}^{n} \frac{g(X_i; t_j)}{1 + \sum_{j=1}^{\omega_n} \acute{\mu}_j g(X_i; t_j)} = 0 \tag{B.23}$$

To express the system of equations described by (B.23) in vectorised form, define the vector

$$\mathbf{g}_i = [g(X_i; t_1),\ g(X_i; t_2),\ \ldots,\ g(X_i; t_{\omega_n})]^T \tag{B.24}$$

Now, as all the elements of $\acute{\boldsymbol{\mu}}$ are non-negative, the $l^1_{\omega_n}$ norm is simply the sum of all elements of $\acute{\boldsymbol{\mu}}$, i.e. $||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} = \sum_{j=1}^{\omega_n} \mu'_j$. This means we can express the vector $\acute{\boldsymbol{\mu}}$ in the form

$$\acute{\boldsymbol{\mu}} = ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \boldsymbol{\theta}\,, \quad \boldsymbol{\theta} \in \mathbb{R}_+^{\omega_n}$$

Under this construction, the j^{th} element of $\boldsymbol{\theta}$ is

$$\theta_j = \frac{\acute{\mu}_j}{\sum_{j=1}^{\omega_n} \acute{\mu}_j}$$

This implies that $\sum_{j=1}^{\omega_n} \theta_j = 1$. The system of equations defined by (B.23) for all $t \in \Delta(\acute{P}_n)$ can be written in the following form

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{g}_i}{1+(\acute{\boldsymbol{\mu}})^T\mathbf{g}_i} = \mathbf{0} \implies \boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{g}_i}{1+(\acute{\boldsymbol{\mu}})^T\mathbf{g}_i}\right) = 0 \tag{B.25}$$

Define the quantity $Y_i = (\acute{\boldsymbol{\mu}})^T\mathbf{g}_i$. Using the manipulation $\frac{1}{1+Y_i} = 1 - \frac{Y_i}{1+Y_i}$ and the fact that $(\acute{\boldsymbol{\mu}})^T\mathbf{g}_i = \mathbf{g}_i^T\acute{\boldsymbol{\mu}}$ in equation (B.25) gives

$$\begin{aligned}
\boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_i \left(1 - \frac{\mathbf{g}_i^T\acute{\boldsymbol{\mu}}}{1+Y_i}\right)\right) &= 0 \\
\boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_i\right) &= \boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{g}_i\mathbf{g}_i^T\acute{\boldsymbol{\mu}}}{1+Y_i}\right) \\
\boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_i\right) &= \boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{g}_i\mathbf{g}_i^T||\acute{\boldsymbol{\mu}}||\boldsymbol{\theta}}{1+Y_i}\right) \\
\therefore \boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_i\right) &= ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{g}_i\mathbf{g}_i^T}{1+Y_i}\right)\boldsymbol{\theta}
\end{aligned} \tag{B.26}$$

We denote the sample analogue estimate of the covariance matrix of measurement functions over the set of all $t \in \Delta(\acute{P}_n)$ by

$$\hat{\Sigma}_{\Delta(\acute{P}_n)} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_i\mathbf{g}_i^T$$

Define $Y_{max} = \max_i |Y_i|$. Note that

$$Y_{max} = \max_i |Y_i| = \max_i \sum_{j=1}^{\omega_n} \acute{\mu}_j |g(X_i;t_j)| \leq \frac{2\bar{t}^{s-1}}{(s-1)!}\sum_{j=1}^{\omega_n} \acute{\mu}_j = \frac{2\bar{t}^{s-1}}{(s-1)!}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}, \tag{B.27}$$

where we have used the uniform boundedness of the moment functions.

Now, consider

$$\begin{aligned}
||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \left(\boldsymbol{\theta}^T\hat{\Sigma}_{\Delta(\acute{P}_n)}\boldsymbol{\theta}\right) &= ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \left(\boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_i\mathbf{g}_i^T\right)\boldsymbol{\theta}\right) \\
&\leq ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \left(\boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{g}_i\mathbf{g}_i^T}{1+Y_i}\right)\boldsymbol{\theta}\right)(1+Y_{max}) \\
&\leq ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \left(\boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{g}_i\mathbf{g}_i^T}{1+Y_i}\right)\boldsymbol{\theta}\right)(1+||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}) \\
\therefore ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \left(\boldsymbol{\theta}^T\hat{\Sigma}_{\Delta(\acute{P}_n)}\boldsymbol{\theta}\right) &\leq \boldsymbol{\theta}^T \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_i\right)(1+||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}) \qquad \text{(B.28)}
\end{aligned}$$

where the last line results from substituting the expression given in (B.26). Rearranging (B.28) gives

$$||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \left[\boldsymbol{\theta}^T\hat{\Sigma}_{\omega_n}\boldsymbol{\theta} - \boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)\right] \leq \boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right) \quad \forall \acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right), \qquad \text{(B.29)}$$

since the derivation above holds for each $\acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right)$. We consider the components of (B.29) to find the required asymptotic bound on $||\acute{\boldsymbol{\mu}}||$. From part (i) of this lemma, for large n we have $\Delta(\acute{P}_n) \subset \Delta(P_0)$. This means for large n, we have that for all $t \in \Delta(\acute{P}_n)$, $E_{P_0}\left[g(X;t_j)\right] = 0$. As a result,

$$\begin{aligned}
\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right) &= \sum_{j=1}^{\omega_n}\theta_j\left(\frac{1}{n}\sum_{i=1}^{n}g(X_i;t_j) - E_{P_0}\left[g(X;t_j)\right]\right) \\
\left|\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)\right| &\leq \sum_{j=1}^{\omega_n}\theta_j\left|\frac{1}{n}\sum_{i=1}^{n}g(X_i;t_j) - E_{P_0}\left[g(X;t_j)\right]\right| \\
&\leq \max_j\left|\frac{1}{n}\sum_{i=1}^{n}g(X_i;t_j) - E_{P_0}\left[g(X;t_j)\right]\right|\left(\sum_{j=1}^{\omega_n}\theta_j\right) \\
&\leq \sup_{t\in[\underline{t},\overline{t}]}\left|\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t\right) - E_{P_0}\left[g\left(\mathbf{X};t\right)\right]\right| \qquad \text{(B.30)}
\end{aligned}$$

The last line follows from the fact that $\sum_{j=1}^{\omega_n}\theta_j = 1$ by construction. The upper bound given by equation (B.30) is $o_P(1)$ uniformly in $\mathcal{P}_0(c_1, c_2)$. This follows from the functions being of Vapnik-Chervonenkis class. The moment functions g belonging to a uniformly bounded Vapnik-Chervonenkis class of functions ensures that class of functions is also uniformly Glivenko-Cantelli.

Now, for sufficiently large n, part (i) of this lemma tells us that $\Delta(\acute{P}_n) \subset \Delta(P_0)$. Therefore, part (ii) of Definition 3 implies that $\boldsymbol{\theta}^T \Sigma_{\Delta(\acute{P}_n)} \boldsymbol{\theta} \geq c_2 > 0$. Using this result and the bound from equation (B.30), we can rewrite (B.29) as

$$||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \leq \frac{o_P(1)}{c + o_P(1)} \quad \forall \acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right), \quad \text{uniformly in} \quad \mathcal{P}_0(c_1, c_2). \tag{B.31}$$

Consequently,

$$\sup_{\acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \leq \frac{o_P(1)}{c + o_P(1)}, \quad \text{uniformly in} \quad \mathcal{P}_0(c_1, c_2). \tag{B.32}$$

Therefore, $\sup_{\acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} = o_P(1)$ uniformly in $\mathcal{P}_0(c_1, c_2)$. Finally, to show that $\sup_{\acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} = O_P(n^{-1/2})$ uniformly in $\mathcal{P}_0(c_1, c_2)$, first note that the expression on the right side of (B.29) has this property. So that

$$\sup_{\acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \leq \frac{O_P(n^{-1/2})}{c + o_P(1)} \quad \text{uniformly in} \quad \mathcal{P}_0(c_1, c_2), \tag{B.33}$$

which implies

$$\sqrt{n} \sup_{\acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \leq \frac{O_P(1)}{c + o_P(1)} \quad \text{uniformly in} \quad \mathcal{P}_0(c_1, c_2). \tag{B.34}$$

Hence, $\sqrt{n} \sup_{\acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}$, a positive random variable, is bounded from above by another variable that is $O_P(1)$, uniformly in $\mathcal{P}_0(c_1, c_2)$. Therefore, we must have that

$$\sqrt{n} \sup_{\acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} = O_P(1) \quad \text{uniformly in} \quad \mathcal{P}_0(c_1, c_2).$$

□

B.4 Relationship Between $E_{\hat{P}_n}[g(\mathbf{X}, \cdot)]$ and $E_{\acute{P}_n}[g(\mathbf{X}, \cdot)]$

The following result implies that the estimator $E_{\acute{P}_n}[g(\mathbf{X}, \cdot)]$ is a uniformly consistent estimator of $E_{P_0}[g(\mathbf{X}, \cdot)]$ under H_0.

Proposition B.1. Given $[c_1, c_2] \in (0, 1] \times (0, +\infty)$ and that $P_0 \in \mathcal{P}_0(c_1, c_2)$, then

$$\sup_{t \in [\underline{t}, \bar{t}]} \left| E_{\hat{P}_n}[g(\mathbf{X}, t)] - E_{\acute{P}_n}[g(\mathbf{X}, t)] \right| = O_P(n^{-1/2}) \quad \text{uniformly over} \quad \mathcal{P}_0(c_1, c_2). \tag{B.35}$$

Proof. The proof follows the direct method. Consider the following derivation:

$$
\begin{aligned}
\left|E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right]-E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right]\right| &= \left|\sum_{i=1}^{n}\frac{1}{n}g(X_i;t)-\sum_{i=1}^{n}\acute{p}_i g(X_i;t)\right| \\
&\le \sum_{i=1}^{n}\left|\left(\frac{1}{n}-\acute{p}_i\right)g(X_i;t)\right| \\
&\le \sum_{i=1}^{n}\left|\frac{1}{n}\left(1-\frac{1}{1+\sum_{j=1}^{N}\acute{\mu}_j g(X_i;t_j)}\right)g\left(\mathbf{X}_i,t\right)\right| \\
&= \sum_{i=1}^{n}\left|\frac{1}{n}\cdot\frac{g\left(\mathbf{X}_i,t\right)\sum_{j=1}^{N}\acute{\mu}_j g(X_i;t_j)}{1+\sum_{j=1}^{N}\acute{\mu}_j g(X_i;t_j)}\right| \\
&= \sum_{i=1}^{n}\left|\acute{p}_i g\left(\mathbf{X}_i,t\right)\sum_{j=1}^{N}\acute{\mu}_j g(X_i;t_j)\right| \\
&\le \left(\frac{2\bar{t}^{s-1}}{(s-1)!}\right)^2\sum_{i=1}^{n}\acute{p}_i\left|\sum_{j=1}^{N}\acute{\mu}_j\right| \\
&\le \left(\frac{2\bar{t}^{s-1}}{(s-1)!}\right)^2\sum_{j=1}^{N}|\acute{\mu}_j| \\
&= \left(\frac{2\bar{t}^{s-1}}{(s-1)!}\right)^2 ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} && \text{(B.36)} \\
&\le \left(\frac{2\bar{t}^{s-1}}{(s-1)!}\right)^2 \sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}, && \text{(B.37)}
\end{aligned}
$$

where $\Lambda_n\left(\acute{P}_n\right)$ is the set of Lagrange multipliers on $\Delta(\acute{P}_n)$ defined in (B.18). Lemma B.5(ii) establishes

$$
\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} = O_P(n^{-1/2}) \quad \text{uniformly over} \quad \mathcal{P}_0(c_1,c_2), \tag{B.38}
$$

which implies the desired result via the inequality (B.37). □

C Technical Lemmas for Theorems 3 and 5

C.1 Theorem 3

This subsection presents two technical lemmas that are useful for proving Theorem 3. They are a consequence of the condition $P\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X}_1;t\right)<0\right]>0$ being true. The first lemma is similar to Lemma B.1, but we now do not constrain P to satisfy H_0.

Lemma C.1. Suppose $P_0 \in \mathcal{P}$ and let I_n^- be given by (B.1). Then $\lim_{n\to\infty} P\left[I_n^- \neq \emptyset\right] = 1$.

Proof. The proof follows similar steps as those in the proof of Lemma B.1. We show that the probability of the complement of $\{I_n^- \neq \emptyset\}$ converges to zero. This set is

$$\left\{I_n^- = \emptyset\right\} = \left\{\text{for each i } \exists t \in [\underline{t},\overline{t}]; g\left(\mathbf{X}_i;t\right) \geq 0\right\}.$$

By the bivariate random sampling assumption on $\{\mathbf{X}_i\}_{i=1}^n$, we have that

$$P_0\left[I_n^- = \emptyset\right] = \left(P_0\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X}_1;t\right) \geq 0\right]\right)^n \tag{C.1}$$

$$= \left(1 - P_0\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X}_1;t\right) < 0\right]\right)^n \to 0 \tag{C.2}$$

$n \to +\infty$ by Condition (i) of Assumption1. □

The second lemma concerns the existence and uniqueness of the constrained empirical likelihood probability vector $\acute{\mathbf{p}}$. Recall that $\mathcal{H}_n = \{p_i, i = 1,\ldots,n; \sum_{i=1}^n p_i = 1, p_i \geq 0, \forall i = 1,\ldots,n\}$ and that its interior is $\mathcal{H}_n^\circ$. Additionally, recall that the constraint is

$$\mathcal{H}_n^0(\mathbf{X}) = \left\{\mathbf{p} \in \mathcal{H}_n : \sum_{i=1}^n p_i g\left(\mathbf{X}_i;t\right) \leq 0 \quad \forall t \in \mathcal{T}_{N(n)}\right\}.$$

As with the previous result, we do not constraint P to satisfy H_0.

Lemma C.2. Suppose $P_0 \in \mathcal{P}$. Then

$$\lim_{n\to\infty} P_0 \left[\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right] = 1.$$

Proof. For large $n,$ Lemma C.1 implies that the event

$$\exists i \in \{1, 2, \ldots, n\} \;\; g\left(\mathbf{X}_i, t\right) < 0 \quad \forall t \in \mathcal{T}_{N(n)} \tag{C.3}$$

occurs with probability approaching 1, since $\mathcal{T}_{N(n)} \subset [\underline{t}, \overline{t}]$ for each $n.$ The rest of the proof proceeds using steps similar to those in the proof of Lemma B.2; therefore, we omit them for brevity. ☐

C.2 Theorems 5

This section presents technical lemmas for the local power analysis of the tests. It relies on the WLLN and Lindeberg-Feller Central limit Theorem for triangular arrays of row-wise IID random variables. These large sample results can be found in Section 27 of Billingsley (1995).

First, we briefly mention a few intermediate useful results regarding constrained estimation under the local alternatives.

Lemma C.3. Suppose that the conditions of Theorem 5 hold. Then

1. $\lim_{n\to+\infty} P_n \left[I_n^- \neq \emptyset\right] = 1,$ where I_n^- is defined in (B.1).
2. $\lim_{n\to+\infty} P_n \left[\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right] = 1.$
3. $\lim_{n\to+\infty} P_n \left[\mathcal{S}_n\right] = 1,$ where $\mathcal{S}_n$ is the event defined in (B.15).
4. $\lim_{n\to+\infty} P_n \left[E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \leq E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \; \forall t \in [\underline{t}, \overline{t}]\right] = 1.$
5. $\sqrt{n} \sup_{\acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} = O_{P_n}(1).$

Proof. Under the conditions of Theorem 5, the steps for proving parts 1 to 4 of this lemma are identical to their counterparts in Appendix B, but with probability computations under the local alternatives; therefore, we omit them for brevity.

We now focus on proving part 5 of this lemma. We will first show that

$$\lim_{n\to+\infty} P_n\left[\Delta(\acute{P}_n)\subset C\right]=1 \quad \text{holds, where} \quad C=\left\{t\in[\underline{t},\overline{t}]:H(t)=0\right\}.$$

The proof will follow steps similar to those of pat (i) of Lemma B.5. Proceeding by contraposition, we need to show that

$$t\notin C \implies t\notin\Delta(\acute{P}_n) \tag{C.4}$$

for large n with probability approaching unity, under the local alternatives. Part 4 of this lemma implies

$$E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right]\leq E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]=E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]-E_{P_n}\left[g\left(\mathbf{X};t\right)\right]+E_{P_n}\left[g\left(\mathbf{X};t\right)\right]. \tag{C.5}$$

Now, consider $t\notin C$. This implies that $\lim_{n\to+\infty}E_{P_n}\left[g\left(\mathbf{X};t\right)\right]=H(t)<0$. By the WLLN for triangular arrays,

$$E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]-E_{P_n}\left[g\left(\mathbf{X};t\right)\right]=o_{P_n}(1). \tag{C.6}$$

It should be noted that the application of the WLLN for triangular arrays is valid since the set of moment functions $\mathcal{F}$ is uniformly bounded. Thus for sufficiently large n, the inequality (C.5) simplifies to

$$E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right]\leq H(t)<0 \quad \text{as} \quad n\to+\infty. \tag{C.7}$$

This shows that $t\notin\Delta(\acute{P}_n)$ for large n with probability approaching unity under the local alternatives.

Using the notation of Lemma B.5, and following identical steps to those up to the inequality (B.29), we have that

$$||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}\left[\boldsymbol{\theta}^T\hat{\Sigma}_{\omega_n}\boldsymbol{\theta}-\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)\right]\leq\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)\quad\forall\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right),\tag{C.8}$$

where $\mathbf{g}_i=[g(X_i;t_1),\,g(X_i;t_2),\,\ldots,\,g(X_i;t_{\omega_n})]^T$, $\Delta(\acute{P}_n)=\{t_1,t_2,\ldots,t_{\omega_n}\}$ and $\boldsymbol{\theta}\in\mathbb{R}_+^{\omega_n}$ with $||\boldsymbol{\theta}||_{l^1_{\omega_n}}=1$. Noting that

$$\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)=\sum_{j=1}^{\omega_n}\theta_j\left(\frac{1}{n}\sum_{i=1}^{n}g(X_i;t_j)-E_{P_n}\left[g(X;t_j)\right]\right)+\sum_{j=1}^{\omega_n}\theta_j(\delta(t_j)/\sqrt{n})=o_{P_n}(1)\tag{C.9}$$

by the WLLN for triangular arrays of random variables that are row-wise IID, we have that

$$\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}\leq\frac{\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)}{\left[\boldsymbol{\theta}^T\hat{\Sigma}_{\omega_n}\boldsymbol{\theta}-\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)\right]}\tag{C.10}$$

since part 1 of this lemma implies that $\lim_{n\to+\infty}P_n\left[\boldsymbol{\theta}^T\hat{\Sigma}_{\omega_n}\boldsymbol{\theta}>0\right]=1.$

Hence,

$$\sqrt{n}\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}\leq\frac{\sqrt{n}\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)}{\left[\boldsymbol{\theta}^T\hat{\Sigma}_{\omega_n}\boldsymbol{\theta}-\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)\right]}.\tag{C.11}$$

To conclude the proof, all we need to do is to show that the numerator on the right side of the inequality (C.11) is $O_{P_n}(1)$. Noting that

$$\sqrt{n}\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)=\sum_{j=1}^{\omega_n}\sigma_n(t_j)\theta_j\left(\frac{1}{\sigma_n(t_j)\sqrt{n}}\sum_{i=1}^{n}g(X_i;t_j)-E_{P_n}\left[g(X;t_j)\right]\right)+\sum_{j=1}^{\omega_n}\theta_j\delta(t_j), \tag{C.12}$$

where we can apply the Lindeberg-Feller Central Limit Theorem for IID triangular arrays (via Lyapounov's condition) to each

$$\frac{1}{\sigma_n(t_j)\sqrt{n}}\sum_{i=1}^{n}g(X_i;t_j)-E_{P_n}\left[g(X;t_j)\right] \quad j=1,\ldots,\omega_n \tag{C.13}$$

since $\mathcal{F}$ is uniformly bounded. Hence, the first term on the right side of (C.12) is $O_{P_n}(1)$ since it is asymptotically a convex combination of $O_{P_n}(1)$ terms. Finally, part (iii) of Assumption 2 implies that the second term on the right side of (C.12) is finite. Therefore,

$$\sqrt{n}\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}=O_{P_n}(1), \tag{C.14}$$

which concludes the proof. ☐

Lemma C.4. Suppose that the conditions of Theorem 5 hold. Then

$$\lim_{n\to+\infty}P_n\left[C\subset\hat{\Delta}_n\right]=1 \quad \text{and} \quad \lim_{n\to+\infty}P_n\left[C\subset\acute{\Delta}_n\right]=1. \tag{C.15}$$

Proof. We first focus on proving $\lim_{n\to+\infty}P_n\left[\hat{\Delta}_n\neq\emptyset\right]=1$. Suppose that $t\in[\underline{t},\overline{t}]$ is such that $H(t)=0$. Consider the event $\left\{-r_n\leq E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]\leq r_n\right\}$. It is equal to

$$\left\{-\frac{\sqrt{n}r_n}{\sigma_n(t)}\leq\frac{\sqrt{n}\left(E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]-E_{P_n}\left[g\left(\mathbf{X},t\right)\right]\right)}{\sigma_n(t)}+\frac{\delta(t)}{\sigma_n(t)}\leq\frac{\sqrt{n}r_n}{\sigma_n(t)}\right\}. \tag{C.16}$$

Noting that $-\frac{\sqrt{n}r_n}{\sigma_n(t)}\to-\infty$, $\frac{\sqrt{n}r_n}{\sigma_n(t)}\to+\infty$, $\frac{\sqrt{n}\left(E_{\hat{P}_n}[g(\mathbf{X};t)]-E_{P_n}[g(\mathbf{X},t)]\right)}{\sigma_n(t)}=O_{P_n}(1)$ as $n\to+\infty$ by the Lindeberg-Feller Central Limit Theorem for IID triangular arrays, and $\delta(t)$ is a uniformly bounded function by part (iii) of Assumption 2, it follows that the probability of $\left\{t \in \hat{\Delta}_n\right\}$ tends to unity. Therefore, $\lim_{n\to+\infty} P_n\left[C \subset \hat{\Delta}_n\right] = 1$.

Now we turn our focus to proving $\lim_{n\to+\infty} P_n\left[\acute{\Delta}_n \neq \emptyset\right] = 1$. Lemma C.3 implies that the random set $\acute{\Delta}_n$ exists for large enough n, with probability approaching unity under the local alternatives. We use an identical argument to that of the first part of this lemma. Consider $t \in [\underline{t}, \overline{t}]$ such that $H(t) = 0$ and the event $\left\{-r_n \leq E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \leq r_n\right\}$. This event is equal to

$$\left\{-\frac{\sqrt{n}r_n}{\sigma_n(t)} \leq \frac{\sqrt{n}\left(E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right]\right)}{\sigma_n(t)} + \frac{\sqrt{n}\left(E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{P_n}\left[g\left(\mathbf{X}, t\right)\right]\right)}{\sigma_n(t)} + \frac{\delta(t)}{\sigma_n(t)} \leq \frac{\sqrt{n}r_n}{\sigma_n(t)}\right\}, \tag{C.17}$$

where

$$\sqrt{n}\left(E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right]\right) = -\sqrt{n}\sum_{t' \in \Delta(\acute{P}_n)} \acute{\mu}_{t'} \sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i; t'\right) g\left(\mathbf{X}_i; t\right). \tag{C.18}$$

As $\sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i; t'\right) g\left(\mathbf{X}_i; t\right)$ is a uniformly bounded function of t and t', part 5 of Lemma C.3 implies the right side of (C.18) is $O_{P_n}(1)$. Now using the result from the first part of this lemma, it follows that $\left\{t \in \acute{\Delta}_n\right\}$ tends to unity under the local alternatives, which implies the desired result. □

Lemma C.5. Suppose that the conditions of Theorem 5 hold, and let

$$B_n(v) = \left\{t \in [\underline{t}, \overline{t}] : \left|E_{P_n}\left[g\left(\mathbf{X}, t\right)\right]\right| \leq v\right\}.$$

Then, for each $\epsilon > 0$

$$\lim_{n\to+\infty} P_n\left[\hat{\Delta}_n \subset B_n((1+\epsilon)r_n)\right] = 1 \quad \text{and} \quad \lim_{n\to+\infty} P_n\left[\acute{\Delta}_n \subset B_n((1+2\epsilon)r_n)\right] = 1.$$

Proof. The proof follows the direct method. Firstly, note that the class of moment functions $\mathcal{F}$ is Donsker and pre-Gaussian, uniformly in the sequence $\{P_n\}$. This implies that the empirical process

$$\left\{\sqrt{n}\left(E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]-E_{P_n}\left[g\left(\mathbf{X},t\right)\right]\right),t\in[\underline{t},\overline{t}]\right\}$$

is asymptotically tight, uniformly in $\{P_n\}$. Therefore,

$$\lim_{n\to+\infty}P_n\left[\sup_{t\in[\underline{t},\overline{t}]}\left|E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]-E_{P_n}\left[g\left(\mathbf{X},t\right)\right]\right|>\epsilon r_n\right]=0 \tag{C.19}$$

by the choice of $r_n\to 0$ and $\sqrt{n}r_n\to+\infty$.

Now we proceed to prove the first part. Given $\epsilon>0$, suppose that $t\in\hat{\Delta}_n$. Consider the following derivation:

$$\begin{aligned}\left|E_{P_n}\left[g\left(\mathbf{X},t\right)\right]\right| &\le \left|E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right]-E_{P_n}\left[g\left(\mathbf{X},t\right)\right]\right|+\left|E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right]\right|\\ &\le \epsilon r_n+r_n=(1+\epsilon)r_n\end{aligned}$$

with probability approaching unity, under the local alternatives. Therefore,

$$\lim_{n\to+\infty}P_n\left[\hat{\Delta}_n\subset B_n((1+\epsilon)r_n)\right]=1.$$

Now we consider the reverse set inclusion. Firstly, not that part 5 of Lemma C.3 implies for each ϵ that

$$\lim_{n\to+\infty}P_n\left[\sup_{t\in[\underline{t},\overline{t}]}\left|E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]-E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right]\right|>\epsilon r_n\right]=0 \tag{C.20}$$

holds. Let $t\in\acute{\Delta}_n$, and consider the following derivation:

$$\begin{aligned}\left|E_{P_n}\left[g\left(\mathbf{X},t\right)\right]\right| \le& \left|E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right]-E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right]\right|+\left|E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right]-E_{P_n}\left[g\left(\mathbf{X},t\right)\right]\right|+\left|E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right]\right|\\ &\le \epsilon r_n+\epsilon r_n+r_n=(1+2\epsilon)r_n\end{aligned}$$

with probability approaching unity, under the local alternatives. Therefore,

$$\lim_{n \to +\infty} P_n \left[\acute{\Delta}_n \subset B_n((1+2\epsilon)r_n) \right] = 1.$$

□