THE UNIVERSITY OF SYDNEY

Economics Working Paper Series

2015 - 15

An Improved Bootstrap Test for Restricted Stochastic Dominance

Thomas M. Lok and Rami V. Tabri

16 January 2018

An Improved Bootstrap Test For Restricted Stochastic Dominance

Thomas M. Lok

Sydney Children's Hospitals Network, Sydney, New South Wales, Australia.

Rami V. Tabri

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

E-mail: rami.tabri@sydney.edu.au

Summary. This paper suggests a method for improving the performance of bootstrap tests for restricted stochastic dominance of any pre-specified order that employ a moment selection procedure: tilting the empirical distribution in the moment selection procedure. It is proposed that the amount of tilting be chosen to maximize the empirical likelihood subject to the restrictions of the null hypothesis, which are the continuum of unconditional moment inequality conditions. This constrained estimation problem is a semi-infinite program, and we propose a data-driven discretization scheme to compute its solution. We characterize sets of population distributions on which a modified test is (i) asymptotically equivalent to its non-modified version to first-order, and (ii) superior to its non-modified version according to large-sample efficiency and deficiency, and normalized deficiency. We report simulation results that show the modified versions of popular tests are noticeably less conservative than their non-modified counterparts and have improved power.

Keywords: Bootstrap Test; Contact Set; Empirical Likelihood; Semi-Infinite Program; Restricted Stochastic Dominance; Efficiency; Deficiency; Normalized Deficiency.

1. Introduction

Postulating a stochastic dominance relation of a pre-specified order between two distributions arises in many theoretical and practical situations. Examples include, the assessment of a treatment on some outcome variable, the ranking of income distributions according to general classes of poverty indices and the formulation of investment decisions (Abadie, 2002, Atkinson, 1987, and Foster and Shorrocks, 1988, Sriboonchitta et al., 2010 and the references therein). Let P_A and P_B be two marginal probability distributions from the bivariate distribution P. Then we say that distribution B, stochastically dominates distribution A at order $s \in \mathbb{Z}_+$, over the interval $[\underline{t}, \overline{t}] \subset \text{Interior}\left(\text{supp}\left(P_A\right) \cup \text{supp}\left(P_B\right)\right)$, if

$$E_P\left[\frac{\left(t - X^B\right)^{s-1}}{(s-1)!} 1\left[X^B \leq t\right] - \frac{\left(t - X^A\right)^{s-1}}{(s-1)!} 1\left[X^A \leq t\right]\right] \leq 0 \quad \forall t \in [\underline{t}, \overline{t}], \quad (1.1)$$

where supp (P_K) is the support of $P_K,\ K = A, B$. As P is not observable in practice, the verification of (1.1) is based upon a statistical test that makes use of distributions estimated from samples.

Several recent papers have proposed bootstrap tests for the conditions (1.1) (under the null) that use an estimator of the "contact set" – the set of all $t \in [\underline{t}, \overline{t}]$ such that the inequalities in (1.1) are binding. For example, Linton et al. (2010) and Donald and Hsu (2016). A major focus of these papers has been to improve upon tests that use a bootstrap critical value from the least favorable case of the model of the null hypothesis (e.g., Abadie, 2002, Barrett and Donald, 2003, and Horváth et al., 2006) because such tests are conservative. The tests they propose are asymptotically less conservative and have higher limiting local power than their least favorable counterparts.

The usefulness of the contact set approach, though, rests upon accurate estimation of the contact set under the null hypothesis. As that type of testing procedure employs the sample-analog estimator of the continuum of moments in (1.1), adjusting this unconstrained estimator by tilting the empirical distribution subject to the constraints of the null hypothesis can improve the test's finite-sample performance under the null and local alternatives. This type of adjustment is a biased-bootstrap technique, suggested by Hall and Presnell (1999) in a general context. The stochastic dominance testing literature hasn't considered this procedure because of the difficulty in imposing a continuum of inequality constraints in estimation.

This paper addresses that gap in the literature. It develops the constrained empirical likelihood estimator (Owen, 2001) of the moments, where the constraints represent the restrictions of the null hypothesis (i.e., (1.1)). Furthermore, it proposes a surgical modification of stochastic dominance tests that use the contact set approach: replace the sample-analog estimator of the moments with the constrained empirical likelihood estimator.

Computing the solution of the constrained empirical likelihood optimization problem is complicated because this optimization problem is a semi-infinite programming problem: for each sample size n, there is a continuum of inequality restrictions on random variables – one for each $t \in [\underline{t}, \overline{t}]$ – and a finite-dimensional choice variable. A common approach for solving semi-infinite programs are *discretization methods* (see, e.g., Shapiro, 2009). In such a method one selects a finite grid $\mathcal{T}_n$, $\mathcal{T}_n \subset [\underline{t}, \overline{t}]$, and solves the constrained empirical likelihood optimization problem that imposes the constraints indexed by $t \in \mathcal{T}_n$. We introduce a data-driven discretization method and show the error between its solution and the solution of the semi-infinite problem converges in probability to zero as the sample size tends to infinity, with uniformity over a collection of distributions. Moreover, we characterize these collections of distributions. To establish this large-sample result, we adapt results on discretization methods from Still (2001) to our setup and use the result of Devroye (1982) concerning the large-sample stochastic behavior of maximal uniform spacings.

The papers that employ a contact set approach to testing (1.1) under the null are Linton et al. (2010) (LSW, hereafter) and Donald and Hsu (2016). The bootstrap test of Andrews and Shi (2017) also applies to testing (1.1) and, more generally, to the setup that has many conditional moment inequalities. Andrews and Shi's model coincides with that of LSW when specialized to (1.1), which places no assumptions on the moment functions beyond the existence of mild moment conditions. By contrast, Donald and Hsu (2016) require the distributions' supports to be closed and bounded intervals, which is restrictive, because it limits the scope of applications (e.g., finance). Hence, this paper omits a discussion of Donald and Hsu's test. It also adopts the statistical framework of LSW and defers a discussion of Andrews and Shi's testing procedure to Appendix C, for ease of exposition.

Interestingly, the LSW test and its modified version are asymptotically equivalent to first-order. However, the modified test confers several advantages over its non-modified version: (i) it is asymptotically less conservative; (ii) has higher large-sample local power, and (iii) is superior according to the criteria: large-sample *efficiency* (Pitman, 1948) and *deficiency* (Hodges and Lehmann, 1970), and *normalized deficiency* (Akahira, 1999), along $\sqrt{n}$ local alternatives. The comparisons using deficiency and normalized deficiency, are more refined than efficiency in large-samples because they discriminate tests on the basis of higher-order asymptotics (e.g., Albers, 1975).

We report Monte Carlo simulation results that compare the LSW and Andrews and Shi tests with their respective modified versions. The simulations use the experimental designs in Section 5 of LSW. Overall, the simulation results verify the superior performance of our method, and that there can be substantial differences in their relative performance.

The literature related to this paper includes numerous papers on inference with unconditional moment inequality models, where there is only a finite number of such inequalities and the parameter of interest in finite-dimensional. For example, Andrews and Soares (2010), Andrews and Guggenberger (2009, 2010), and Canay (2010), among others. In contrast, the parameter of interest in this paper is infinite-dimensional and there is a continuum of inequality restrictions.

The literature on shape-constrained estimation via tilting the empirical distribution, overlaps with this paper. It focuses on nonparametric density and regression estimation, for example, Hall and Huang (2001, 2002), Carroll et al. (2011), and Du et al. (2013) among others, where enforcement of the constraints is on a predesignated grid of points which the practitioner must set. In contrast, the constrained estimation method this paper proposes is for distributions and uses a data-driven grid of points. To the best of our knowledge, our paper is the first in that literature to propose a data-driven grid, albeit for stochastic dominance constraints.

There are also tests for restricted stochastic dominance that posit a null and alternative of non-dominance and restricted dominance, respectively. For example: Berger (1988), Davidson and Duclos (2013), and lvarez Esteban et al. (2017). 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 do complement each other.

We organize the paper as follows. Section 2 illustrates the tests' relative performance using a simple example. Section 3 introduces the statistical framework of LSW. Section 4 introduces the constrained empirical likelihood optimization problem, the data-driven discretization method and its asymptotic properties, and the proposed contact set estimator. Section 5 presents the asymptotic null properties of the proposed contact set estimator and modified LSW test. Section 6 present the asymptotic power properties of the modified LSW test. Section 7 reports the findings of Monte Carlo simulation experiments. Section 8 concludes and Section 9 collates the acknowledgements of the individuals and institutions who provided help during the research.

2. Illustrative Example

This section illustrates the inference results of this paper in the context of testing on a bivariate Gaussian mean. The hypothesis testing problem is an example of what Sen and Silvapulle (2001) refer to as a Type B testing problem in Section 3.2 of their book.

2.1. Statistical Model, Testing Problem, and Test Statistic

Let $\mathbf{X} = [X_1, X_2] \sim N(\mu, \Omega)$, where $\mu = [\mu_1, \mu_2]$ and Ω is the 2×2 correlation matrix with correlation coefficient ρ. Denote by $\mathcal{M}$ the statistical model consisting of the set of all Gaussian distributions P of $\mathbf{X}$ that satisfy the following assumption.

Assumption 2.1. *(i)* $\{\mathbf{X}_i\}_{i=1}^n$ *i.i.d.* P, *(ii)* $\mu \in \mathbb{R}^2$, *and (iii)* $\rho_0 = 1/2$.

We set ρ_0 as positive to mimic the inherent correlational structure of the moment functions in (1.1) that are indexed by the contact set, which we discuss in Section 3.

The hypothesis testing problem of interest in this example is

$$H_0 : \mu_1 \leq 0 \text{ and } \mu_2 \leq 0 \quad \text{versus} \quad H_1 : \text{either } \mu_1 > 0 \text{ or } \mu_2 > 0 \text{ or both.} \tag{2.1}$$

The model of the null hypothesis is defined as $\mathcal{M}_0 = \{P \in \mathcal{M} : \mu_j \leq 0, j = 1, 2\}$.

Following Example 1 of LSW, consider the test statistic $T_n = \max\{\sqrt{n}\hat{\mu}_1, \sqrt{n}\hat{\mu}_2\}$, where $\hat{\mu}_j = n^{-1}\sum_{i=1}^n X_{ji}$ for $j = 1, 2$. We select this statistic to illustrate the results of our paper because the rejection probabilities under the null and local alternatives can be computed numerically. Let $F_{\mu,\Omega_0}(\cdot,\cdot)$ denote the cumulative distribution function (CDF) of $N(\mu, \Omega_0)$, where Ω_0 is the correlation matrix with $\rho = \rho_0$. The CDF of T_n is $F_{\sqrt{n}\mu,\Omega_0}(u, u)$, for $u \in \mathbb{R}$.

The pointwise-asymptotic null distribution of T_n is non-degenerate provided the contact set, $\Delta(P) = \{j \in \{1, 2\} : \mu_j = 0\}$, is nonempty:

$$\lim_{n\to+\infty} F_{\sqrt{n}\mu,\Omega_0}(u, u) = \begin{cases} F_{\mathbf{0},\Omega_0}(u, u) & \text{if } \mu_1 = \mu_2 = 0 \\ \Phi(u) & \text{if } \mu_1 = 0, \mu_2 < 0 \\ \Phi(u) & \text{if } \mu_1 < 0, \mu_2 = 0, \\ 1, & \text{if } \mu_1, \mu_2 < 0, \end{cases} \quad \forall u \in \mathbb{R}, \tag{2.2}$$

where $\mathbf{0} \in \mathbb{R}^2$ denotes the zero vector and $\Phi(\cdot)$ is the CDF of $N(0, 1)$. Hence, T_n exhibits a discontinuity in its asymptotic null distribution as a function of the underlying distribution, P, with respect to the topology of weak convergence. This type of asymptotic behavior motivates the use of generalized moment selection (GMS) testing procedures.

2.2. *Testing Procedures*

This section introduces a GMS testing procedure that is a special case of the procedure Andrews and Soares (2010) introduce, and its modified version that is based on the proposal of this paper. The former test is defined as $\hat{\tau}_n = 1\left[T_n > \hat{c}\right]$, where $1[\cdot]$ is the indicator function, and $\hat{c}$ is the GMS critical value. This critical value depends on a localization parameter through the GMS function. This parameter is a sequence $\{r_n\}_n$ of positive numbers such that (i) $r_n \to 0$ and (ii) $\sqrt{n} r_n \to +\infty$, as $n \to +\infty$. The GMS function is the vector $\hat{\varphi} = [\hat{\varphi}_1, \hat{\varphi}_2]$ whose components are defined as follows

$$\hat{\varphi}_j = \begin{cases} 0 & \text{if } |\hat{\mu}_j| < r_n \\ +\infty & \text{if } |\hat{\mu}_j| \geq r_n \end{cases} \qquad j = 1, 2.$$

The GMS critical value is defined as

$$\hat{c} = \begin{cases} \inf\left\{u \in \mathbb{R} : F_{\mathbf{0},\Omega_0}\left(u + \hat{\varphi}_1, u + \hat{\varphi}_2\right) \geq 1 - \alpha\right\}, & \text{if } \hat{\varphi} \neq [+\infty, +\infty] \\ \inf\left\{u \in \mathbb{R} : F_{\mathbf{0},\Omega_0}\left(u, u\right) \geq 1 - \alpha\right\}, & \text{otherwise,} \end{cases}$$

where α is the given nominal level.

The modified version of $\hat{\tau}_n$ is $\tilde{\tau}_n = 1\left[T_n > \tilde{c}\right]$, where the critical value $\tilde{c}$ is defined in exactly the same way as $\hat{c}$, except that the constrained maximum likelihood estimator of μ,

$$\tilde{\mu} = \arg\max\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{X}_i - \mu\right)' \Omega^{-1}\left(\mathbf{X}_i - \mu\right); \, \mu \in \mathbb{R}^2_-\right\}, \tag{2.3}$$

replaces the estimator $\hat{\mu}$ in the GMS function $\hat{\varphi}$. The constraints in the definition of $\tilde{\mu}$ are the restrictions of the null hypothesis in (2.1).

2.3. *Relative Performance of Tests Under The Null and $\sqrt{n}$ Local Alternatives*

For ease of exposition, Appendix B.1.1 shows the tests' exact sizes are approximately equal to $\alpha = 0.05$ and are within 0.0003 of each other, using $r_n = \sqrt{\log(n)/n}$ and sample sizes $n = 100, 101, 102, \ldots, 10000$. Therefore, when $\alpha = 0.05$, it is feasible to compare their rejection probabilities along local alternatives for $n \geq 100$ without adjusting them so that they have the same exact size.

Consider local alternatives $\{P_n\}_n \subset \mathcal{M} - \mathcal{M}_0$ in which P_n satisfies the hypothesis

$$H_n : \mu_{1n} = -0.1 \quad \text{and} \quad \mu_{2n} = 2/\sqrt{n}, \quad \forall n. \tag{2.4}$$

Figure 1 reports the local powers of the tests for $n = 100, 101, 102, \ldots, 7000$. These numerical results indicate that the local powers satisfy $\lim_{n\to+\infty} E_{P_n}\hat{\tau}_n = \lim_{n\to+\infty} E_{P_n}\tilde{\tau}_n = 1 - \Phi(c-2) \approx 0.6388$, where $c = \Phi^{-1}(1-\alpha)$ is the common probability limit of $\hat{c}$ and $\tilde{c}$, and $E_{P_n}\hat{\tau}_n < E_{P_n}\tilde{\tau}_n \; \forall n$.

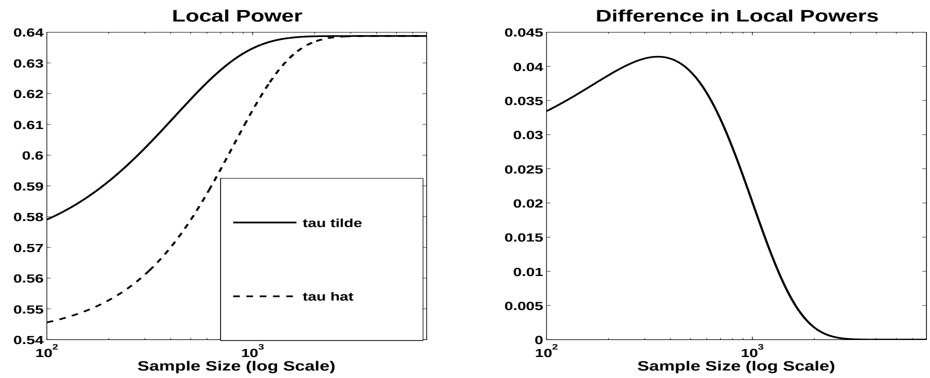

Fig. 1. The left panel plots the local powers $E_{P_n}\hat{\tau}_n$ and $E_{P_n}\tilde{\tau}_n$. The right panel plots the difference $E_{P_n}\tilde{\tau}_n - E_{P_n}\hat{\tau}_n$.

Table 1. Relative Performance Criteria

	Efficiency	Deficiency	$\sqrt{n}$-Normalized Deficiency
Finite-sample	k_n/n	$d_n = k_n - n$	$d_n/\sqrt{n}$
Asymptotic	$\lim_{n\to+\infty}(k_n/n)$	$\lim_{n\to+\infty} d_n$	$\lim_{n\to+\infty}(d_n/\sqrt{n})$

In consequence, the equality of the limiting local power functions does not reflect the large-sample situation. In this case, the standard practice is to compare the tests using the criteria: asymptotic efficiency and deficiency. Furthermore, when these criteria fail to separate the tests, normalized deficiency suitably compares them. In this numerical example, asymptotic efficiency and deficiency of $\hat{\tau}_n$ relative to $\tilde{\tau}_n$, do not separate the tests as they are equal to unity and $+\infty$, respectively. We show $\sqrt{n}$-normalized deficiency of $\hat{\tau}_n$ relative to $\tilde{\tau}_n$, separates the tests. These criteria are functions of $k_n = \min\{N \in \mathbb{Z}_+ : E_{P_m}\hat{\tau}_n \geq E_{P_n}\tilde{\tau}_n \, \forall m \geq N\}$, which is the minimal sample size $\hat{\tau}_n$ needs to attain the local power that is at least as large as the local power of $\tilde{\tau}_n$ at sample size n. Table 1 defines these criteria. A value $(k_n/n) > 1$ or $d_n > 0$ indicates that $\tilde{\tau}_n$ is superior to $\hat{\tau}_n$, because $\hat{\tau}_n$ requires more observations than n to achieve at least the same local power of $\tilde{\tau}_n$.

Figure 2 reports the numerical values of the criteria described Table 1. The leftmost panel indicates $k_n/n > 1 \; \forall n \geq 100$, and that it is decreasing to unity with the sample size. Therefore, asymptotic efficiency does not reflect the relative behavior of the tests' local powers in large-samples, suggesting that it is insensitive to small power differences in moderate and large sample sizes. The middle panel

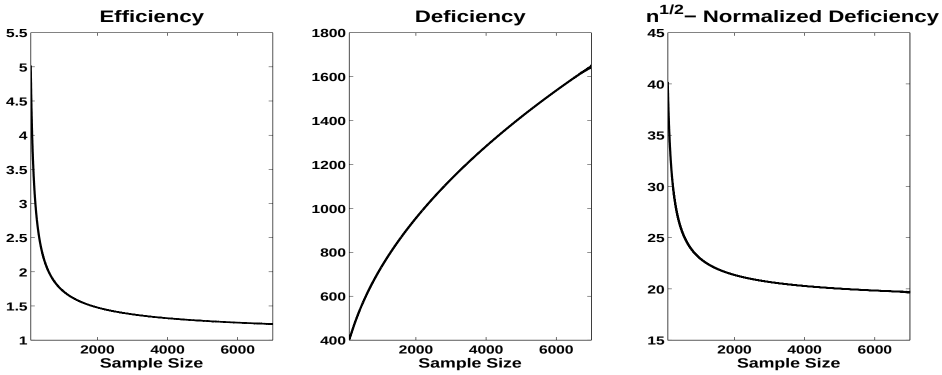

Fig. 2. The panels, from left to right, plot the efficiency k_n/n, deficiency d_n, and normalized deficiency $d_n/\sqrt{n}$ of $\hat{\tau}_n$ with respect to $\tilde{\tau}_n$, respectively.

indicates that $d_n > 0\ \forall n$ and diverges to $+\infty$, albeit slowly. The rightmost panel reports the $\sqrt{n}$-normalized deficiency criterion, indicating that $\lim_{n\to+\infty}(d_n/\sqrt{n}) \approx 19$.

The conclusion from using normalized deficiency is the following. In large sample sizes, the number of additional observation needed to compensate for the amount by which $E_{P_n}\hat{\tau}_n$. falls short of $E_{P_n}\tilde{\tau}_n$ is approximately $19\sqrt{n}$. Note that such a result is typically much stronger than the mere assertion that k_n/n tends to 1 as $n \to +\infty$. This result is not unusual because the local alternatives (2.4) become progressively more difficult to detect as more observations are available. Results of this sort exist in the statistics literature; see, for example, Albers (1975) for results similar to our numerical findings, but in the context of the symmetry problem.

2.4. Synthesis

The marriage of constrained estimation and Condition (iii) of Assumption 2.1 (i.e., positive correlation) produces the relative asymptotic behavior of the tests $\hat{\tau}_n$ and $\tilde{\tau}_n$ described in the previous sections. From the characterization of $\tilde{\mu}$, given by (B.1) in Appendix B.1.2, this marriage yields

$$\tilde{\mu}_j \leq \hat{\mu}_j \quad \text{for} \quad j = 1, 2, \tag{2.5}$$

with probability one under any $P \in \mathcal{M}$. Hence, there are sample realizations that satisfy $\{\tilde{\mu}_{j_1} \leq -r_n < \hat{\mu}_{j_1} < 0\} \cap \{0 < \hat{\mu}_{j_2} < r_n\}$, under the null and local alternatives of the form (2.4). These realizations imply $\tilde{c} < \hat{c}$ occurs because the modified GMS function omits and retains the dimension j_1 and j_2, respectively, to compute $\tilde{c}$, whereas $\hat{\varphi}$ retains both dimensions to compute $\hat{c}$. Thus, when $\mu_j < 0$ holds (under the null or alternative) $\tilde{\tau}_n$ detects this configuration more easily than $\tilde{\tau}_n$ because $\tilde{\mu}_j$ is a more reliable estimator than $\hat{\mu}_j$, and therefore, takes it into account by delivering a critical value that is suitable for the case where this moment inequality is omitted.

While the illustration focuses on local alternatives (2.4), the superiority of $\tilde{\tau}_n$ over $\hat{\tau}_n$ also holds in directions where $\lim_{n\to+\infty}\mu_{jn} = 0$ for $j = 1, 2$. In these directions

k_n does not exist in large-samples because the local powers are decreasing to their common limit; see Appendix B.1.3 for a numerical illustration.

The results of this section carry over to the more complicated setup of restricted stochastic dominance when combining constrained empirical likelihood estimation with the non-negative correlational structure of the moment functions indexed by the contact set. That setup is more complicated because the statistical model is nonparametric and there is a continuum of moment inequality restrictions that characterize the null hypothesis. Consequently, the discontinuity in the asymptotic null distribution of conventional test statistics is more complex than (2.2), and can have a profound effect on the relative performance of the LSW test and its modified version.

3. Setup

3.1. Statistical Model, Null Hypothesis and A Property of the Moment Functions

Let P_0 denote the "true" distribution of 2-dimensional random vector $\mathbf{X} = [X^A, X^B]$. For ease of exposition, let $\{\mathbf{x} \mapsto g(\mathbf{x}; t), t \in [\underline{t}, \overline{t}]\}$ denote the set of moment functions in (1.1), where

$$g(\mathbf{x}; t) = \frac{(t - x^B)^{s-1}}{(s-1)!} 1\left[x^B \leq t\right] - \frac{(t - x^A)^{s-1}}{(s-1)!} 1\left[x^A \leq t\right].$$

Implicit in this notation for the moment functions is the order of stochastic dominance, which is fixed by the null hypothesis. Given $s \in \mathbb{Z}_+$ and the interval $[\underline{t}, \overline{t}]$, the testing problem of main interest takes the following form:

$$H_0 : E_{P_0}[g(\mathbf{X}; t)] \leq 0 \, \forall t \in [\underline{t}, \overline{t}] \text{ versus } H_1 : \exists t \in [\underline{t}, \overline{t}] \text{ such that } E_{P_0}[g(\mathbf{X}; t)] > 0.$$

The statistical model this paper considers is denoted by $\mathcal{P}$. It is the set of all potential continuous distributions of $\mathbf{X}$ that satisfies the following assumption.

Assumption 3.1. *(i)* $[\underline{t}, \overline{t}] \subset$ *Interior* $(supp(P_A) \cup supp(P_B))$; *(ii)* $supp(P) \subseteq \mathbb{R}^2$; *(iii)* $\{\mathbf{X}_i\}_{i=1}^n$ *is i.i.d.* P, *and (iv)* $\sup_{P \in \mathcal{P}} E_P\left[|X^K|^{2((s-1)\vee 1)+\delta}\right] < +\infty$ *for* $K = A, B$, *for some* $\delta > 0$.

Define $\mathcal{P}_0 = \left\{P \in \mathcal{P} : E_P[g(\mathbf{X}; t)] \leq 0 \quad \forall t \in [\underline{t}, \overline{t}]\right\}$. This paper characterizes submodels of $\mathcal{P}_0$ that serve as models of the null hypothesis for which the proposed testing procedure has asymptotically exact size and is asymptotically similar in the sense of Definition 1 of LSW. We repeat this definition here for convenience.

Definition 3.1. *Suppose that* Ω *is the model of the null hypothesis. (i) A test* τ *with a nominal level* α *is said to have an* asymptotically exact size *if there exists a nonempty subset* $\Omega' \subset \Omega$ *such that:*

$$\limsup_{n \to +\infty} \sup_{P \in \Omega} E_P \tau \leq \alpha, \quad \textit{and} \tag{3.1}$$

$$\limsup_{n \to +\infty} \sup_{P \in \Omega'} |E_P \tau - \alpha| = 0. \tag{3.2}$$

(ii) When a test τ satisfies (3.2), we say that the test is asymptotically similar *on Ω'.*

At the heart of the desirable properties that emanate from the proposed modification is the combination of constrained estimation and the inherent non-negative covariance structure of $\{g(\mathbf{X};t) : t \in \Delta(P)\}$, where $\Delta(P) = \{t \in [\underline{t}, \overline{t}] : E_P[g(\mathbf{X};t)] = 0\}$ is the contact set with respect to $P \in \mathcal{P}$. The non-negative covariance structure is due to the following property of $\left\{\mathbf{x} \mapsto g(\mathbf{x};t),\ t \in [\underline{t}, \overline{t}]\right\}$.

Property 1. *For each $s \in \mathbb{Z}_+$, the class of functions $\left\{\mathbf{x} \mapsto g(\mathbf{x};t),\ t \in [\underline{t}, \overline{t}]\right\}$ 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. It is a consequence of the moment functions having the form $g(\mathbf{x};t) = h(x^B;t) - h(x^A;t)$, where $h(x^k;t)$ is weakly monotonically increasing in its second argument for a given $x^k \in \text{supp}(P_K)$, $K = A, B$. This property of the moment functions implies that for any $P \in \mathcal{P}$, $\text{Cov}_P(g(\mathbf{X};t_1), g(\mathbf{X};t_2)) = E_P[g(\mathbf{X};t_1)\, g(\mathbf{X};t_2)] \geq 0\ \forall t_1, t_2 \in \Delta(P)$.

3.2. Test Statistic, Asymptotic Theory and LSW Bootstrap Procedure

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 $\hat{T}_n = n \int_{\underline{t}}^{\overline{t}} \max\{E_{\hat{P}_n}[g(\mathbf{X};t)], 0\}^2\, dt$, where $\hat{P}_n = n^{-1} \sum_{i=1}^n \delta_{\mathbf{X}_i}$ is the empirical measure based on the random sample $\{\mathbf{X}_i\}_{i=1}^n$, and $E_{\hat{P}_n}$ denotes the expectation under $\hat{P}_n$. Theorem 1 of LSW establishes the pointwise-asymptotic null distribution of $\hat{T}_n$ using the Donsker property of $\left\{\mathbf{x} \mapsto g(\mathbf{x};t), t \in [\underline{t}, \overline{t}]\right\}$ for each $s \in \mathbb{Z}_+$:

$$\hat{T}_n \xrightarrow{d} \begin{cases} \int_{\Delta(P)} \max\{\nu(t), 0\}^2\, dt, & \text{if } P \in \mathcal{P}_{00}, \\ 0, & \text{if } P \in \mathcal{P}_0 - \mathcal{P}_{00}, \end{cases} \tag{3.3}$$

where $\mathcal{P}_{00} = \left\{P \in \mathcal{P}_0 : \int_{\Delta(P)} dt > 0\right\}$, and $\nu(\cdot)$ is a zero-mean Gaussian process on $[\underline{t}, \overline{t}]$ with covariance kernel $C(t_1, t_2) = \text{Cov}_P(g(\mathbf{X};t_1), g(\mathbf{X};t_2))$, for $t_1, t_2 \in \Delta(P)$.

The asymptotic null distribution of $\hat{T}_n$ depends on the form of contact set $\Delta(P_0)$. Hence, it exhibits a discontinuity in the underlying probability P that generates the data. The consequence of this large sample behavior of $\hat{T}_n$ is that it invalidates the use of the canonical bootstrap for testing H_0 (e.g., Andrews, 2000). For this reason, LSW propose a bootstrap testing procedure that uses a contact set estimator. Their contact set estimator is defined as

$$\hat{\Delta}_n = \{t \in [\underline{t}, \overline{t}] : |E_{\hat{P}_n}[g(\mathbf{X};t)]| < r_n\}, \tag{3.4}$$

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 procedure follows these steps. Using the data, compute $\hat{P}_n$ and $\hat{T}_n$. Then draw a random sample of size n, $\{\mathbf{X}^\star_{i,l}\}_{i=1}^n$, for $l = 1, \ldots, B$, using resampling with replacement from $\hat{P}_n$. Then for each bootstrap sample, compute the bootstrap test statistic as follows:

$$\hat{T}^\star_n = \begin{cases} \int_{\underline{t}}^{\overline{t}} \left(\max\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[g\left(\mathbf{X}^\star_{i,l}; t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right], 0 \right\}\right)^2 dt, & \text{if } \int_{\hat{\Delta}_n} dt = 0, \\ \int_{\hat{\Delta}_n} \left(\max\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[g\left(\mathbf{X}^\star_{i,l}; t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right], 0 \right\}\right)^2 dt, & \text{if } \int_{\hat{\Delta}_n} dt > 0, \end{cases} \tag{3.5}$$

where $\hat{\Delta}_n$ is defined in (3.4). After that, compute the approximate bootstrap p-value $\hat{\Upsilon}_B = B^{-1} \sum_{l=1}^B 1[\hat{T}^\star_{n,l} \geq \hat{T}_n]$. Finally, reject H_0 if $\hat{\Upsilon}_B \leq \alpha$, where $\alpha \in (0, 1/2)$ is a given nominal level.

The modification of the LSW test this paper proposes follows the steps of their bootstrap procedure, but replaces $\hat{\Delta}_n$ with a different set estimator of $\Delta(P_0)$. As their contact set estimator (3.4) employs the empirical measure $\hat{P}_n$, which is the unrestricted empirical likelihood estimator of P_0, this paper proposes to replace $\hat{P}_n$ with the constrained empirical likelihood estimator of P_0 that imposes the restrictions of the null hypothesis. The next section introduces this procedure and the contact set estimator based upon it.

4. Empirical Likelihood And Contact Set Estimation

Consider the following constrained empirical likelihood optimization problem:

$$\tilde{\mathbf{p}} = \arg\max \left\{ \sum_{i=1}^n \log\left(p_i\right); \sum_{i=1}^n p_i g\left(\mathbf{X}_i; t\right) \leq 0 \, \forall t \in [\underline{t}, \overline{t}], \sum_{i=1}^n p_i = 1, \, p_i \geq 0 \, \forall i \right\}, \tag{4.1}$$

where $\tilde{\mathbf{p}} = [\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_n] \in \mathbb{R}^n$. Given a sample size, n, there is a continuum of constraints being imposed in estimation and there is a finite number of choice variables. This feature of the optimization problem classifies it as a (strictly concave) semi-infinite programming problem (SIP) with a random constraint set.

For the case $s = 1$, the optimization problem (4.1) is, in fact, a finite programming problem: there is a finite subset of the constraints $\sum_{i=1}^n p_i g\left(\mathbf{X}_i; t\right) \leq 0$ $\forall t \in [\underline{t}, \overline{t}]$ for which $\tilde{\mathbf{p}}$ is also the solution of the optimization problem that replaces this continuum of inequality constraints with that finite subset. However, it is, in general, not possible to find a such a finite subset of the continuum of inequality constraints for orders of stochastic dominance $s > 1$.

We propose a data-driven discretization scheme for the SIP problem (4.1). For $s > 1$, the solution of the discretized SIP problem approximates the SIP's solution $\tilde{\mathbf{p}}$, and in the case of $s = 1$ it is equal to $\tilde{\mathbf{p}}$. The discretization scheme uses a finite subset of the constraints $\sum_{i=1}^{n} p_i g(\mathbf{X}_i; t) \leq 0 \ \forall t \in [\underline{t}, \overline{t}]$ whose index set, $\mathcal{T}_n = \{t_{(j)}\}_{j=0}^{N}$, comprises the order statistics of $\left\{X_i^A, X_i^B\right\}_{i=1}^{n} \cap (\underline{t}, \overline{t}) \cup \left\{\underline{t}, \overline{t}\right\}$, where $t_{(0)} = \underline{t}$ and $t_{(N)} = \overline{t}$. The solution of the discretized SIP problem is defined as

$$\acute{\mathbf{p}} = \arg\max\left\{\sum_{i=1}^{n}\log(p_i)\,;\sum_{i=1}^{n}p_i g(\mathbf{X}_i;t) \leq 0\,\forall t \in \mathcal{T}_n, \sum_{i=1}^{n}p_i = 1, p_i \geq 0\,\forall i\right\}. \quad (4.2)$$

As the discretization is data-driven, there is an additional layer of randomness that must be accounted for in deriving the subsequent large-sample results.

Next, we characterize subsets of probabilities in $\mathcal{P}$ under which the optimization problems in (4.1) and (4.2) are well-posed, with uniformity.

DEFINITION 4.1. *Let $f(x^A, x^B)$ denote the joint density function corresponding to a joint distribution $P \in \mathcal{P}$. For each $c_1 \in (0,1)$ and $c_2 \in \left(0, (\overline{t}-\underline{t})^{-2}\right)$ define the sets of probabilities*

$$\mathcal{P}_1(c_1) = \left\{P \in \mathcal{P} : P\left[\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X};t) < 0\right] \geq c_1\right\} \quad \textit{and} \quad (4.3)$$

$$\mathcal{P}_2(c_2) = \left\{P \in \mathcal{P} : \inf_{(x^A,x^B)\in[\underline{t},\overline{t}]\times[\underline{t},\overline{t}]} f(x^A, x^B) \geq c_2\right\}. \quad (4.4)$$

The set $\mathcal{P}_1(c_1)$ restricts $\mathcal{P}$ by excluding distributions that are arbitrarily close to distributions that place zero probability on the event $\{\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X};t) < 0\}$. This condition begets the uniform asymptotic existence of the probabilities $\tilde{\mathbf{p}}$ and $\acute{\mathbf{p}}$ and Lagrange multipliers for these optimization problems that solve Karush-Kuhn-Tucker conditions. The set $\mathcal{P}_2(c_2)$ also restricts $\mathcal{P}$, but by excluding distributions whose joint densities are arbitrarily close to zero on the square $[\underline{t},\overline{t}] \times [\underline{t},\overline{t}]$. This condition begets the uniform convergence of the sequence $\{\mathcal{T}_n\}_{n\geq 1}$ to the the interval $[\underline{t},\overline{t}]$ in the Hausdorff metric. See Appendices B.2.1 and B.2.2 for a further discussion of these sets.

We have the following result.

THEOREM 4.1. *Let $\|\cdot\|$ denote the Euclidean metric on $\mathbb{R}^n$. The following statements hold.*

(a) Let $s = 1$. For any $c_1 \in (0,1)$, $\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)} P[\acute{\mathbf{p}} = \tilde{\mathbf{p}}] = 1$.

(b) Let $s > 1$. For each $c_1 \in (0,1), c_2 \in \left(0, (\overline{t}-\underline{t})^{-2}\right)$ and for each $\epsilon > 0$,

$$\lim_{n\to+\infty} \sup_{P\in\mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2)} P[\|\acute{\mathbf{p}} - \tilde{\mathbf{p}}\| > \epsilon] = 0.$$

PROOF. See Appendix D.1.

Part (a) of Theorem 4.1 indicates that for $s = 1$ and large enough n, the vector of probabilities $\tilde{\mathbf{p}}$ and $\acute{\mathbf{p}}$ coincide with uniformity over sets of probabilities $\mathcal{P}_1(c)$. Part

(b) of Theorem 4.1 establishes the uniform convergence of the error between $\tilde{\mathbf{p}}$ and $\acute{\mathbf{p}}$ to zero over sets of probabilities $\mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2)$, for any $s>1$.

The computational cost of $\acute{\mathbf{p}}$ is rather low, even when n is large. For brevity, we relegate to Appendix B.2.3 a discussion of this point and show how to improve the numerical accuracy of the optimization problem (4.2).

The contact set estimator this paper proposes is defined as

$$\acute{\Delta}_n = \{t \in [\underline{t},\overline{t}] : |E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right]| < r_n\}, \tag{4.5}$$

where $\acute{P}_n = \sum_{i=1}^n \acute{p}_i \delta_{X_i}$ and $\acute{\mathbf{p}}$ is given by (4.2). As the empirical measure $\hat{P}_n$ is the solution of the empirical likelihood problem that does not impose the continuum of inequality constraints, the estimator of the moments it engenders $E_{\hat{P}_n}\left[g\left(\mathbf{X};\cdot\right)\right]$ does not necessarily satisfy the restrictions of the null hypothesis. By contrast, the definition of $\acute{P}_n$ implies $E_{\acute{P}_n}\left[g\left(\mathbf{X};\cdot\right)\right]$ satisfies the constraints (1.1) when $s=1$, and approximately satisfies these constraints when $s>1$, but with the approximation error disappearing asymptotically, with uniformity.

The next section presents the uniform asymptotic null properties of the contact set estimator $\acute{\Delta}_n$, and the asymptotic size properties of the proposed testing procedure which uses it instead of $\hat{\Delta}_n$ in the LSW bootstrap procedure.

5. Asymptotic Null Properties

5.1. Behavior of The Contact Set Estimator

The following definition characterizes subsets of $\mathcal{P}_0$ on which we establish the uniform asymptotic properties of $\acute{\Delta}_n$ under the null hypothesis.

DEFINITION 5.1. *Let $s\in\mathbb{Z}_+$. For each $c_1\in(0,1), c_2\in\left(0,(\overline{t}-\underline{t})^{-2}\right)$ and $c_3\in(0,+\infty)$,*

$$\mathcal{P}_0^s(c_1,c_2,c_3) = \begin{cases} \{P\in\mathcal{P}_0\bigcap\mathcal{P}_1(c_1) : \inf\limits_{t\in\Delta(P)} E_P\left[g^2\left(\mathbf{X};t\right)\right]\geq c_3\}, & \textit{if } s=1 \\ \{P\in\mathcal{P}_0\bigcap\mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2) : \inf\limits_{t\in\Delta(P)} E_P\left[g^2\left(\mathbf{X};t\right)\right]\geq c_3\}, & \textit{if } s>1. \end{cases} \tag{5.1}$$

Definition 5.1 distinguishes the cases $s=1$ and $s>1$ because there is an additional layer of randomness due to the discretization scheme that affects the asymptotic behavior of $\acute{\Delta}_n$ when $s>1$ which does not arise in the case $s=1$. The defining condition $\inf_{t\in\Delta(P)} E_P\left[g^2\left(\mathbf{X};t\right)\right]\geq c_3$ in this Definition restricts $\mathcal{P}_0$ by excluding probabilities whose variances of the moment functions that are indexed by the contact set are arbitrarily close to zero. It begets the convergence in probability to zero of the Lagrange multipliers from the problem in (4.2), uniformly over $\mathcal{P}_0^s(c_1,c_2,c_3)$. For brevity, we relegate the formal statement and proof of this technical intermediate result to Appendix F. This type of condition is common in models with a finite-dimensional parameter and a finite number of moment inequality restrictions for achieving uniform asymptotic validity of pure and Generalized Empirical Likelihood tests using the dual approach. See, for example, Condition

(iii) of Definition 3.1 in Canay (2010), and Condition (b) of Assumption GEL in Andrews and Guggenberger (2009).

We have the following result.

THEOREM 5.1. *For each $s \in \mathbb{Z}_+$, let the contact set estimators $\hat{\Delta}_n$ and $\acute{\Delta}_n$ be given by (3.4) and (4.5), respectively, and $\mathcal{P}_{000} = \{P \in \mathcal{P}_0 : \Delta(P) \neq \emptyset\}$. For each $s \in \mathbb{Z}_+$, $c_1 \in (0,1), c_2 \in \left(0, (\bar{t} - \underline{t})^{-2}\right)$ and $c_3 \in (0, +\infty)$, let $\mathcal{P}_0^s(c_1, c_2, c_3)$ be given by (5.1) in Definition 5.1, and the following statements hold.*

(a) $\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)} P\left[E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right] \leq E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]\ \forall t \in [\underline{t}, \bar{t}]\right] = 1$. *Moreover,* $\{\acute{P}_n \neq \hat{P}_n\} \implies \{E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right] < E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]\ \forall t \in [\underline{t}, \bar{t}]\}$.

(b) $\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)} P\left[\Delta\left(P\right) \subseteq \acute{\Delta}_n \subseteq \hat{\Delta}_n\right] = 1$.

(c) $\lim_{n\to+\infty} \sup_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)} P\left[\hat{\Delta}_n \subsetneq \acute{\Delta}_n\right] = 0$.

(d) $\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)\bigcap\mathcal{P}_{000}} P\left[\acute{\Delta}_n \subsetneq \hat{\Delta}_n\right] \geq \frac{1}{2}$.

(e) $\forall \epsilon > 0$, $\lim_{n\to+\infty} \inf_{P\in\mathcal{P}} P\left[\hat{\Delta}_n \subseteq \left\{t \in [\underline{t}, \bar{t}] : |E_P\left[g\left(\mathbf{X};t\right)\right]| \leq (1+\epsilon) r_n\right\}\right] = 1$.

PROOF. See Appendix D.2.

REMARK 5.1. *Part (a) of Theorem 5.1 is the consequence of the marriage between Property 1 of the moment function and constrained empirical likelihood estimation introduced in Section 4. This result mirrors the inequalities (2.5) arising in the illustrative example, and is essential in the proofs of Parts (b) - (d) of Theorem 5.1.*

REMARK 5.2. *Parts (b) and (c) of Theorem 5.1 imply the LSW bootstrap test statistic $T_n^{\star}$, described in (3.5), weakly dominates its modified counterpart stochastically at the first-order, conditional on the sample $\{\mathbf{X}_i\}_{i=1}^n$, when n is large enough. Moreover, this ordering holds strictly on the event $\{\acute{\Delta}_n \subsetneq \hat{\Delta}_n\}$. Part (d) of Theorem 5.1 indicates the probability of $\{\acute{\Delta}_n \subsetneq \hat{\Delta}_n\}$ in large samples, with uniformity. This result is derived using Lemma F.1 which uses the classical Berry-Esseen bound for $s = 1$, and the generalized Berry-Esseen bound of Feller (1968) when $s > 1$. For $s > 1$, the proof does not require the existence of higher-order absolute moments of the moment functions, i.e., $\delta < 1$ in Condition (iv) of Assumption 3.1 is feasible.*

REMARK 5.3. *The set $\left\{t \in [\underline{t}, \bar{t}] : |E_P\left[g\left(\mathbf{X};t\right)\right]| \leq (1+\epsilon) r_n\right\}$ is an enlargement of $\Delta(P)$ that shrinks to it as the sample size tends to infinity because $\{(1+\epsilon) r_n\}_{n\geq1}$ is a null sequence for each $\epsilon > 0$. Therefore, Parts (b) and (e) of Theorem 5.1 imply that the two contact set estimators converge to $\Delta(P)$, uniformly over $\mathcal{P}_0^s(c_1, c_2, c_3)$. These results of this theorem drive the uniform asymptotic equivalence of the testing procedures under the null, which the next section presents.*

5.2. Asymptotic Size Properties

This section introduces the asymptotic size properties of the proposed modification of the LSW test and characterizes the sets of probabilities under null for which the proposed test has asymptotically exact size. LSW pay attention to the control of asymptotic rejection probabilities uniform in $P \in \mathcal{P}$. They introduce a regularity condition on the asymptotic Gaussian process ν in (3.3) with respect to r-enlargements of the contact set $B(r) = \left\{t \in [\underline{t}, \overline{t}] : |E_P\left[g\left(\mathbf{X}; t\right)\right]| \leq r\right\}$. For ease of exposition, this condition is given by Definition B.1 in Appendix B.3.1 along with a discussion.

The set of probabilities under which the proposed bootstrap test has asymptotically exact size is given by the following definition.

Definition 5.2. *(i) For each* $\epsilon > 0$, $s \in \mathbb{Z}_+$, $c_1 \in (0,1), c_2 \in \left(0, (\overline{t} - \underline{t})^{-2}\right)$ *and* $c_3 \in (0, +\infty)$, *let* $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$ *be the collection of probabilities in* $\mathcal{P}_0^s(c_1, c_2, c_3)$ *under which* ν *in (3.3) is* regular *on* B_n *in the sense of Definition B.1, for each* $n \geq 1$, *where*

$$B_n = \begin{cases} B\left((1-\epsilon)r_n\right), & \textit{if } \int_{B((1+\epsilon)r_n)} dt > 0, \textit{ and} \\ [\underline{t}, \overline{t}], & \textit{if } \int_{B((1+\epsilon)r_n)} dt = 0; \end{cases} \tag{5.2}$$

(ii) Given $\xi_n \to 0$, *let* $\mathcal{P}_{00}^s(c_1, c_2, c_3, \epsilon, \{\xi_n\})$ *be the collection of probabilities in* $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$ *under which for each* $n > 1/\epsilon$, ν *in (3.3) is* regular *on* $B\left(n^{-1/2}\xi_n\right)$ *in the sense of Definition B.1,*

$$\int_{B((1-\epsilon)r_n)} dt > 0 \quad \textit{and} \quad \int_{B((1+\epsilon)r_n) - B(n^{-1/2}\xi_n)} dt \leq \xi_n. \tag{5.3}$$

Denote by $\acute{\Upsilon}_B = B^{-1}\sum_{l=1}^{B} 1[\acute{T}_{n,l}^{\star} \geq \hat{T}_n]$ the bootstrap p-value that uses the bootstrap test statistics $\{\acute{T}_{n,l}^{\star}\}_{l=1}^{B}$ computed as in (3.5) of the LSW procedure, but with $\hat{\Delta}_n$ replaced by $\acute{\Delta}_n$.

Theorem 5.2. *For each* n, *let* $\mathcal{A}_n$ *denote the Borel sigma-algebra generated by the random sample* $\{\mathbf{X}_i\}_{i=1}^n$. *For each* $\epsilon > 0$, $s \in \mathbb{Z}_+$, $c_1 \in (0,1), c_2 \in \left(0, (\overline{t} - \underline{t})^{-2}\right)$ *and* $c_3 \in (0, +\infty)$, $\hat{\Upsilon}_B - \acute{\Upsilon}_B \xrightarrow{P} 0$ *conditional on* $\mathcal{A}_n$ *uniformly in* $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$.

Proof. See Appendix D.3.

Theorem 5.2 establishes the asymptotic equivalence of the bootstrap p-values $\hat{\Upsilon}_B$ and $\acute{\Upsilon}_B$, uniformly over $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$. For each $\epsilon > 0$, $s \in \mathbb{Z}_+$, $c_1 \in (0,1), c_2 \in \left(0, (\overline{t} - \underline{t})^{-2}\right)$ and $c_3 \in (0, +\infty)$, the LSW test has asymptotically exact size, in the sense of Definition 3.1, uniformly over a superset of $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$; therefore, it also has this property over $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$.

Consequently, the proposed test inherits the uniform asymptotic properties of the LSW test over the sets $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$. By applying Theorem 2 of LSW in the setup of our paper, these properties are: for each $\epsilon > 0$ $s \in \mathbb{Z}_+$, $c_1 \in (0,1), c_2 \in \left(0, (\overline{t} - \underline{t})^{-2}\right)$ and $c_3 \in (0, +\infty)$, (i) $\limsup_{n \to +\infty} \sup_{P \in \mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)} P[\hat{\Upsilon}_\infty \leq \alpha] \leq \alpha$, and (ii) $\limsup_{n\to+\infty} \sup_{P\in\mathcal{P}_{00}^{s}(c_1,c_2,c_3,\epsilon,\{\xi_n\})} |P[\hat{\Upsilon}_\infty \leq \alpha]-\alpha| = 0$, for each decreasing sequence $\xi_n \to 0$. Therefore, Theorem 5.2 implies that these properties hold with $\hat{\Upsilon}_\infty$ replaced by $\acute{\Upsilon}_\infty$.

Remark 5.4. *The subsets of $\mathcal{P}_0$ that establish the validity of the testing procedure this paper proposes (i.e., Definition 5.1) are subsets of their LSW counterparts. They are sets of probabilities that can be included in the empirical likelihood framework. The extent to which they are smaller than their LSW counterparts depends on the specific choices of the constants c_1, c_2 and c_3. In general, the closer these constants are to the lower bounds of their domains, the subsets of $\mathcal{P}_0$ demarcated by Definition 5.1 are closer to the LSW counterparts. For ease of exposition, Appendix B.3.2 elaborates on this point, and Appendices B.2.2 and B.3.3 discusses the choice of the constants in practice.*

6. Asymptotic Power Properties

6.1. Test Consistency

This section establishes consistency of the proposed testing procedure.

Theorem 6.1. *Given $s \in \mathbb{Z}_+$, suppose $P \in \mathcal{P} - \mathcal{P}_0$ is such that $\int_{\underline{t}}^{\overline{t}} \max\{E_P[g(\mathbf{X};t)],0\}^2\, dt > 0$. Then, $\lim_{n\to+\infty} P[\acute{\Upsilon}_B \leq \alpha] = 1$.*

Proof. See Appendix D.4.

Therefore, the proposed test is consistent against all alternatives. This property is also shared by the LSW test; see Theorem 3 of their paper. The next section focuses on the local asymptotic power property of the proposed test and compares it to the LSW test.

6.2. Power Against Local Alternatives

Following LSW, consider a sequence of probabilities $\{P_n\}_{n\geq 1} \subset \mathcal{P} - \mathcal{P}_0$ such that

$$E_{P_n}[g(\mathbf{X};t)] = H(t) + \delta(t)/\sqrt{n} \text{ and } \sigma_{P_n}^2(t) = E_{P_n}\left[g^2(\mathbf{X};t)\right] - \left(E_{P_n}[g(\mathbf{X};t)]\right)^2, \tag{6.1}$$

where the functions $H(t), \delta(t)$ and $\sigma_{P_n}^2(t)$ satisfy the following conditions.

Assumption 6.1. *(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) $\inf_{t\in[\underline{t},\overline{t}],n\in\mathbb{N}} \sigma_{P_n}^2(t) > 0$.*

Except for Part (iv), Assumption 6.1 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 i.i.d. random variables. subsubsectionBehavior of Contact Set Estimator Under Local Alternatives

We first present a result that characterizes the behavior of the contact set estimators under the sequences demarcated by Assumption 6.1.

THEOREM 6.2. *Under the local alternatives $\{P_n\}_{n\geq 1} \subset \mathcal{P} - \mathcal{P}_0$ satisfying the conditions in (6.1) and Assumption 6.1, the following statements hold.*

(a) $\lim_{n\to+\infty} P_n \left[E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right] \leq E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right] \; \forall t \in [\underline{t}, \overline{t}] \right] = 1.$

(b) $\lim_{n\to+\infty} P_n \left[C \subseteq \acute{\Delta}_n \subseteq \hat{\Delta}_n \right] = 1.$

(c) $\lim_{n\to+\infty} P_n \left[\hat{\Delta}_n \subsetneq \acute{\Delta}_n \right] = 0.$

(d) $\forall \epsilon > 0$, $\lim_{n\to+\infty} P_n \left[\hat{\Delta}_n \subseteq \left\{ t \in [\underline{t}, \overline{t}] : |E_{P_n}\left[g\left(\mathbf{X};t\right)\right]| \leq (1+\epsilon) r_n \right\} \right] = 1.$

PROOF. See Appendix D.5.

The results of Theorem 6.2 are similar to those of Theorem 5.1 but under local alternatives. Parts 2 and 3 of Theorem 6.2 imply that $\acute{T}_n^{\star} \leq \hat{T}_n^{\star}$ holds with probability tending to unity, conditional on the data. Thus, when some moment inequality is satisfied under the alternative and is sufficiently far from being an equality, then the proposed procedure detects this configuration more easily than the LSW procedure, and therefore, take it into account by delivering a bootstrap p-value that is suitable for the case where this moment inequality is omitted.

6.2.1. *Comparison of Local Asymptotic Power Functions*

Next, we compare the local asymptotic power functions of the LSW test and its modified counterpart, as well as the direct comparison of their local powers at equal sample sizes. This sort of comparison of the tests requires that we specify a set of probabilities satisfying Part (i) of Definition 5.2 as the model of the null hypothesis, where the limit of the local alternatives (in some topology on $\mathcal{P}$), $\lim_{n\to+\infty} P_n$, belongs to the boundary of this set. We have the following Corollary to Theorem 6.2.

COROLLARY 6.1. *Let $s \in \mathbb{Z}_+$ and suppose the conditions of Theorem 6.2 hold. Then, for every $\epsilon > 0$, $c_1 \in (0,1), c_2 \in \left(0, (\overline{t} - \underline{t})^{-2}\right)$ and $c_3 \in (0, +\infty)$, where $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$ is the null hypothesis with $\lim_{n\to+\infty} P_n \in \mathcal{P}_{00} \bigcap \mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$, the following statements hold.*

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

(b) $P_n \left[\acute{\Upsilon}_\infty \leq \alpha \right] \geq P_n \left[\hat{\Upsilon}_\infty \leq \alpha \right]$ *for large n, and with strict inequality if* $P_n[\acute{\Delta}_n \subsetneq \hat{\Delta}_n] > 0.$

PROOF. See Appendix D.6.

The first result of Corollary 6.1 establishes the equality of the limiting local power functions of the two testing procedures. The second result of Corollary 6.1 implies that the modified test has local power no less than that of the LSW test in large samples, and it can be strictly larger.

A remarkable point concerning Part 1 of Corollary 6.1 is that the testing procedures are asymptotically equivalent in test problems where the null models are demarcated by Part (i) of Definition 5.2, and the local alternatives satisfy Assumption 6.1. By contrast, the consequence of Part 2 of Corollary 6.1 is that the testing procedures will differ in large-samples.

Thus, a comparison of the tests that stops at the result of Part 1 of Corollary 6.1, is misleading. The result of Part 2 of Corollary 6.1 implies that a more informative comparison is in terms of the criteria: efficiency, deficiency, and normalized deficiency. The ensuing section presents results for these types of comparisons.

6.2.2. *Efficiency, Deficiency, and Normalized Deficiency*

Next, we turn our attention to assessing the relative asymptotic performance of these testing procedures. We compare them using their minimal required sample sizes for achieving a predesignated magnitude of power along the sequences of local alternatives demarcated by Assumption 6.1, subject to the constraint that the tests are asymptotically level α (i.e., the tests satisfy the condition in (3.1) in Definition 3.1). As with Corollary 6.1, this constraint necessitates that we specify a set of probabilities satisfying Part (i) of Definition 5.2 as the model of the null hypothesis and have $\lim_{n\to+\infty} P_n$ as an element in the boundary of this set. For every sample size n, define

$$k_n = \min\{N \in \mathbb{Z}_+ : P_m[\hat{\Upsilon}_\infty \leq \alpha] \geq P_n[\acute{\Upsilon}_\infty \leq \alpha]\, \forall m \geq N\}. \tag{6.2}$$

i.e. the minimal sample size the LSW test needs to attain local power which is at least as large as its modified counterpart along these local alternative sequences. The relative asymptotic performance of the two testing procedures is based on comparing k_n with n, when the former exists.

The next result describes the large sample behavior of k_n.

THEOREM 6.3. *Let $s \in \mathbb{Z}_+$ and suppose the conditions of Theorem 6.2 hold. Furthermore, let k_n be defined as in (6.2). Then, for every $\epsilon > 0$, $c_1 \in (0,1), c_2 \in \big(0, (\bar{t} - \underline{t})^{-2}\big)$ and $c_3 \in (0, +\infty)$, where $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$ is the model of the null hypothesis with $\lim\limits_{n\to+\infty} P_n \in \mathcal{P}_{00} \bigcap \mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$, the following statements hold.*

(a) If $P_n[\acute{\Delta}_n \subsetneq \hat{\Delta}_n] > 0$ for large n and $P_n[\hat{\Upsilon}_\infty \leq \alpha] \nearrow p_\infty \in (0,1)$ as $n \to +\infty$, then k_n exists for large enough n.

(b) If $P_n[\acute{\Delta}_n \subsetneq \hat{\Delta}_n] > 0$ for large n and $P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] \searrow p_\infty \in (0,1)$ as $n \to +\infty$, then k_n does not exist for large n.

PROOF. See Appendix D.7.

Theorem 6.3 distinguishes the cases when k_n exists and does not exist, in large samples. Part (a) indicates the local power functions $n \mapsto P_n[\acute{\Upsilon}_\infty \leq \alpha]$ and $n \mapsto P_n[\hat{\Upsilon}_\infty \leq \alpha]$, when n is large enough, are increasing towards a horizontal asymptote at $p_\infty \in (0,1)$. This large-sample behavior of the local power functions implies that k_n exists and satisfies $k_n \geq n+1$. In contrast, Part (b) indicates the power functions are decreasing towards a horizontal asymptote at $p_\infty \in (0,1)$, when n is large enough. In this contingency, k_n does not exist in large enough samples, which does not preclude the comparison of the local powers. The only deduction is that the LSW test can never outperform its modified version according to any measure of relative performance that uses k_n.

The next result characterizes the large sample behavior of efficiency, deficiency and normalized deficiency under the local alternatives demarcated by Assumption 6.1, and the conditions of Part 1 of Theorem 6.3.

COROLLARY 6.2. *Let $s \in \mathbb{Z}_+$, k_n be defined as in (6.2), and suppose the conditions for Part 1 of Theorem 6.3 hold. Then, for every $\epsilon > 0$, $c_1 \in (0,1), c_2 \in \left(0, (\bar{t} - \underline{t})^{-2}\right)$ and $c_3 \in (0,+\infty)$, where $\mathcal{P}_0^s(c_1,c_2,c_3,\epsilon)$ is the model of the null hypothesis with $\lim_{n\to+\infty} P_n \in \mathcal{P}_{00} \bigcap \mathcal{P}_0^s(c_1,c_2,c_3,\epsilon)$, the following statements hold.*

(a) $k_n/n > 1$ and $d_n > 0$ for large n, and hence, $\liminf_{n\to+\infty}(k_n/n) \geq 1$ and $\liminf_{n\to+\infty} d_n \geq 0$.

(b) Suppose that $\lim_{n\to+\infty} d_n = +\infty$, and let the sequence $\{q_n\}_{n\geq 1}$ satisfy

$$\lim_{n\to+\infty} \frac{n\left(P_{k_n}[\hat{\Upsilon}_\infty \leq \alpha] - P_n[\hat{\Upsilon}_\infty \leq \alpha]\right)}{q_n\left(P_n[\hat{\Upsilon}_\infty \leq \alpha] - b_n\right)} = \ell \in \mathbb{R}_{++}, \tag{6.3}$$

where, for each n, b_n is the intercept of the linear function that passes through the points $\left(n, P_n[\hat{\Upsilon}_\infty \leq \alpha]\right)$ and $\left(k_n, P_{k_n}[\hat{\Upsilon}_\infty \leq \alpha]\right)$. Then, $\lim_{n\to+\infty}(d_n/q_n) = \ell$.

PROOF. See Appendix D.8.

In general, it is not possible to find k_n/n for fixed values of n because the exact values of $P_n[\hat{\Upsilon}_\infty \leq \alpha]$ and $P_n[\acute{\Upsilon}_\infty \leq \alpha]$ are not known. However, the result of Corollary 6.2 provides strong evidence for the superior performance of the proposed testing procedure over the LSW test. It indicates that, in large-samples, the proposed test is superior to the LSW test since more observations than n are required by the LSW test to achieve the same local power as the proposed test at sample size n. In the circumstance that the deficiency diverges to infinity, Part (b) of Corollary 6.2 establishes that, in large-samples, $d_n \approx \ell q_n$. Thus, the price for using the LSW test is increasing with the sample size at the rate ℓq_n.

7. Monte Carlo Experiments

This section reports the results of Monte Carlo experiments that compares the performance of the LSW and Andrews and Shi (2017) (AS, hereafter) tests with their modified counterparts. Appendix C.1 presents the AS procedure in the context of unconditional restricted stochastic dominance testing and explains how to modify it using our approach. The experimental setup is the same as the one in Section 5 of LSW who focus on testing for first-order stochastic dominance. 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%, $r_n(t) = \hat{\sigma}_t\sqrt{\log(n)/n}$, where $\hat{\sigma}_t^2 = E_{\hat{P}_n}[g(\mathbf{X};t)]^2 - (E_{\hat{P}_n}[g(\mathbf{X};t)])^2$ and $t \in [\underline{t}, \overline{t}]$. This choice for r_n is the Bayesian Information Criterion (BIC) choice. An alternative choice is $r_n = an^{-1/2}\log\log n$, which LSW use, 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 this localization parameter. Andrews and Soares (2010) suggest the BIC choice for use in practice, and we follow their lead. We set $\underline{t} = 0.05$ and $\overline{t} = 0.95$. Finally, the number of Monte Carlo replications was set to 10000 in each simulation experiment, and the number of bootstrap replications was 499.

7.1. Simulation Under H_0

We compare the type I error rate properties of the LSW and AS tests, and their modified versions. 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 (1.1) 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 size was $n = 500$. 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.

Figure 3 the empirical rejection frequencies of the tests along with their pointwise 95% confidence intervals. For each value of c_0, the discrepancy between the performances of the LSW and AS tests and their modified versions are not large when 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 the modified tests are statistically closer to the 5% nominal level than the ones based on the LSW and AS tests. These results suggest the bias of the LSW and AS tests is larger than their modified versions.

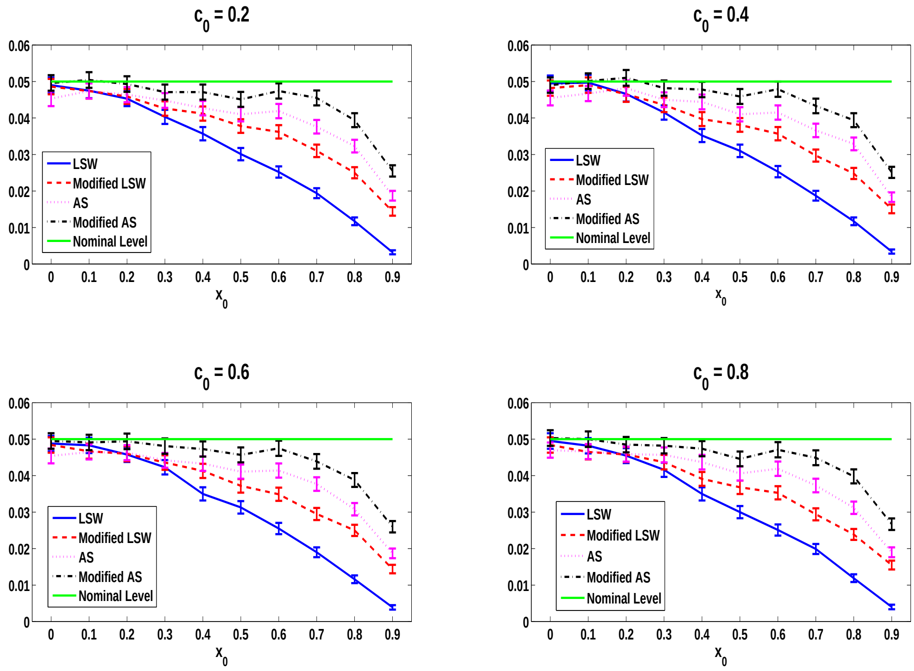

Fig. 3. The empirical rejection probabilities under the null.

7.2. *Simulation Under H_1*

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 $X^B = (U - a_0 b_1)1[a_0 b_1 \leq U \leq x_0] + (U + a_0 b_2)1[x_0 < U \leq 1 - a_0 b_2]$ 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 $a_0 \in \{0, 0.05, 0.1, 0.15, 0.2, \ldots, 0.45\}$. 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 (1.1). We considered the following sample sizes $n = 256, 512, 1024$, 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.

Figure 4 reports the simulation results, 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 tests, which is expected since they are all consistent. For moderate values of a_0, the modified versions of the LSW and AS tests have statistically higher power than their non-modified versions, and the power differences can be large. For example, when $n = 256$ and $a_0 = 0.15$, the difference in powers for the LSW and AS type tests are approximately 10% and 21%, respectively. And when $n = 1024$ and $a_0 = 0.1$, these differences are approximately 25% and 12%, respectively.

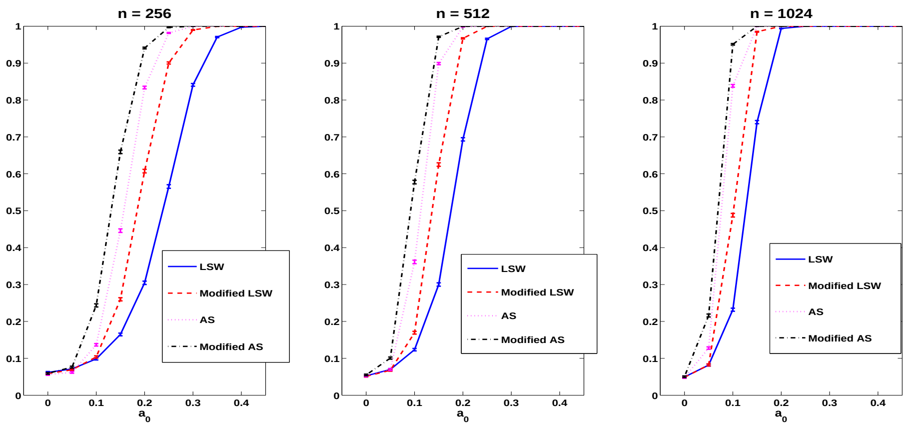

Fig. 4. The empirical rejection probabilities under the alternative: directions of non-dominance.

These findings suggest that the modified tests can better detect DGPs in H_1 that are "close" to H_0, when the sample size is moderately large. Overall, the simulation results show that our method yields tests that perform better than their non-modified versions.

8. Conclusion

While the paper focuses almost exclusively on the LSW test, the method we propose extends to restricted stochastic dominance testing within the framework of Andrews and Shi (2017). The simulation evidence we report suggests that similar results will be obtained for their test and its modified version. Appendix C outlines these results.

The method this paper presents easily extends to the partial identification setup in which testing is on a finite-dimensional parameter defined by a finite number of unconditional moment inequality conditions; for example, as in the general framework of Andrews and Soares (2010). Within their setup, the surgical modification simplifies to solving a finite program: constrained empirical likelihood subject to a finite number of inequality constraints. However, a detailed theoretical analysis of the merits this approach may offer goes beyond the intended scope of this paper.

9. Acknowledgments

The paper has greatly benefited from feedback received in seminar presentations at McGill University, Monash University, UCSD, UCLA, UC Davis, and the Harvard-MIT Econometrics Workshop. Rami V. Tabri thanks Garry Barrett, Russell Toth, Peter Exterkate, Yi Sun, Mervyn Silvapulle, Andres Santos and Yoon-Jae Whang for helpful comments and discussions. The authors acknowledge the Sydney Informatics Hub and the University of Sydney's high performance computing cluster Artemis for providing the high performance computing resources that have contributed to the research results reported within this paper.

References

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association 97*(457), 284–292.

Akahira, M. (1999). The concept of normalized deficiency and its applications. *Statistics & Risk Modeling 17*, 403–412.

Albers, W. (1975). Efficiency and deficiency considerations in the symmetry problem. *Statistica Neerlandica 29*(3), 81–92.

Andrews, D. W. and P. Guggenberger (2009, 6). Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. *Econometric Theory 25*, 669–709.

Andrews, D. W. and P. Guggenberger (2010, 4). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory 26*, 426–468.

Andrews, D. W. and X. Shi (2017). Inference based on many conditional moment inequalities. *Journal of Econometrics 196*(2), 275 – 287.

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

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

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

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

Berger, R. L. (1988). A nonparametric, intersection-union test for stochastic order. In *Statistical decision theory and related topics, IV, Vol. 2 (West Lafayette, Ind., 1986)*, pp. 253–264. Springer, New York.

Billingsley, P. (1995). *Probability and Measure* (Third ed.). New York: John Wiley & Sons.

Canay, I. A. (2010). EL Inference for Partially Identified Models:Large Deviations Optimality and Bootstrap Validity. *Journal of Econometrics 156*(2), 408–425.

Carroll, R. J., A. Delaigle, and P. Hall (2011). Testing and estimating shape-constrained nonparametric density and regression in the presence of measurement error. *Journal of the American Statistical Association 106*(493), 191–202.

Chernozhukov, V., S. Lee, and A. Rosen (2013). Intersection Bounds: Estimation and Inference. *Econometrica 82*.

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

Devroye, L. (1982, 08). A log log law for maximal uniform spacings. *Ann. Probab. 10*(3), 863–868.

Donald, S. G. and Y.-C. Hsu (2016). Improving the power of tests of stochastic dominance. *Econometric Reviews 35*(4), 553–585.

Du, P. J., C. Parmeter, and J. Racine (2013). Nonparametric Kernel Regression with Multiple Predictors and Multiple Shape Constraints. *Statistica Sinica 23*, 1347–1371.

Feller, W. (1968, Sep). On the berry-esseen theorem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 10*(3), 261–268.

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

Hall, P. and L.-S. Huang (2001, 06). Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics 29*(3), 624–647.

Hall, P. and L.-S. Huang (2002). Unimodal density estimation using kernel methods. *Statistica Sinica 12*(4), 965–990.

Hall, P. and B. Presnell (1999). Intentionally biased bootstrap methods. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 61*(1), pp. 143–158.

Hodges, J. L. and E. L. Lehmann (1970, 06). Deficiency. *Ann. Math. Statist. 41*(3), 783–801.

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

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

lvarez Esteban, P. C., E. del Barrio, J. A. Cuesta-Albertos, and C. Matrn (2017, 08). Models for the assessment of treatment improvement: The ideal and the feasible. *Statist. Sci. 32*(3), 469–485.

Mordukhovich, B. and T. Nghia (2013). Constraint qualifications and optimality conditions for nonconvex semi-infinite and infinite programs. *Mathematical Programming 139*(1-2), 271–300.

Owen, A. (2001). *Empirical Likelihood*, Volume 92 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC.

Pitman, E. (1948). *Lecture Notes on Nonparametric Statistical Inference: Lectures Given for the University of North Carolina*. Mathematisch Centrum.

Sen, P. K. and M. J. Silvapulle (2001). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. Wiley.

Shapiro, A. (2009). Semi-Infinite Programming, Duality, Discretization, and Optimality Conditions. *Optimization 8*(2), 133–161.

Sriboonchitta, S., W.-K. Wong, S. Dhompongsa, and H. Nguyen (2010). *Stochastic Dominance and Applications to Finance, Risk, and Economics* (First ed.). Chapman and Hall/CRC, Taylor and Francis Group.

Still, G. (2001). Discretization in semi-infinite programming: the rate of convergence. *Mathematical programming 91*(1), 53–69.

van der Vaart, A. W. and J. Wellner (1996). *Weak Convergence and Empirical Processes* (First ed.). Springer Series in Statistics. Springer.

A. Outline

This Appendix provides supplementary material to this paper. It is organized as follows.

- Appendix B covers further discussions of points raised in the paper.
- Appendix C presents (i) the framework of Andrews and Shi (2017), but in the context of testing for restricted stochastic dominance; (ii) the modification of Andrews and Shi's test based on this paper's proposal; and (iii) the comparison of Andrews and Shi's test and its modification under the null and local alternatives.
- Appendix D presents the proofs the results in the paper: Theorems 4.1, 5.1, 5.2, 6.1, 6.2, and Corollaries 6.1, 6.2.
- Appendix E presents technical lemmas used in the proof of Theorem 4.1.
- Appendix F presents technical lemmas used in the proofs of Theorems 5.1 and 5.2.
- Appendix G presents technical lemmas using in the proofs of Theorems 6.1 and 6.2.

B. Further Discussion

B.1. Section 2

B.1.1. Behavior of The Tests Under H_0

Setting $\alpha = 0.05$ and $r_n = \sqrt{\log(n)/n}$, the following numerical results establish the test's exact sizes are approximately equal to 0.05 for sample sizes $n = 100, 101, 102, \ldots, 10000$. The left panel in Figure 5 reports exact sizes of the tests, which are defined as $\sup_{P\in\mathcal{M}_0} E_P\hat{\tau}_n$ and $\sup_{P\in\mathcal{M}_0} E_P\tilde{\tau}_n$. They were computed using numerical integration and optimization packages in Matlab.

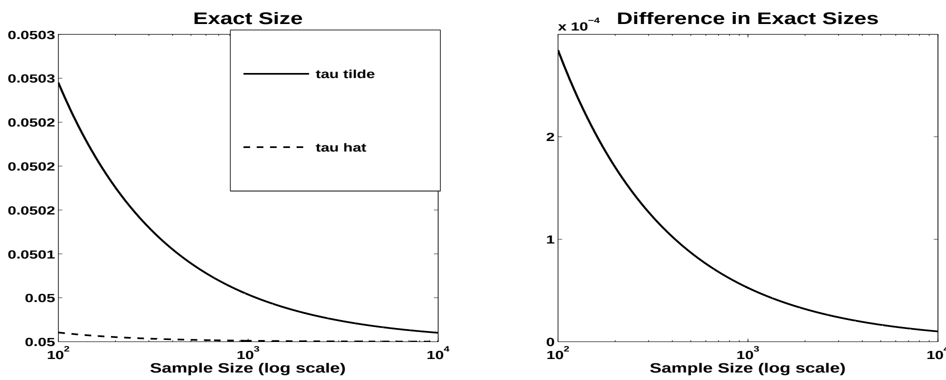

Fig. 5. The first panel reports the exact sizes of the tests $\hat{\tau}_n$ and $\tilde{\tau}_n$. The second panel reports the difference $\sup_{P\in\mathcal{M}_0} E_P\hat{\tau}_n - \sup_{P\in\mathcal{M}_0} E_P\tilde{\tau}_n$.

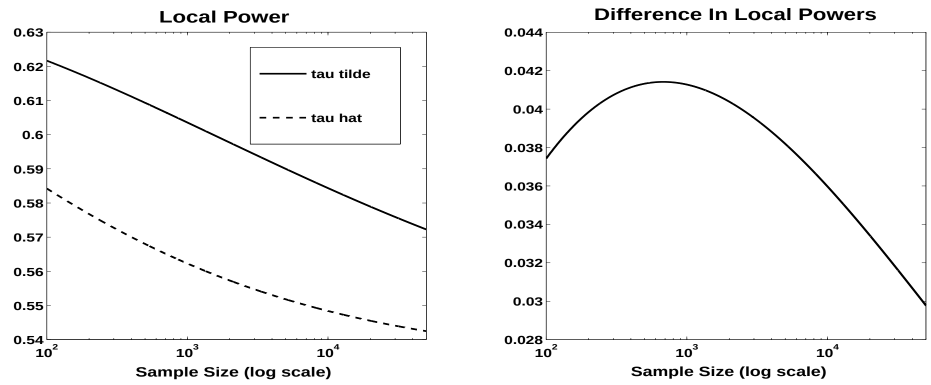

Fig. 6. The left panel plots the local powers $E_{P_n}\hat{\tau}_n$ and $E_{P_n}\tilde{\tau}_n$. The right panel plots the difference $E_{P_n}\tilde{\tau}_n - E_{P_n}\hat{\tau}_n$.

The right panel in Figure 5 reports the difference $\sup_{P\in\mathcal{M}_0} E_P\hat{\tau}_n - \sup_{P\in\mathcal{M}_0} E_P\tilde{\tau}_n$. Overall, the exact sizes of the tests are approximately equal to the nominal level and they are within 0.0003 of each other. Hence, it is feasible to compare the tests' rejection probabilities along local alternatives for sample sizes $n \geq 100$ without adjusting them so that they have the same exact size.

B.1.2. Characterization of $\tilde{\mu}$

Using the Karush-Kuhn-Tucker conditions for the constrained optimization problem in (2.3), it is a straightforward task to deduce that

$$\tilde{\mu} = [\tilde{\mu}_1, \tilde{\mu}_2] = \begin{cases} [\hat{\mu}_1, \hat{\mu}_2] & \text{if } \hat{\mu}_1, \hat{\mu}_2 \leq 0 \\ [\hat{\mu}_1 - \rho_0\hat{\mu}_2, 0] & \text{if } \hat{\mu}_1 \leq 0, \hat{\mu}_2 > 0 \\ [0, \hat{\mu}_2 - \rho_0\hat{\mu}_1] & \text{if } \hat{\mu}_1 > 0, \hat{\mu}_2 \leq 0 \\ [0, 0] & \text{if } \hat{\mu}_1, \hat{\mu}_2 > 0. \end{cases} \tag{B.1}$$

We omit the statement of these first-order conditions for brevity.

B.1.3. Relative Performance of Tests: A Different Type of $\sqrt{n}$ Local-Alternatives

Consider local alternatives $\{P_n\}_n \subset \mathcal{M} - \mathcal{M}_0$ in which P_n satisfies the hypothesis

$$H_n : \mu_{1n} = -2/\sqrt{n} \quad \text{and} \quad \mu_{2n} = 2/\sqrt{n}, \quad \forall n = 100, 101, \ldots, 50000. \tag{B.2}$$

Figure 6 reports the numerical results. They show $E_{P_n}\tilde{\tau}_n > E_{P_n}\hat{\tau}_n$ for all the values of n under consideration; therefore, k_n does not exists for these sample sizes. This result does not preclude the comparison of the the tests. The only deduction is that $\hat{\tau}_n$ can never outperform $\tilde{\tau}_n$ according to any measure of relative performance that uses k_n; for example, the criteria described in Table 1.

B.2. Section 4

B.2.1. Definition 4.1

Definition 4.1 specifies two subsets of $\mathcal{P}$, which are $\mathcal{P}_1(c_1)$ and $\mathcal{P}_2(c_2)$. Focusing firstly on $\mathcal{P}_1(c_1)$, we describe the event $\left\{\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X};t)<0\right\}$. It is given by

$$\left\{\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X};t)<0\right\} = \begin{cases} \left\{X^A<\underline{t}, X^B>\overline{t}\right\}, & \text{if } s=1, \\ \left\{X^A<\underline{t}, X^B>\overline{t}\right\} \cup \left\{X^A<\underline{t}, X^A<X^B\leq\overline{t}\right\}, & \text{if } s\geq 2. \end{cases} \tag{B.3}$$

Hence, this event is a tail event, and Part (i) of Assumption 3.1 implies

$$P\left[\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X};t)<0\right]>0\ \forall P\in\mathcal{P}.$$

Consequently, the set $\mathcal{P}_1(c_1)$ excludes distributions that place probability less than c_1 on the respective tail events in (B.3) when $s=1$, and when $s>1$.

The sets of the form $\mathcal{P}_2(c_2)$ also restrict $\mathcal{P}$. They are subsets over which the convergence of the grid $\mathcal{T}_n$ to the interval $[\underline{t},\overline{t}]$, is uniform. These subsets exclude distributions whose joint densities are arbitrary close to zero on the box $[\underline{t},\overline{t}]\times[\underline{t},\overline{t}]$. This condition is essential for controlling the size of the derivative of the quantile function, with uniformity. Specifically, we express the elements of the grid in terms of $U(0,1)$ random variables using the Mean Value Theorem, and the reciprocals of the marginal densities of P arise in it. This representation of the grid allows us to use the theory of maximal uniform spacings (e.g. Devroye, 1982) to establish the uniform convergence of the grid to the interval $[\underline{t},\overline{t}]$. For ease of exposition, these intermediate technical results are relegated to Appendix E.

B.2.2. Verifying The Conditions of $\mathcal{P}_1(c_1)$ and $\mathcal{P}_2(c_2)$.

The conditions of Definition 4.1 are, in fact, verifiable in practice by means of statistical testing. In the case of $\mathcal{P}_1(c_1)$, given $c_1\in(0,1)$ the testing problem is

$$H_0: P_0\left[\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X};t)<0\right]\geq c_1 \quad \text{versus} \quad H_1: P_0\left[\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X};t)<0\right]<c_1. \tag{B.4}$$

A t-test based on the statistic $\frac{1}{n}\sum_{i=1}^n 1\left[\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X}_i;t)<0\right]$ and a least favorable critical value, yields a valid testing procedure, in which the critical value can be derived using, for example, the canonical bootstrap in a standard way.

In the case of $\mathcal{P}_2(c_2)$, given $c_2\in(0,(\overline{t}-\underline{t})^{-2})$, the testing problem is

$$H_0: \inf_{(x^A,x^B)\in[\underline{t},\overline{t}]\times[\underline{t},\overline{t}]} f(x^A,x^B)\geq c_2 \quad \text{versus} \quad H_1: \inf_{(x^A,x^B)\in[\underline{t},\overline{t}]\times[\underline{t},\overline{t}]} f(x^A,x^B)<c_2,$$

where $f(x^A,x^B)$ is the joint density whose probability measure is P_0. This testing problem is a test on an intersection bound for the joint density $f(x^A,x^B)$. Hence, the estimation and inference procedures Chernozhukov et al. (2013) introduce apply to this testing problem.

B.2.3. Computational Aspects

The data-driven discretization scheme, introduced in Section 4, is a sequence of finite programming optimization problems. The optimization problems in the scheme can be easily implemented using standard numerical computing packages and built-in optimization routines (e.g., fmincon in Matlab), even when n is large. That's because the constraints in (4.2) are linear in the choice variables $p_1, \ldots, p_n$. Furthermore, these problems are strictly concave; therefore, it is sufficient to compute only a local optimum when searching for the global optimum (which is unique).

Property 1 implies the solution $\acute{\mathbf{p}}$, given by (4.2), satisfies the following property

$$E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right] \leq E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right] \quad \forall t \in [\underline{t}, \overline{t}], \tag{B.5}$$

when it exists. Therefore, the inequalities (B.5) can be used to increase the numerical accuracy and speed of computation in the optimization problem (4.2) by replacing the constraints $\sum_{i=1}^{n} p_i g\left(\mathbf{X}_i;t\right) \leq 0 \; \forall t \in \mathcal{T}_n$, with the following:

$$\sum_{i=1}^{n} p_i g\left(\mathbf{X}_i;t\right) \leq 0 \quad \forall t \in \mathcal{T}_n \cap \left\{t \in [\underline{t}, \overline{t}] : E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right] \geq 0\right\} \quad \text{and} \tag{B.6}$$

$$\sum_{i=1}^{n} p_i g\left(\mathbf{X}_i;t\right) \leq E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right] \quad \forall t \in \mathcal{T}_n \cap \left\{t \in [\underline{t}, \overline{t}] : E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right] < 0\right\}, \tag{B.7}$$

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

B.3. Section 5

B.3.1. Regularity On The Asymptotic Gaussian Process ν

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 in (3.3) which is given by Definition 2 of their paper. In the context of the present work, this condition is the following.

Definition B.1. *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*

$$P\left[\int_A \max\left\{\nu(t), 0\right\}^2 \, dt < \overline{\epsilon}\right] < 1 - \alpha \tag{B.8}$$

and for any $c > 0$,

$$\limsup_{\eta \downarrow 0} \sup_{P \in \mathcal{P}_0} P\left[\left|\int_A \max\left\{\nu(t), 0\right\}^2 \, dt - c\right| \leq \eta\right] = 0. \tag{B.9}$$

Condition (B.8) is a weak requirement. It restricts the null parameter space by excluding probabilities for which the bootstrap p-value has a large mass point at zero that exceeds α. To understand this condition, note that the bootstrap empirical process in the definition of the bootstrap statistic $\hat{T}_n^{\star}$ converges to the

Gaussian process ν at the $\sqrt{n}$ rate; therefore, the behavior of the sequence $\{r_n\}_n$ implies the asymptotic behavior of $\hat{T}_n^\star$ (conditional upon the data) is equivalent to $\int_{A_n} \max\{\nu(t), 0\}^2 \, dt$, where $A_n = \{t \in [\underline{t}, \overline{t}] : |E_P[g(\mathbf{X}; t)]| \leq r_n\}$. Condition (B.9) requires all points $c > 0$ to be points at which the distribution of

$$\int_A \max\{\nu(t), 0\}^2 \, dt$$

is continuous, uniformly in $\mathcal{P}_0$.

B.3.2. Proposed and LSW Null Parameter Spaces

The sets of probabilities on which the proposed test has asymptotically correct size are defined in Part (i) of Definition 5.2. 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 5.2. These sets are subsets of their LSW counterparts that can be included in the empirical likelihood framework described in Sections 4, 5.1 and 5.2. The sets of probabilities included in the LSW framework that yield asymptotically correct size are: for each $\epsilon > 0$

$$\mathcal{P}_0(\epsilon) = \{P \in \mathcal{P}_0 : \nu \text{ in } (3.3) \text{ is } \textit{regular} \text{ on } B_n \, \forall n \geq 1\},$$

where B_n is defined in (5.2), in Part (i) of Definition 5.2. And the sets of probabilities in their framework that yield asymptotic similarity are: for each $\epsilon > 0$ and decreasing sequence $\xi_n \to 0$,

$$\begin{aligned} \mathcal{P}_0(\epsilon, \{\xi_n\}) = \{&P \in \mathcal{P}_0(\epsilon) : \nu \text{ in } (3.3) \text{ is } \textit{regular} \text{ on } B(n^{-1/2}\xi_n), \\ &\text{and (5.3) holds } \forall n \geq 1/\epsilon\}. \end{aligned}$$

A natural question to raise at this point is by how much the set of probabilities in Parts (i) and (ii) of Definition 5.2 are more restrictive than their LSW counterparts. That is, given $\epsilon > 0$ and a decreasing sequence $\xi_n \to 0$, how large are the set differences

$$\mathcal{P}_0(\epsilon) - \mathcal{P}_0^s(c_1, c_2, c_3, \epsilon) \quad \text{and} \quad \mathcal{P}_0(\epsilon, \{\xi_n\}) - \mathcal{P}_0^s(c_1, c_2, c_3, \epsilon, \{\xi_n\}), \tag{B.10}$$

where $c_1 \in (0, 1)$, $c_2 \in (0, (\overline{t} - \underline{t})^{-2})$ and $c_3 \in (0, +\infty)$. The answer to this question depends on the values of the constants c_j for $j = 1, 2, 3$. It is easily seen that these set differences (B.10) tend to the empty set as $c_j \to 0$ for $j = 1, 2, 3$. In consequence, the sets of probabilities that are relevant to our framework can be made arbitrarily close to their LSW counterparts by setting c_1, c_2 and c_3 arbitrarily close to zero. Otherwise, for values of the constants close to the upper bounds of their domains, the set differences (B.10) can be non-negligible. In practice, the appropriate values for these constants can be determined via statistical tests; Appendices B.2.2 and B.3.3 elaborate on this point.

B.3.3. Verifying The Regularity Condition in Definitions 5.1

The first step in setting up the model of the null hypothesis, defined in Part (i) of Definition 5.2, is to set values for $c_1 \in (0,1)$, $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$ and $c_3 \in (0, +\infty)$, so as to designate sets of probabilities $\mathcal{P}_j(c_j), j = 1,2,3$, respectively. Appendix B.2.2 establishes that the conditions that define the sets of probabilities $\mathcal{P}_j(c_j), j = 1,2$ are, in fact, verifiable in practice by means of statistical testing. In this section, we discuss how to test for the condition that defines the set $\mathcal{P}_3(c_3)$.

Given $c_3 \in (0, +\infty)$, consider the testing problem

$$H_0 : \inf_{t \in \Delta(P_0)} E_{P_0}\left[g^2(\mathbf{X}; t)\right] \geq c_3 \quad \text{versus} \quad H_1 : \inf_{t \in \Delta(P_0)} E_{P_0}\left[g^2(\mathbf{X}; t)\right] < c_3. \quad \text{(B.11)}$$

It is also testing problem on an intersection bound, but for the variances of the moment functions indexed by the contact set $\Delta(P_0)$. It should be noted that the methods Chernozhukov et al. (2013) introduce must be appropriately adjusted for this intersection bound testing problem because the set over which the infimum is being taken depends on P_0, which is a case that their work does not cover. Such a modification of their testing procedures is, however, beyond the intended scope of this paper.

C. The Framework of Andrews and Shi (2017)

This section specializes the framework of Andrews and Shi (2017) to the case of restricted stochastic dominance described in Section 3. For each $s \in \mathbb{Z}_+$, their statistical model coincides with $\mathcal{P}$ when the envelope and scale functions in their model are specified as

$$M(\mathbf{x}) = s \max\left\{|\underline{t}|, |\bar{t}|\right\}^{s-1}\left(|x^B|^{s-1} + |x^A|^{s-1}\right) \quad \text{and} \quad \text{(C.1)}$$

$$\sigma_P(0) = 1 \quad \text{for} \quad P \in \mathcal{P}, \quad \text{(C.2)}$$

respectively.

A distinguishing feature of their testing procedure is that their test statistic and it bootstrap version employ Studentization. Specifically, they are functions of Studentized empirical processes. The next section details their testing procedure.

C.1. Bootstrap Testing Procedure

We outline the steps of the bootstrap procedure Andrews and Shi (2017) (AS, hereafter) propose. The Monte Carlo experiments implement their test in Section 7 for testing restricted stochastic dominance. They propose a Kolmogorov-Smirnov and Cramér-von-Mises test statistics for inference on possibly infinite number of conditional moment inequality conditions. Recall that the setting of this paper considers a continuum of unconditional moment inequality conditions, which the AS procedure covers as a special case. In this setting, the AS test statistics are identical, and given by

$$\hat{T}_n = \sup_{t\in[\underline{t},\overline{t}]} \left(\max\left\{ \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t\right)\right)/\hat{\sigma}(t), 0\right\}\right)^2, \quad \text{where} \tag{C.3}$$

$$\hat{\sigma}^2(t) = \frac{1}{n}\sum_{i=1}^{n} g^2\left(\mathbf{X}_i;t\right) - \left[\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t\right)\right]^2. \tag{C.4}$$

Next we describe the steps for computing the AS bootstrap GMS critical value in the setting of this paper. The critical value is obtained through the following steps.

(a) Compute $\overline{\varphi}_n(t)$ for $t \in [\underline{t},\overline{t}]$, where $\overline{\varphi}_n(t)$ is defined as follows. Let

$$\xi_n(t) = \kappa_n^{-1}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t\right)\right)/\hat{\sigma}(t), \tag{C.5}$$

where $\kappa_n = (0.3\log(n))^{1/2}$. Define

$$\overline{\varphi}_n(t) = \hat{\sigma}(t)B_n 1\left[\xi_n(t) < -1\right] \quad \text{and} \quad B_n = (0.4\log(n)/\log\log(n))^{1/2}. \tag{C.6}$$

(b) Generate B bootstrap samples $\left\{\mathbf{X}^\star_{i,l}\right\}_{i=1}^n$ for $l = 1, \ldots B$ using the ECDF on the data.

(c) For each bootstrap sample, compute $\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}^\star_{i,l};t\right)$, and $\hat{\sigma}^2_l(t)$ just as $\hat{\sigma}^2(t)$ is computed but with the bootstrap sample in place of the original sample.

(d) For each bootstrap sample, compute the bootstrap test statistic $\hat{T}^\star_{n,l}$ as $\hat{T}_n$ is computed in (C.3) but with $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}_i;t\right)\right)$ replaced by $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}^\star_{i,l};t\right) - \frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}_i;t\right) - \overline{\varphi}_n(t)\right)$ and with $\hat{\sigma}^2(t)$ replaced by $\hat{\sigma}^2_l(t)$.

(e) Take the bootstrap GMS critical value $c_{n,1-\alpha}$ to be the $1-\alpha+\eta$ sample quantile of the bootstrap test statistics $\left\{\hat{T}^\star_{n,l}, l = 1, \ldots, B\right\}$ plus η, where $\eta = 10^{-6}$.

For a given nominal level $\alpha \in (0, 1/2)$, the AS test rejects H_0 if $\hat{T}_n > c_{n,1-\alpha}$. Denote their test by $\hat{\tau}^{AS} = 1[\hat{T}_n > c_{n,1-\alpha}]$.

C.2. Model of The Null Hypothesis

This section describes the subsets of $\mathcal{P}_0$ on which $\hat{\tau}^{AS}$ satisfies (3.1) in Definition 3.1. Define the scaled covariance kernel function

$$h_{2P}(t_1,t_2) = \frac{\mathrm{Cov}_P\left(g\left(\mathbf{X};t_1\right), g\left(\mathbf{X};t_2\right)\right)}{\sqrt{\mathrm{VAR}_P\left(g\left(\mathbf{X};t_1\right)\right)\mathrm{VAR}_P\left(g\left(\mathbf{X};t_2\right)\right)}} \quad t_1,t_2 \in [\underline{t},\overline{t}], \tag{C.7}$$

and consider the set of covariance kernels that correspond to $\mathcal{P}_0$ given by

$$\mathcal{C} = \{h_{2P}(\cdot,\cdot) : P \in \mathcal{P}_0\}. \tag{C.8}$$

On the set $\mathcal{C}$ define the uniform metric

$$d\left(h_2^{(1)}, h_2^{(2)}\right) = \sup_{t_1,t_2\in[\underline{t},\overline{t}]} \left| h_2^{(1)}(t_1,t_2) -, h_2^{(2)}(t_1,t_2)\right|. \tag{C.9}$$

According to Theorem 5.1 of AS, the subsets of $\mathcal{P}_0$ on which $\hat{\tau}^{AS}$ satisfies (3.1) are of the form

$$\{P \in \mathcal{P}_0 : h_{2P} \in \mathcal{C}_{cpt}\}, \tag{C.10}$$

where $\mathcal{C}_{cpt}$ is a compact subset of $\mathcal{C}$ with respect to the uniform metric $d(\cdot,\cdot)$. That is, given $s \in \mathbb{Z}_+$,

$$\limsup_{n\to+\infty} \sup_{P\in\{P\in\mathcal{P}_0 : h_{2P}\in\mathcal{C}_{cpt}\},} E_P \hat{\tau}^{AS} \leq \alpha, \tag{C.11}$$

for every compact subset $\mathcal{C}_{cpt}$ of $\mathcal{C}$.

C.3. Proposed Modification of Andrews and Shi's Bootstrap Procedure

The proposed modification this paper suggests alters the AS testing procedure by replacing the sample-analog estimator of the moments in (C.5) with the constrained empirical likelihood estimator described in Section 4. Let $\acute{\varphi}_n(t)$ be constructed in the same way as $\overline{\varphi}_n(t)$, but with $\sum_{i=1}^n \acute{p}_i g(\mathbf{X}_i; t)$ in place of $\frac{1}{n}\sum_{i=1}^n g(\mathbf{X}_i; t)$. Then, the contact set estimator in the AS procedure and its modified version are given by

$$\hat{\Delta}_n^{\text{AS}} = \left\{t \in [\underline{t},\overline{t}] : \overline{\varphi}_n(t) = 0\right\} \quad \text{and} \quad \acute{\Delta}_n^{\text{AS}} = \left\{t \in [\underline{t},\overline{t}] : \acute{\varphi}_n(t) = 0\right\}, \tag{C.12}$$

respectively.

Property 1 of the moment functions implies that

$$[\underline{t},\overline{t}] - \hat{\Delta}_n^{\text{AS}} \subseteq [\underline{t},\overline{t}] - \acute{\Delta}_n^{\text{AS}} \tag{C.13}$$

holds numerically when $\acute{\mathbf{p}}$ exists and is characterisable by Lagrange multipliers, which is equivalent to $\acute{\Delta}_n^{\text{AS}} \subseteq \hat{\Delta}_n^{\text{AS}}$. These set inclusions imply

$$\acute{\varphi}_n(t) = \overline{\varphi}_n(t) \quad \text{for } t \in \acute{\Delta}_n^{\text{AS}}, \tag{C.14}$$

$$\acute{\varphi}_n(t) > \overline{\varphi}_n(t) \quad \text{for } t \in \hat{\Delta}_n^{\text{AS}} - \acute{\Delta}_n^{\text{AS}}, \text{ and} \tag{C.15}$$

$$\acute{\varphi}_n(t) = \overline{\varphi}_n(t) \quad \text{for } t \in [\underline{t},\overline{t}] - \hat{\Delta}_n^{\text{AS}}, \tag{C.16}$$

when $\acute{\mathbf{p}}$ exists and is characterisable in terms of Lagrange multipliers. Thus, conditional upon the sample, the inequalities (C.14) - (C.16) yield $\hat{T}^\star_{n,s} \geq \acute{T}^\star_{n,s}$, where $\acute{T}^\star_{n,s}$ is computed in exactly the same way as $\hat{T}^\star_{n,s}$ but with $\acute{\varphi}_n(\cdot)$ in place of $\overline{\varphi}_n(\cdot)$. Moreover, if $\acute{\Delta}_n^{\text{AS}} \subsetneq \hat{\Delta}_n^{\text{AS}}$ and $\acute{\Delta}_n^{\text{AS}} \neq \emptyset$, then $\hat{T}^\star_{n,s} > \acute{T}^\star_{n,s}$ holds with positive probability conditional on the sample. In consequence, we expect that results analogous to those in the paper regarding the LSW test and its modification would also hold for the AS test.

C.4. Equivalence Under The Null Hypothesis

Appendix C.2 describes subsets of $\mathcal{P}_0$ on which the AS test is asymptotically level α. These subsets differ from the subsets on which the LSW test is asymptotically level α. Hence, we must characterize subsets of $\mathcal{P}_0$ on which the AS test and its modification are asymptotically equivalent.

Intuitively, we specify subsets of $\mathcal{P}_0$ on which the AS procedure is asymptotically valid that are relevant for the empirical likelihood framework this paper introduces. For each $s \in \mathbb{Z}_+,\ c_1 \in (0,1), c_2 \in \left(0, (\bar{t}-\underline{t})^{-2}\right),\ c_3 \in (0,+\infty)$ and $\mathcal{C}_{cpt}$ compact subset of $\mathcal{C}$, these subsets are

$$\mathcal{P}_0^s(c_1,c_2,c_3,\mathcal{C}_{cpt}) = \{P \in \mathcal{P}_0 : h_{2P} \in \mathcal{C}_{cpt}\} \cap \mathcal{P}_0^s(c_1,c_2,c_3), \tag{C.17}$$

where $\mathcal{P}_0^s(c_1,c_2,c_3)$ is given by Definition 4.1.

The characterization (C.17) follows from an application of Proposition F.1 in Appendix F.3 to

$$\acute{\xi}_n(t) = \kappa_n^{-1}\sqrt{n}\left(\sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i;t\right)\right)/\hat{\sigma}(t) \tag{C.18}$$

$$= \xi_n(t) + \kappa_n^{-1}\sqrt{n}\left(\sum_{i=1}^{n} (\acute{p}_i - \frac{1}{n}) g\left(\mathbf{X}_i;t\right)\right)/\hat{\sigma}(t) \tag{C.19}$$

appearing in the GMS function $\acute{\varphi}_n(t)$. It yields

$$\acute{\xi}_n(t) = \xi_n(t) + O_P\left(\kappa_n^{-1}\right) \text{ uniformly in } \mathcal{P}_0^s(c_1,c_2,c_3,\mathcal{C}_{cpt}), \tag{C.20}$$

$$= \xi_n(t) + o_P(1) \text{ uniformly in } \mathcal{P}_0^s(c_1,c_2,c_3,\mathcal{C}_{cpt}). \tag{C.21}$$

The large-sample behavior (C.21) implies that with probability tending to unity, $\acute{\Delta}_n^{\mathrm{AS}}$ and $\hat{\Delta}_n^{\mathrm{AS}}$ tend to $\Delta(P_0)$ as $n \to +\infty$, uniformly in $\mathcal{P}_0^s(c_1,c_2,c_3,\mathcal{C}_{cpt})$.

C.5. Relative Behavior Under Local Alternatives

Given $s \in \mathbb{Z}_+$, $c_1 \in (0,1), c_2 \in \left(0, (\bar{t}-\underline{t})^{-2}\right)$, $c_3 \in (0,+\infty)$ and $\mathcal{C}_{cpt}$ compact subset of $\mathcal{C}$, let $\mathcal{P}_0^s(c_1,c_2,c_3,\mathcal{C}_{cpt})$ be the model of the null hypothesis. Consider a sequence of local alternatives $\{P_n\}_{n\geq 1} \subset \mathcal{P} - \mathcal{P}_0$ such that $\lim_{n\to+\infty} P_n \in \partial\mathcal{P}_0^s(c_1,c_2,c_3,\mathcal{C}_{cpt})$, where $\partial\mathcal{P}_0^s(c_1,c_2,c_3,\mathcal{C}_{cpt})$ denotes the boundary of the null model.

Denote the modified AS test by $\acute{\tau}^{AS}$. It was shown in Appendix C.3 that the bootstrap critical value from $\acute{\tau}^{AS}$ is never larger than that from $\hat{\tau}^{AS}$, for any $P \in \mathcal{P}$ when n is large enough. Because the test statistics in these tests are identical, the ordering of their bootstrap critical values implies that

$$E_{P_n}\acute{\tau}^{AS} \geq E_{P_n}\hat{\tau}^{AS}, \tag{C.22}$$

holds, for large enough n. Moreover, in light of the discussion in Appendix C.3, this inequality in their local powers holds strictly whenever

$$P_n\left[\acute{\Delta}_n^{\mathrm{AS}} \subsetneq \hat{\Delta}_n^{\mathrm{AS}} \text{ and } \acute{\Delta}_n^{\mathrm{AS}} \neq \emptyset\right] > 0, \tag{C.23}$$

for large enough n.

In consequence, results analogous to Theorem 6.3 and Corollary 6.2 for the LSW test and its modified version hold for comparison of between $\acute{\tau}^{AS}$ and $\hat{\tau}^{AS}$. A remarkable point is that additional regularity conditions on $\{P_n\}_{n\geq 1}$ are not required for these results, which is in contrast to the comparison of the LSW test and its modified version (see Assumption6.1), which is due to the set inclusion $\acute{\Delta}_n^{\mathrm{AS}} \subset \hat{\Delta}_n^{\mathrm{AS}}$ holding for large enough n.

D. Proofs of Results

D.1. Theorem 4.1

Proof. The proof proceeds by the direct method. The following random sets are used in the proof:

$$\mathcal{H}^0(\mathbf{X}) = \left\{ \mathbf{p} \in \mathcal{H}_n : \sum_{i=1}^n p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t \in [\underline{t}, \overline{t}] \right\}, \tag{D.1}$$

$$\mathcal{H}_n^0(\mathbf{X}) = \left\{ \mathbf{p} \in \mathcal{H}_n : \sum_{i=1}^n p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t \in \mathcal{T}_n \right\}, \tag{D.2}$$

$$\mathcal{H}_n = \left\{ \mathbf{p} \in \mathbb{R}^n : \sum_{i=1}^n p_i = 1, p_i \geq 0 \, \forall i \right\} \quad \text{and} \tag{D.3}$$

$$\mathcal{H}_n^\circ = \left\{ \mathbf{p} \in \mathbb{R}^n : \sum_{i=1}^n p_i = 1, p_i > 0 \, \forall i \right\}. \tag{D.4}$$

Part 1. This part of the proof covers the case of first-order stochastic dominance i.e. $s = 1$. The moment functions in this case are therefore of the following form:

$$g(X_i; t) = 1\left[X_i^B \leq t\right] - 1\left[X_i^A \leq t\right] \quad t \in [\underline{t}, \overline{t}]. \tag{D.5}$$

As the difference between $\mathcal{H}_n^0(\mathbf{X})$ and $\mathcal{H}^0(\mathbf{X})$ is that the former constraint set is based on a subset of the inequality constraints that define the latter constraint set, it follows that $\mathcal{H}^0(\mathbf{X}) \subseteq \mathcal{H}_n^0(\mathbf{X})$ holds regardless of the underlying probability. To conclude the proof, we need to show the reverse set inclusion, and then apply Lemma E.6 to establish that $\mathcal{H}^0(\mathbf{X})$ is asymptotically non-empty, with uniformity.

Now we will show that the event $\mathcal{H}_n^0(\mathbf{X}) \subseteq \mathcal{H}^0(\mathbf{X})$ occurs regardless of the underlying probability. We have that $\forall t \in [\underline{t}, \overline{t}] - \mathcal{T}_n \; \exists j \in \{1, 2, \ldots, N\}$ such that

$$t_{(j)} \leq t \leq t_{(j+1)}. \tag{D.6}$$

Because the moment functions are of the form (D.5) for each $i = 1, \ldots, n$, it follows that

$$g(X_i; t) = g(X_i; t_{(j)}) \quad \forall t \in [\underline{t}, \overline{t}] - \mathcal{T}_n. \tag{D.7}$$

Hence, $\forall \mathbf{p} \in \mathcal{H}_n^0(\mathbf{X})$

$$\sum_{i=1}^{n} p_i g(X_i;t) = \sum_{i=1}^{n} p_i g(X_i;t_{(j)}) \leq 0 \quad t \in [\underline{t},\overline{t}] - \mathcal{T}_n. \tag{D.8}$$

Furthermore, $\mathbf{p} \in \mathcal{H}_n^0(\mathbf{X})$ implies $\sum_{i=1}^n g(X_i;t) \leq 0 \ \forall t \in \mathcal{T}_n$. Putting these two parts together yields $\forall \mathbf{p} \in \mathcal{H}_n^0(\mathbf{X})$ that $\sum_{i=1}^n p_i g(X_i;t) \leq 0 \quad \forall t \in [\underline{t},\overline{t}]$.

So we just proved that $\mathcal{H}_n^0(\mathbf{X}) \subseteq \mathcal{H}^0(\mathbf{X})$ holds, which now in conjunction with the set inclusion $\mathcal{H}^0(\mathbf{X}) \subseteq \mathcal{H}_n^0(\mathbf{X})$, implies $\mathcal{H}_n^0(\mathbf{X}) = \mathcal{H}^0(\mathbf{X})$. Lemma E.5 establishes that when the extrema of the SIP and its discretized counterpart exist, then the solution sets are both singletons equal to $\tilde{\mathbf{p}}$ and $\acute{\mathbf{p}}$, respectively. The constraint set equality we just proved implies the equality $\tilde{\mathbf{p}} = \acute{\mathbf{p}}$. Therefore, to conclude the proof, we need to show that given $c_1 \in (0,1)$, $\tilde{\mathbf{p}}$ exists and is the unique solution of the SIP problem with probability approaching unity, uniformly over the set of probabilities $\mathcal{P}_1(c_1)$. Lemma E.6 establishes this result:

$$1 = \lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)} P\left[\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right] \leq \lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)} P\left[\tilde{\mathbf{p}} \text{ exists and is unique}\right] \tag{D.9}$$

because $\left\{\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right\} \subseteq \left\{\tilde{\mathbf{p}} \text{ exists and is unique}\right\}$, where $\mathcal{H}_n^\circ$ denotes the interior of the $n-1$ simplex $\mathcal{H}_n$. This concludes the proof for this part of the theorem.

Part 2. We first present a sketch of the proof because the main steps in it use the intermediate technical results presented in Appendix E.

Sketch of Proof. The proof proceeds using the direct method. We derive an upper bound on $\|\acute{\mathbf{p}} - \tilde{\mathbf{p}}\|$ which converges to zero in probability, with uniformity over set of probabilities $\mathcal{P}_1(c_1) \bigcap \mathcal{P}_2(c_2)$. The derivations in the proof are based on the the occurrence of the event

$$\left\{\tilde{\mathbf{p}} \text{ exists and is unique}\right\}.$$

That is, on this event, we construct the upper bound on $\|\acute{\mathbf{p}} - \tilde{\mathbf{p}}\|$. The result then follows directly since Lemma E.6 in Appendix E.2 establishes that, for each $c_1 \in (0,1)$, the probability of $\{\tilde{\mathbf{p}} \text{ exists and is unique}\}$ converges to unity with uniformity over the set of probabilities $\mathcal{P}_1(c_1)$. The proof proceeds in 4 steps, where the details of Steps 1,2 and 3 are based on the contents in Appendices E.2, E.5 and E.6, respectively.

Step 1. On the event that $\{\tilde{\mathbf{p}} \text{ exists and is unique}\}$, Part 2 of Lemma E.5 in Appendix E.2 implies the occurrence of the event $\{\acute{\mathbf{p}} \text{ exists and is unique}\}$, because the set inclusion $\mathcal{H}^0(\mathbf{X}) \subseteq \mathcal{H}_n^0(\mathbf{X})$, holds.

Step 2. By Lemma E.9 in Appendix E.5, we can construct a probability vector $\check{\mathbf{p}}$ in $\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ$ that is nearby to $\acute{\mathbf{p}}$. The consequence of this step is that by the triangular inequality we can conclude

$$\|\acute{\mathbf{p}} - \tilde{\mathbf{p}}\| \leq \|\check{\mathbf{p}} - \tilde{\mathbf{p}}\| + \|\acute{\mathbf{p}} - \check{\mathbf{p}}\|, \quad \text{where} \quad \|\acute{\mathbf{p}} - \check{\mathbf{p}}\| \leq \frac{\acute{p}_{(1)}}{2n^{3/2}}. \tag{D.10}$$

Step 3. Since $\breve{\mathbf{p}} \in \mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^{\circ}$, there exists a large enough $\epsilon > 0$ such that $\|\breve{\mathbf{p}} - \tilde{\mathbf{p}}\| \leq \epsilon$. Then Lemma E.10 in Appendix E.6 establishes, via a quadratic growth condition in a neighborhood of $\tilde{\mathbf{p}}$, that

$$\|\breve{\mathbf{p}} - \tilde{\mathbf{p}}\|^2 \leq \frac{h(\breve{\mathbf{p}}) - h(\tilde{\mathbf{p}})}{K_0} = \frac{h(\breve{\mathbf{p}}) - h(\acute{\mathbf{p}})}{K_0} + \frac{h(\acute{\mathbf{p}}) - h(\tilde{\mathbf{p}})}{K_0}, \tag{D.11}$$

where $K_0 = 1/(2\epsilon^2 + 10)$ and $h(\mathbf{p}) = -\sum_{i=1}^n \log(p_i)$. Then noting that $h(\acute{\mathbf{p}}) - h(\tilde{\mathbf{p}}) \leq 0$ holds as $\mathcal{H}^0(\mathbf{X}) \subseteq \mathcal{H}_n^0(\mathbf{X})$, it follows that

$$\|\breve{\mathbf{p}} - \tilde{\mathbf{p}}\|^2 \leq \frac{h(\breve{\mathbf{p}}) - h(\acute{\mathbf{p}})}{K_0}. \tag{D.12}$$

Step 4. Combining the inequalities in (D.10) and (D.12), yields

$$\|\acute{\mathbf{p}} - \tilde{\mathbf{p}}\| \leq \left[\frac{h(\breve{\mathbf{p}}) - h(\acute{\mathbf{p}})}{K_0}\right]^{1/2} + \frac{\acute{p}_{(1)}}{2n^{3/2}}. \tag{D.13}$$

In consequence, to prove the result of this part of the theorem, we need to show that the two parts on the right side of (D.13) converge to zero in probability, with uniformity over sets of probabilities $\mathcal{P}_1(c_1) \bigcap \mathcal{P}_2(c_2)$. Observe that by the inequality $\log(1+y) \leq y \; \forall y > -1$

$$h(\breve{\mathbf{p}}) - h(\acute{\mathbf{p}}) = \sum_{i=1}^n \log\left(1 + \frac{\acute{p}_i - \breve{p}_i}{\breve{p}_i}\right) \leq \sum_{i=1}^n \frac{\acute{p}_i - \breve{p}_i}{\breve{p}_i}. \tag{D.14}$$

Then using the construction of the probability vector in Part 2 of Lemma E.9 i.e. $\breve{p}_i \geq \acute{p}_{(1)}/2$ and $|\acute{p}_i - \breve{p}_i| \leq \frac{\acute{p}_{(1)}}{2n^2} \;\; \forall i$, and the Cauchy-Schwartz inequality for sums yields

$$\sum_{i=1}^n \frac{\acute{p}_i - \breve{p}_i}{\breve{p}_i} \leq \sqrt{\sum_{i=1}^n \frac{1}{\breve{p}_i^{-2}}}\, \|\acute{\mathbf{p}} - \breve{\mathbf{p}}\| \leq \frac{2\sqrt{n}}{\acute{p}_{(1)}} \|\acute{\mathbf{p}} - \breve{\mathbf{p}}\| \leq 1/n. \tag{D.15}$$

Hence, from the inequality (D.13) it follows that

$$\|\acute{\mathbf{p}} - \tilde{\mathbf{p}}\| \leq (K_0 n)^{-1/2} + n^{-3/2}, \tag{D.16}$$

because $\acute{p}_{(1)} < 1$.

Therefore, the steps described above yields the following event inclusion

$$\{\tilde{\mathbf{p}} \text{ exists and is unique}\} \subseteq \left\{\|\acute{\mathbf{p}} - \tilde{\mathbf{p}}\| \leq (K_0 n)^{-1/2} + n^{-3/2}\right\}. \tag{D.17}$$

Hence, given $c_1 \in (0,1)$ and $c_2 \in \left(0, \frac{1}{(\bar{t}-\underline{t})^2}\right)$, Lemmas E.5 and E.6 establish

$$1 = \lim_{n\to+\infty} \inf_{P \in \mathcal{P}_1(c_1)} P[\tilde{\mathbf{p}} \text{ exists and is unique}] \tag{D.18}$$

$$\leq \lim_{n\to+\infty} \inf_{P \in \mathcal{P}_1(c_1) \bigcap \mathcal{P}_2(c_2)} P\left[\|\acute{\mathbf{p}} - \tilde{\mathbf{p}}\| \leq (K_0 n)^{-1/2} + n^{-3/2}\right], \tag{D.19}$$

which implies the desired result and concludes the proof.

D.2. Theorem 5.1

PROOF. **Part 1**. The proof proceeds by the direct method. Lemma F.2 implies that the $\acute{p}_i$ can be characterized in terms of Lagrange multipliers as in (F.31). Without loss of generality, let

$$\Delta(\acute{P}_n) = \{t_1, t_2, \ldots, t_{\omega_n}\}\,. \tag{D.20}$$

Therefore, the probabilities (F.31) can be expressed as

$$\acute{p}_i = \frac{1}{n}\left(1 + \sum_{j=1}^{\omega_n} \acute{\mu}_j g(X_i; t_j)\right)^{-1}. \tag{D.21}$$

Therefore,

$$E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right] - E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right] = -\sum_{i=1}^{n} \acute{p}_i \sum_{t' \in \Delta(\acute{P}_n)} \acute{\mu}_{t'} g\left(\mathbf{X}_i; t'\right) g\left(\mathbf{X}_i; t\right), \tag{D.22}$$

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

$$g\left(\mathbf{x}; t'\right) g\left(\mathbf{x}; t\right) \geq 0 \quad \forall (t, \mathbf{x}) \in [\underline{t}, \overline{t}] \times \operatorname{supp}(P)\,. \tag{D.23}$$

On the event $\left\{\acute{P}_n \neq \hat{P}_n\right\}$, the Lagrange multipliers cannot all be equal to zero. Then Lemma E.1 implies that there exists X_i such that $g\left(\mathbf{X}_i; t'\right) g\left(\mathbf{X}_i; t\right) > 0\ \forall t$, which concludes the proof.

Part 2. First, we prove the probability of the event $\left\{\acute{\Delta}_n \subseteq \hat{\Delta}_n\right\}$ (in $\mathcal{A}_n$) converges to unity in probability uniformly over $\mathcal{P}_0^s(c_1, c_2, c_3)$, 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 $t \in \acute{\Delta}_n$,

$$\begin{aligned} E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right] &= E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right] - E_{\acute{P}_n}\left[g\left(\mathbf{X}, t\right)\right] + E_{\acute{P}_n}\left[g\left(\mathbf{X}, t\right)\right] \\ &< E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right] - E_{\acute{P}_n}\left[g\left(\mathbf{X}, t\right)\right] + r_n \\ &= O_P(n^{-1/2}) + r_n \quad \text{uniformly over} \quad \mathcal{P}_0^s(c_1, c_2, c_3), \end{aligned} \tag{D.24}$$

where (D.24) follows by Proposition F.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 $P[E_{\hat{P}_n}[g(\mathbf{X},t)] \in [r_n, O_P(n^{-1/2})+r_n)]$, which is equal to

$$P\left[\sqrt{n}\left(E_{\hat{P}_n}[g(\mathbf{X},t)] - E_P[g(\mathbf{X},t)]\right) + \sqrt{n}E_P[g(\mathbf{X},t)] \in [\sqrt{n}r_n, O_P(1)+\sqrt{n}r_n)\right]. \tag{D.25}$$

For $t \in \Delta(P)$, this probability is equal to

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

and the Uniform Central Limit Theorem establishes that $\sqrt{n}E_{\hat{P}_n}[g(\mathbf{X},t)] = O_p(1)$, uniformly over $\mathcal{P}_0^s(c_1,c_2,c_3)$. Because $\sqrt{n}r_n \to +\infty$, it follows that

$$\sup_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)} P\left[\sqrt{n}E_{\hat{P}_n}[g(\mathbf{X},t)] \in [\sqrt{n}r_n, O_P(1)+\sqrt{n}r_n)\right] \to 0. \tag{D.26}$$

Therefore, if $t \in \Delta(P)$, then $\inf_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)} P\left[E_{\hat{P}_n}[g(\mathbf{X},t)] < r_n\right] \to 1$.

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

$$\sqrt{n}E_{\hat{P}_n}[g(\mathbf{X},t)] = \sqrt{n}\left(E_{\hat{P}_n}[g(\mathbf{X},t)] - E_P[g(\mathbf{X},t)]\right) + \sqrt{n}E_P[g(\mathbf{X},t)] \tag{D.27}$$

$$O_P(1) + \sqrt{n}E_P[g(\mathbf{X},t)] \quad \text{uniformly over} \quad \mathcal{P}_0^s(c_1,c_2,c_3). \tag{D.28}$$

Note that $\sqrt{n}E_P[g(\mathbf{X},t)]$ diverges to $-\infty$, $\sqrt{n}E_{\hat{P}_n}[g(\mathbf{X},t)]$ also diverges to $-\infty$, but uniformly over $\mathcal{P}_0^s(c_1,c_2,c_3)$. Combining this result with the fact that $\sqrt{n}r_n \to +\infty$, implies that the probability (D.25) tends to zero with uniformity. Therefore, $\inf_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)} P\left[E_{\hat{P}_n}[g(\mathbf{X},t)] < r_n\right] \to 1$, which concludes the proof of this part of the theorem.

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

$$\left\{-r_n < E_{\acute{P}_n}[g(\mathbf{X},t)] < r_n\right\} = \left\{-r_n < E_{\acute{P}_n}[g(\mathbf{X},t)] - E_{\hat{P}_n}[g(\mathbf{X},t)] + E_{\hat{P}_n}[g(\mathbf{X},t)] < r_n\right\} \tag{D.29}$$

$$= \left\{-r_n < O_P(n^{-1/2}) + E_{\hat{P}_n}[g(\mathbf{X},t)] < r_n\right\} \tag{D.30}$$

$$= \left\{-\sqrt{n}r_n < O_P(1) + \sqrt{n}E_{\hat{P}_n}[g(\mathbf{X},t)] < \sqrt{n}r_n\right\}, \tag{D.31}$$

uniformly over $\mathcal{P}_0^s(c_1,c_2,c_3)$ by Proposition F.1. As $t \in \Delta(P)$, we have that $\sqrt{n}E_{\hat{P}_n}[g(\mathbf{X},t)] = O_P(1)$ uniformly over $\mathcal{P}_0^s(c_1,c_2,c_3)$, by the Uniform Central Limit Theorem. Therefore, the event (F.60) is equal to

$$\{-\sqrt{n}r_n \le O_P(1) \le \sqrt{n}r_n\}, \tag{D.32}$$

whose probability tends to unity uniformly over $\mathcal{P}_0^s(c_1,c_2,c_3)$, because $\sqrt{n}r_n \to +\infty$. This concludes part 2.

Part 3. The proof proceeds by the direct method. We have that

$$\begin{aligned}\left\{\hat{\Delta}_n \subsetneq \acute{\Delta}_n\right\} &= \left\{\hat{\Delta}_n \subsetneq \acute{\Delta}_n \text{ and } \hat{P}_n \neq \acute{P}_n\right\} \\ &\subseteq \left\{\exists t \in \acute{\Delta}_n; E_{\acute{P}_n}\left[g\left(\mathbf{X}, t\right)\right] < r_n \leq E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right]\right\} \\ &= \left\{\exists t \in \acute{\Delta}_n; E_{\acute{P}_n}\left[g\left(\mathbf{X}, t\right)\right] - E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right] < r_n - E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right] \leq 0\right\} \\ &= \left\{\exists t \in \acute{\Delta}_n; O_P\left(n^{-1/2}\right) < r_n - O_P\left(n^{-1/2}\right) - E_P\left[g\left(\mathbf{X}, t\right)\right] \leq 0\right\} \\ &= \left\{\exists t \in \acute{\Delta}_n; O_P\left(1\right) < \sqrt{n} r_n - O_P\left(1\right) - \sqrt{n} E_P\left[g\left(\mathbf{X}, t\right)\right] \leq 0\right\},\end{aligned}$$

for any $t \in [\underline{t}, \overline{t}]$, uniformly over $\mathcal{P}_0^s(c_1, c_2, c_3)$ by Proposition F.1. Since for any $t \in [\underline{t}, \overline{t}]$,

$$\sqrt{n} r_n - O_P\left(1\right) - \sqrt{n} E_P\left[g\left(\mathbf{X}, t\right)\right]$$

diverges to $+\infty$ as $n \to +\infty$, uniformly over $\mathcal{P}_0^s(c_1, c_2, c_3)$, it implies that the event of it being non-positive tends to zero, with uniformity; i.e.,

$$\lim_{n\to+\infty} \sup_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)} P\left[\exists t \in \acute{\Delta}_n; O_P\left(1\right) < \sqrt{n} r_n - O_P\left(1\right) - \sqrt{n} E_P\left[g\left(\mathbf{X}, t\right)\right] \leq 0\right] = 0.$$

This concludes the proof since, for each $P \in \mathcal{P}_0^s(c_1, c_2, c_3)$, the above set inclusions imply that

$$P\left[\hat{\Delta}_n \subsetneq \acute{\Delta}_n\right] \leq P\left[\exists t \in \acute{\Delta}_n; O_P\left(1\right) < \sqrt{n} r_n - O_P\left(1\right) + \sqrt{n} E_P\left[g\left(\mathbf{X}, t\right)\right] \leq 0\right],$$

holds for each n.

Part 4. The proof proceeds by the direct method. First note the following decomposition of the event $\left\{\acute{\Delta}_n \subsetneq \hat{\Delta}_n\right\}$:

$$\left\{\acute{\Delta}_n \subsetneq \hat{\Delta}_n\right\} = \left\{\acute{P}_n \neq \hat{P}_n\right\} \bigcap \left\{\exists t \in \hat{\Delta}_n; \left|E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right| \geq r_n\right\} \bigcap \left\{\acute{\Delta}_n \subseteq \hat{\Delta}_n\right\}. \quad \text{(D.33)}$$

Noting that $\left\{\forall t \in \hat{\Delta}_n; \left|E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right| < r_n\right\} = \left\{\hat{\Delta}_n \subseteq \acute{\Delta}_n\right\}$, the probability of the complement of (D.33) is

$$P\left[\left\{\acute{P}_n = \hat{P}_n\right\} \bigcup \left\{\forall t \in \hat{\Delta}_n; E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right] > -r_n\right\} \bigcup \left\{\acute{\Delta}_n \not\subseteq \hat{\Delta}_n\right\}\right],$$

which equals

$$P\left[\acute{P}_n = \hat{P}_n\right] + P\left[\hat{\Delta}_n \subseteq \acute{\Delta}_n\right] + P\left[\acute{\Delta}_n \not\subseteq \hat{\Delta}_n\right] - P\left[\acute{\Delta}_n = \hat{\Delta}_n\right] - P\left[\hat{\Delta}_n \subsetneq \acute{\Delta}_n\right]$$

and simplifies to $P\left[\acute{P}_n = \hat{P}_n\right] + P\left[\acute{\Delta}_n \not\subseteq \hat{\Delta}_n\right]$. Then,

$$\begin{aligned}\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)\bigcap\mathcal{P}_{000}} P\left[\acute{\Delta}_n \subsetneq \hat{\Delta}_n\right] &= 1 - \lim_{n\to+\infty} \sup_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)\bigcap\mathcal{P}_{000}} P\left[\acute{P}_n = \hat{P}_n\right] \\ &\quad - \lim_{n\to+\infty} \sup_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)\bigcap\mathcal{P}_{000}} P\left[\acute{\Delta}_n \not\subseteq \hat{\Delta}_n\right] \\ &\geq \frac{1}{2}\end{aligned}$$

because Lemma F.1 establishes that $\lim_{n\to+\infty}\sup_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)\bigcap\mathcal{P}_{000}} P\left[\acute{P}_n=\hat{P}_n\right]\leq 1/2$, and by Part 2 of Theorem 5.1

$$\lim_{n\to+\infty}\sup_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)\bigcap\mathcal{P}_{000}} P\left[\acute{\Delta}_n\not\subseteq\hat{\Delta}_n\right]\leq\lim_{n\to+\infty}\sup_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)} P\left[\acute{\Delta}_n\not\subseteq\hat{\Delta}_n\right]=0.$$

Part 5. The proof follows identical steps to those in the second part of the proof of Claim 1, in LSW on page 200.

D.3. Theorem 5.2

Proof. The proof proceeds by the direct method. As the test statistic is the same, it is sufficient to show that the proposed and LSW bootstrap test statistics are asymptotically equal with uniformity. Let

$$\gamma_n^\star(t)=\left(\max\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[g\left(\mathbf{X}_{i,l}^\star;t\right)-E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]\right],0\right\}\right)^2, \tag{D.34}$$

then consider the following,

$$\left|\hat{T}_{n,l}^\star-\acute{T}_{n,l}^\star\right|=\begin{cases}\int_{[\underline{t},\overline{t}]-\acute{\Delta}_n}\gamma_n^\star(t)\,dt & \text{if } \int_{\acute{\Delta}_n}dt>0,\int_{\hat{\Delta}_n}dt=0\\ \int_{[\underline{t},\overline{t}]-\hat{\Delta}_n}\gamma_n^\star(t)\,dt & \text{if } \int_{\acute{\Delta}_n}dt=0,\int_{\hat{\Delta}_n}dt>0\\ \int_{\acute{\Delta}_n\ominus\hat{\Delta}_n}\gamma_n^\star(t)\,dt & \text{if } \int_{\acute{\Delta}_n}dt>0,\int_{\hat{\Delta}_n}dt>0\\ 0 & \text{if } \int_{\acute{\Delta}_n}dt=0,\int_{\hat{\Delta}_n}dt=0,\end{cases} \tag{D.35}$$

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

$$\left|\hat{T}_{n,l}^\star-\acute{T}_{n,l}^\star\right|\leq\begin{cases}\left(\sup_{t\in[\underline{t},\overline{t}]}\gamma_n^\star(t)\right)\int_{[\underline{t},\overline{t}]-\acute{\Delta}_n}dt & \text{if } \int_{\acute{\Delta}_n}dt>0,\int_{\hat{\Delta}_n}dt=0\\ \left(\sup_{t\in[\underline{t},\overline{t}]}\gamma_n^\star(t)\right)\int_{[\underline{t},\overline{t}]-\hat{\Delta}_n}dt & \text{if } \int_{\acute{\Delta}_n}dt=0,\int_{\hat{\Delta}_n}dt>0\\ \left(\sup_{t\in[\underline{t},\overline{t}]}\gamma_n^\star(t)\right)\int_{\acute{\Delta}_n\ominus\hat{\Delta}_n}dt & \text{if } \int_{\acute{\Delta}_n}dt>0,\int_{\hat{\Delta}_n}dt>0,\\ 0 & \text{if } \int_{\acute{\Delta}_n}dt=0,\int_{\hat{\Delta}_n}dt=0.\end{cases} \tag{D.36}$$

To prove the result we need to prove that $\left(\sup_{t\in[\underline{t},\overline{t}]}\gamma_n^\star(t)\right)$ is $O_P(1)$ conditional on $\mathcal{A}_n$, uniformly in $\mathcal{P}_0^s(c_1,c_2,c_3,\epsilon)$. and then apply Theorem 5.1 to the integrals in (D.36). Since the set of moment functions $\left\{\mathbf{x}\mapsto g(\mathbf{x},t),t\in[\underline{t},\overline{t}]\right\}$ is uniform Donsker with respect to $\mathcal{P}_0^s(c_1,c_2,c_3,\epsilon)$, 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 $\left(\sup_{t\in[\underline{t},\overline{t}]}\gamma_n^\star(t)\right)$ yields the desired result.

Parts 3 and 2 of Theorem 5.1 imply that $\hat{\Delta}_n$ and $\acute{\Delta}_n$ are consistent estimators of $\Delta(P)$ uniformly in $\mathcal{P}_0^s(c_1,c_2,c_3)$, as $\mathcal{P}_0^s(c_1,c_2,c_3,\epsilon)\subset\mathcal{P}_0^s(c_1,c_2,c_3)$. Noting that $P\in\mathcal{P}_0^s(c_1,c_2,c_3,\epsilon)$ implies that $\int_{\Delta(P)}dt>0$, for large n, we must have $\acute{\Delta}_n\neq\emptyset,\hat{\Delta}_n\neq\emptyset$ with probability tending to one, uniformly in $\mathcal{P}_0^s(c_1,c_2,c_3,\epsilon)$. Applying Part 2 of Theorem 5.1 to this case in (D.36) implies $\acute{\Delta}_n\ominus\hat{\Delta}_n=\hat{\Delta}_n-\acute{\Delta}_n$ with probability approaching unity uniformly in $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$. Consequently, by Parts 2 and 3 of Theorem 5.1

$$\left(\sup_{t\in[\underline{t},\overline{t}]} \gamma_n^{\star}(t)\right) \int_{\acute{\Delta}_n \ominus \hat{\Delta}_n} dt \xrightarrow{P} 0 \tag{D.37}$$

conditional on $\mathcal{A}_n$ uniformly in $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$. Therefore, $\hat{T}^{\star}_{n,l} - \acute{T}^{\star}_{n,l} \xrightarrow{P} 0$ conditional on $\mathcal{A}_n$ uniformly in $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$. This concludes the proof.

D.4. Theorem 6.1

PROOF. The proof proceeds by the direct method. Because Theorem 3 of LSW shows the test statistic

$$\hat{T}_n = n \int_{\underline{t}}^{\overline{t}} \max\left\{E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right], 0\right\}^2 dt, \tag{D.38}$$

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.

Corollary E.1 and Part 2 of Lemma E.5 implies 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

$$\int_{\underline{t}}^{\overline{t}} \left(\max\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[g\left(\mathbf{X}_i^{\star}; t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right], 0\right\}\right)^2 dt, \tag{D.39}$$

which converges in distribution (conditional on $\mathcal{A}_n$) to the distribution of

$$\int_{\underline{t}}^{\overline{t}} \left(\max\left\{\nu(t), 0\right\}\right)^2 dt,$$

it follows that $\acute{T}_n^{\star} = O_P(1)$ conditional on $\mathcal{A}_n$. This concludes the proof.

D.5. Theorem 6.2

PROOF. **Part 1**. The proof follows steps identical to those in Part 1 of Theorem 5.1 except that we are taking limits under the local alternatives, which is based on Lemma G.4. We omit the details for brevity.

Part 2. The proof proceeds by the direct method. We first focus on proving $\lim_{n\to+\infty} P_n\left[C \subseteq \acute{\Delta}_n\right] = 1$. Lemma G.4 implies that the random set $\acute{\Delta}_n$ exists for large enough n, with probability approaching unity under the local alternatives. 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

$$\left\{-\frac{\sqrt{n}r_n}{\sigma_n(t)} \leq \frac{\sqrt{n}\left(E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right]\right)}{\sigma_n(t)} + \frac{\sqrt{n}\left(E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{P_n}\left[g\left(\mathbf{X}, t\right)\right]\right)}{\sigma_n(t)} + \frac{\delta(t)}{\sigma_n(t)} \leq \frac{\sqrt{n}r_n}{\sigma_n(t)}\right\}, \tag{D.40}$$

where

$$\sqrt{n}\left(E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{\hat{P}_n}\left[g\left(\mathbf{X}, t\right)\right]\right) = -\sqrt{n}\sum_{t' \in \Delta(\acute{P}_n)} \acute{\mu}_{t'} \sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i; t'\right) g\left(\mathbf{X}_i; t\right). \tag{D.41}$$

We can follow steps identical to those in Proposition F.1 to deduce that

$$\sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i; t'\right) g\left(\mathbf{X}_i; t\right) = O_{P_n}(1)$$

under the local alternatives, and we omit them for brevity. In consequence, Part 5 of Lemma G.4 implies the right side of (D.41) is $O_{P_n}(1)$. Noting that

$$-\frac{\sqrt{n}r_n}{\sigma_n(t)} \to -\infty, \frac{\sqrt{n}r_n}{\sigma_n(t)} \to +\infty, \frac{\sqrt{n}\left(E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{P_n}\left[g\left(\mathbf{X}, t\right)\right]\right)}{\sigma_n(t)} = O_{P_n}(1)$$

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 6.1, it follows that

$$\lim_{n \to +\infty} P_n\left[C \subseteq \acute{\Delta}_n\right] = 1.$$

Next we focus on proving $\lim_{n \to +\infty} P_n\left[\acute{\Delta}_n \subseteq \hat{\Delta}_n\right] = 1$. Parts 1 and 2 of Lemma G.4 implies that asymptotically the set estimator $\acute{\Delta}_n$ exist with probability approaching unity under the local alternatives. Suppose that $t \in \acute{\Delta}_n$, then by Part 4 of Lemma G.4, $\lim_{n \to +\infty} P_n\left[E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] > -r_n\right] = 1$. Thus, to prove the result we need to show that $\lim_{n \to +\infty} P_n\left[E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] < r_n\right] = 1$.

We will show that $\lim_{n \to +\infty} P_n\left[E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \geq r_n\right] = 0$. Noting that

$$E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] = E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{P_n}\left[g\left(\mathbf{X}; t\right)\right] + E_{P_n}\left[g\left(\mathbf{X}; t\right)\right], \text{ and} \tag{D.42}$$

$$\sqrt{n}\left(E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{P_n}\left[g\left(\mathbf{X}; t\right)\right]\right)/\sigma_{P_n}(t) = O_{P_n}(1) \tag{D.43}$$

by the Lindeberg-Feller Central Limit Theorem for triangular arrays of row-wise IID random variables, it follows that $\lim_{n \to +\infty} P_n\left[E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \geq r_n\right]$ equals

$$\lim_{n \to +\infty} P_n\left[O_{P_n}(1) + \frac{\sqrt{n}}{\sigma_{P_n}(t)} E_{P_n}\left[g\left(\mathbf{X}; t\right)\right] \geq \sqrt{n}r_n/\sigma_{P_n}(t)\right]. \tag{D.44}$$

As $\sqrt{n}r_n/\sigma_{P_n}(t) \to +\infty$, for $t \in C$, the term $\sqrt{n}E_{P_n}\left[g\left(\mathbf{X};t\right)\right]/\sigma_{P_n}(t) = \delta(t)/\sigma_{P_n}(t)$, is uniformly bounded with uniformity over n; hence, the limit (D.44) is equal to zero. Furthermore, for $t \in [\underline{t}, \overline{t}] - C$, the term

$$\sqrt{n}E_{P_n}\left[g\left(\mathbf{X};t\right)\right]/\sigma_{P_n}(t) = (\sqrt{n}H(t) + \delta(t))/\sigma_{P_n}(t) \to -\infty \tag{D.45}$$

as $H(t) < 0$ and because $\delta(t)/\sigma_{P_n}(t)$ is uniformly bounded in n and t by the conditions of Assumption 6.1. Therefore, the limit (D.44) is equal to zero for such t since the divergence to $-\infty$ in (D.45) is at rate $\sqrt{n}$ whereas the divergence of $\sqrt{n}r_n/\sigma_{P_n}(t)$ to $+\infty$ is slower than $\sqrt{n}$. This concludes the proof.

Part 3. The proof proceeds by the direct method and follows steps identical to those in Part 3 of Theorem 5.1, except that the limits are taken under the sequence of local alternatives. We have that

$$\begin{aligned}
\left\{\hat{\Delta}_n \subsetneq \acute{\Delta}_n\right\} &= \left\{\hat{\Delta}_n \subsetneq \acute{\Delta}_n \text{ and } \hat{P}_n \neq \acute{P}_n\right\} \\
&\subseteq \left\{\exists t \in \acute{\Delta}_n; E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right] < r_n \leq E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right]\right\} \\
&= \left\{\exists t \in \acute{\Delta}_n; E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right] - E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] < r_n - E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] \leq 0\right\} \\
&= \left\{\exists t \in \acute{\Delta}_n; O_{P_n}\left(n^{-1/2}\right) < r_n - O_{P_n}\left(n^{-1/2}\right) - E_{P_n}\left[g\left(\mathbf{X},t\right)\right] \leq 0\right\} \\
&= \left\{\exists t \in \acute{\Delta}_n; O_{P_n}\left(1\right) < \sqrt{n}r_n - O_{P_n}\left(1\right) - \sqrt{n}E_{P_n}\left[g\left(\mathbf{X},t\right)\right] \leq 0\right\},
\end{aligned}$$

for any $t \in [\underline{t}, \overline{t}]$, under the sequence of local alternatives by Proposition G.1. For $t \in C$, we have

$$\sqrt{n}r_n - O_{P_n}\left(1\right) - \sqrt{n}E_{P_n}\left[g\left(\mathbf{X},t\right)\right] = \sqrt{n}r_n - O_{P_n}\left(1\right) - \delta(t) \to +\infty \tag{D.46}$$

as $n \to +\infty$, because $\sqrt{n}r_n \to +\infty$ as $n \to +\infty$ and $\delta(t)$ is finite. Similarly, for $t \in [\underline{t}, \overline{t}] - C$, we have

$$\sqrt{n}r_n - O_{P_n}\left(1\right) - \sqrt{n}E_{P_n}\left[g\left(\mathbf{X},t\right)\right] = \sqrt{n}r_n - O_{P_n}\left(1\right) - \sqrt{n}H(t) - \delta(t) \to +\infty \tag{D.47}$$

as $n \to +\infty$, because $\sqrt{n}r_n \to +\infty$ and $-\sqrt{n}H(t) \to +\infty$ as $n \to +\infty$ and $\delta(t)$ is finite. In consequence, the event (D.46) tends to the empty set under the sequence of local alternatives, which implies that

$$\lim_{n \to +\infty} P_n\left[\hat{\Delta}_n \subsetneq \acute{\Delta}_n\right] = 0,$$

and concludes the proof.

Part 4. The proof proceeds by the direct method. Given $\epsilon > 0$ and $t \in \hat{\Delta}_n$, observe that

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

by the triangular inequality and the definition of $\hat{\Delta}_n$. Hence, to conclude the proof, we need to establish that

$$\lim_{n\to+\infty} P_n\left[\sup_{t\in[\underline{t},\overline{t}]}\left|E_{P_n}[g(\mathbf{X};t)] - E_{\hat{P}_n}[g(\mathbf{X};t)]\right| \le \epsilon r_n\right] = 1, \tag{D.48}$$

holds under the conditions of the theorem. In fact, the empirical process

$$\sqrt{n}\left[E_{P_n}[g(\mathbf{X};t)] - E_{\hat{P}_n}[g(\mathbf{X};t)]\right] \tag{D.49}$$

is *tight* under the sequence of local alternatives, which implies that

$$\lim_{n\to+\infty} P_n\left[\sup_{t\in[\underline{t},\overline{t}]}\sqrt{n}\left|E_{P_n}[g(\mathbf{X};t)] - E_{\hat{P}_n}[g(\mathbf{X};t)]\right| > \epsilon\sqrt{n}r_n\right] = 0, \tag{D.50}$$

holds. This limit result implies the limit (D.48) since it is the limit of the probabilities of the complementary event under the sequence of local alternatives. This concludes the proof.

D.6. Corollary 6.1

PROOF. The proof proceeds by the direct method. Since the test statistics are the same, it suffices to compare the bootstrap p-values of the tests as $B \to +\infty$. They are

$$\acute{\Upsilon}_\infty = P^\star\left[\acute{T}_n^\star \ge \hat{T}_n\right] \quad \text{and} \quad \hat{\Upsilon}_\infty = P^\star\left[\hat{T}_n^\star \ge \hat{T}_n\right], \tag{D.51}$$

which are conditional on $\mathcal{A}_n$. By Theorem 6.2, the bootstrap test statistics $\acute{T}_n^\star$ and $\hat{T}_n^\star$ converge in distribution conditional on $\mathcal{A}_n$ to $\int_{\acute{\Delta}_n} \max\{\nu(t),0\}^2\, dt$ and $\int_{\hat{\Delta}_n} \max\{\nu(t),0\}^2\, dt$, respectively, under the sequence of local alternatives. Furthermore, Theorem 6.2 implies that asymptotically

$$\int_{\acute{\Delta}_n} \max\{\nu(t),0\}^2\, dt \le \int_{\hat{\Delta}_n} \max\{\nu(t),0\}^2\, dt \tag{D.52}$$

holds with probability tending to unity under the sequence of local alternatives. So that

$$\acute{\Upsilon}_\infty = P^\star\left[\int_{\acute{\Delta}_n} \max\{\nu(t), 0\}^2 \geq \hat{T}_n\right] \leq P^\star\left[\int_{\hat{\Delta}_n} \max\{\nu(t), 0\}^2 \geq \hat{T}_n\right] = \hat{\Upsilon}_\infty \tag{D.53}$$

holds asymptotically, conditional on $\mathcal{A}_n$, which implies the following relationship between the rejection events $\left\{\hat{\Upsilon}_\infty \leq \alpha\right\} \subseteq \left\{\acute{\Upsilon}_\infty \leq \alpha\right\}$ holds conditional on $\mathcal{A}_n$, with probability tending to unity under the sequence of local alternatives. In consequence,

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

Furthermore, Part 4 of Theorem 6.2 implies that the contact set estimators $\acute{\Delta}_n$ and $\hat{\Delta}_n$ converge to the set C; therefore, we must have the equality

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

Finally, on the event $\left\{\acute{\Delta}_n \subsetneq \hat{\Delta}_n\right\}$, the inequalities (D.52) and (D.53) hold strictly. As for large n the probability $P_n\left[\acute{\Delta}_n \subsetneq \hat{\Delta}_n\right] > 0$, these inequalities hold strictly with positive probability, which implies that

$$P_n\left[\left(\left\{\acute{\Upsilon}_\infty \leq \alpha\right\} - \left\{\hat{\Upsilon}_\infty \leq \alpha\right\}\right)\bigcap\left\{\acute{\Delta}_n \subsetneq \hat{\Delta}_n\right\}\right] > 0. \tag{D.54}$$

D.7. Theorem 6.3

PROOF. The proof proceeds by the direct method. Let $s \in \mathbb{Z}_+$, $\epsilon > 0$, $c_1 \in (0,1)$, $c_2 \in \left(0, \frac{1}{(\bar{t}-\underline{t})^2}\right)$ and $c_3 \in (0,+\infty)$, be given, and suppose that $\lim_{n\to\infty} P_n \in \lim_{n\to+\infty} P_n \in \mathcal{P}_{00} \bigcap \mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$.

Part 1. By Part 2 of Corollary 6.1, the condition $P_n\left[\acute{\Delta}_n \subsetneq \hat{\Delta}_n\right] > 0$ for large n, implies that $P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] > P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right]$ holds for large n. In conjunction with this implication, the condition $P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right] \nearrow p_\infty \in (0,1)$ as $n \to +\infty$, further implies $P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] \nearrow p_\infty$ as $n \to +\infty$. In consequence,

$$p_\infty > P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] > P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right] \quad \text{for large} \quad n. \tag{D.55}$$

From the convergence of $P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right]$ to p_∞ as $n \to +\infty$, by setting $\gamma_n = p_\infty - P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right]$, $\exists N_{\gamma_n} \in \mathbb{Z}_+$ such that $p_\infty - P_m\left[\hat{\Upsilon}_\infty \leq \alpha\right] < \gamma_n \quad \forall m \geq N_{\gamma_n}$. Substituting out γ_n yields

$$P_m\left[\hat{\Upsilon}_\infty \leq \alpha\right] > P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] \quad \forall m \geq N_{\gamma_n}. \tag{D.56}$$

Hence,

$$\left\{N \in \mathbb{Z}_+ : P_m\left[\hat{\Upsilon}_\infty \leq \alpha\right] \geq P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] \; \forall m \geq N\right\} \neq \emptyset, \tag{D.57}$$

as N_{γ_n} is an element of it. Moreover, the set in (D.57) is bounded from below by the integer 1; therefore, the infimum

$$k_n' = \inf\left\{N \in \mathbb{Z}_+ : P_m\left[\hat{\Upsilon}_\infty \leq \alpha\right] \geq P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] \; \forall m \geq N\right\}, \tag{D.58}$$

exists. Consequently, k_n is equal to the smallest integer that is greater than k_n'. This ends the proof for this part of the theorem.

Part 2. By Part 2 of Corollary 6.1, the condition $P_n\left[\acute{\Delta}_n \subsetneq \hat{\Delta}_n\right] > 0$ for large n, implies that $P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] > P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right]$ holds for large n. In conjunction with this implication, the condition $P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] \searrow p_\infty \in (0,1)$ as $n \to +\infty$, further implies $P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right] \searrow p_\infty$ as $n \to +\infty$. In consequence,

$$P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] > P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right] > p_\infty \quad \text{for large} \quad n. \tag{D.59}$$

From the convergence of $P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right]$ to p_∞ as $n \to +\infty$, by setting

$$\gamma_n = P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] - p_\infty,$$

$\exists N_{\gamma_n} \in \mathbb{Z}_+$ such that $P_m\left[\hat{\Upsilon}_\infty \leq \alpha\right] - p_\infty < \gamma_n \quad \forall m \geq N_{\gamma_n}$. Substituting out γ_n yields

$$P_m\left[\hat{\Upsilon}_\infty \leq \alpha\right] < P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] \quad \forall m \geq N_{\gamma_n}. \tag{D.60}$$

Hence,

$$\left\{N \in \mathbb{Z}_+ : P_m\left[\hat{\Upsilon}_\infty \leq \alpha\right] \geq P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right] \; \forall m \geq N\right\} = \emptyset, \tag{D.61}$$

which implies that k_n does not exist. This concludes the proof.

D.8. Corollary 6.2

PROOF. **Part 1**. The proof proceeds by the direct method. Under the conditions of the corollary, when n is large enough, Part 1 of Theorem 6.3 establishes the existence of k_n, and therefore, $P_{k_n}\left[\hat{\Upsilon}_\infty \leq \alpha\right]$ exists. The strict inequality, $k_n > n$, holds, since $P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right] < P_n\left[\acute{\Upsilon}_\infty \leq \alpha\right]$; hence, $k_n/n > 1$ and $d_n > 0$, when n is large enough. In consequence, $\liminf_{n\to+\infty}(k_n/n) \geq 1$ and $\liminf_{n\to+\infty} d_n \geq 0$.

Part 2. The proof proceeds by the direct method. The proof makes use of the following linear function of $m \in \mathbb{R}_+$ that passes through the points $\left(n, P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right]\right)$ and $\left(k_n, P_{k_n}\left[\hat{\Upsilon}_\infty \leq \alpha\right]\right)$:

$$Y = b_n + \left(\frac{P_{k_n}\left[\hat{\Upsilon}_\infty \leq \alpha\right] - P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right]}{d_n}\right) m, \tag{D.62}$$

where at $m = n$, $Y = P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right]$, and at $m = k_n$, $Y = P_{k_n}\left[\hat{\Upsilon}_\infty \leq \alpha\right]$. These two points define the intercept b_n as either

$$b_n = P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right] - \left(\frac{P_{k_n}\left[\hat{\Upsilon}_\infty \leq \alpha\right] - P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right]}{d_n}\right) n, \quad \text{or} \tag{D.63}$$

$$b_n = P_{k_n}\left[\hat{\Upsilon}_\infty \leq \alpha\right] - \left(\frac{P_{k_n}\left[\hat{\Upsilon}_\infty \leq \alpha\right] - P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right]}{d_n}\right) k_n. \tag{D.64}$$

Solving for d_n in equation (D.63) yields

$$d_n = \left(\frac{P_{k_n}\left[\hat{\Upsilon}_\infty \leq \alpha\right] - P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right]}{P_n\left[\hat{\Upsilon}_\infty \leq \alpha\right] - b_n}\right) n. \tag{D.65}$$

Then, the result follows upon dividing both sides by q_n and taking limits with respect to n. This concludes the proof for this part of the corollary.

E. Intermediate Technical Results for Theorem 4.1

This section presents intermediate technical results that are used in the proofs of Theorem 4.1.

E.1. Consequences of Definition 4.1

Define the events

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

$$I_n^+ = \left\{i \in \{1, \ldots, n\} : g\left(\mathbf{X}_i; t\right) \geq 0 \;\forall t \in [\underline{t}, \overline{t}]\right\} \quad \text{and} \tag{E.2}$$

$$I_n = \left\{(i, K) \in \{1, \ldots, n\} \times \{A, B\} : X_i^K \in [\underline{t}, \overline{t}]\right\}. \tag{E.3}$$

The results of this subsection concern the large-sample behavior of the likelihoods of the events $\{I_n^- \neq \emptyset\}, \{I_n^+ \neq \emptyset\}$ and $\{I_n \neq \emptyset\}$ with uniformity over probabilities in $\mathcal{P}_1(c_1)$ and $\mathcal{P}_2(c_2)$. Furthermore, we will show that probabilities in $\mathcal{P}_2(c_2)$ have marginal densities bounded from below by c_2 over the interval $[\underline{t}, \overline{t}]$.

We first focus on the event $\{I_n^- \neq \emptyset\}$.

LEMMA E.1. *Given* $c_1 \in (0,1)$ *and recall that* $\mathcal{P}_1(c_1)$ *is defined in (4.3). Then*

$$\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)} P\left[I_n^- \neq \emptyset\right] = 1.$$

PROOF. The proof proceeds by the direct method. We observe that

$$\inf_{P\in\mathcal{P}_1(c_1)} P\left[I_n^- \neq \emptyset\right] = 1 - \sup_{P\in\mathcal{P}_1(c_1)} P\left[I_n^- = \emptyset\right]. \tag{E.4}$$

We show that the probability of the complement of $\{I_n^- \neq \emptyset\}$ converges to zero uniformly in $\mathcal{P}_1(c_1)$. The 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

$$\begin{aligned}
\sup_{P\in\mathcal{P}_1(c_1)} P\left[I_n^- = \emptyset\right] &= \sup_{P\in\mathcal{P}_1(c_1)} \left(P\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X}_1;t\right) \geq 0\right]\right)^n & \text{(E.5)}\\
&= \sup_{P\in\mathcal{P}_1(c_1)} \left(1 - P\left[\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X}_1;t\right) < 0\right]\right)^n & \text{(E.6)}\\
&\leq (1-c_1)^n \to 0 & \text{(E.7)}
\end{aligned}$$

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

Next, we focus on the event $\{I_n \neq \emptyset\}$.

LEMMA E.2. *Given* $c_2 \in \left(0, \frac{1}{(\overline{t}-\underline{t})^2}\right)$ *and recall that* $\mathcal{P}_2(c_2)$ *is defined in (4.4). Then*

$$\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_2(c_2)} P\left[I_n \neq \emptyset\right] = 1.$$

PROOF. The proof proceeds by the direct method. We will make use of the joint density $f(x^A, x^B)$ of probability $P \in \mathcal{P}_2(c_2)$ in the proof. We observe that

$$\inf_{P\in\mathcal{P}_2(c_2)} P\left[I_n \neq \emptyset\right] = 1 - \sup_{P\in\mathcal{P}_2(c_2)} P\left[I_n = \emptyset\right]. \tag{E.8}$$

We show that the probability of the complement of $\{I^- \neq \emptyset\}$ converges to zero uniformly in $\mathcal{P}_2(c_2)$. The complement of this event is

$$\{I_n = \emptyset\} = \left\{\forall (i,K) \in \{1,\ldots,n\} \times \{A,B\} : X_i^K \notin [\underline{t},\overline{t}]\right\}.$$

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

$$\sup_{P\in\mathcal{P}_2(c_2)} P\left[I_n=\emptyset\right] = \sup_{P\in\mathcal{P}_2(c_2)} \left(P\left[X_i^A, X_i^B \notin [\underline{t},\overline{t}]\right)\right)^n \tag{E.9}$$

$$= \sup_{P\in\mathcal{P}_2(c_2)} \left(1-P\left[\bigcap_{K=A,B}\left\{X_i^K\in[\underline{t},\overline{t}]\right\}\right]\right)^n \tag{E.10}$$

$$= \sup_{P\in\mathcal{P}_2(c_2)} \left(1-\int_{[\underline{t},\overline{t}]\times[\underline{t},\overline{t}]} f(x^A,x^B)\,dx^A dx^B\right)^n \tag{E.11}$$

$$\leq \left(1-c_2(\overline{t}-\underline{t})^2\right)^n \to 0 \tag{E.12}$$

as $n\to+\infty$, since $c_2\in\left(0,\frac{1}{(\overline{t}-\underline{t})^2}\right)$.

Now we focus on the event $\{I_n^+\neq\emptyset\}$.

Lemma E.3. *Given* $c_2\in\left(0,\frac{1}{(\overline{t}-\underline{t})^2}\right)$ *and recall that* $\mathcal{P}_2(c_2)$ *is defined in (4.4). Then*

$$\lim_{n\to+\infty}\inf_{P\in\mathcal{P}_2(c_2)} P\left[I_n^+\neq\emptyset\right]=1.$$

Proof. The proof proceeds by the direct method. Given $[\underline{t},\overline{t}]$ and $s\in\mathbb{Z}_+$, observe that given by

$$\left\{\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right)<0\right\} = \begin{cases}\left\{X^A<\underline{t}, X^B>\overline{t}\right\}, & \text{if } s=1,\\ \left\{X^A<\underline{t}, X^B>\overline{t}\right\}\cup\left\{X^A<\underline{t}, X^A<X^B\leq\overline{t}\right\}, & \text{if } s\geq 2.\end{cases} \tag{E.13}$$

This representation of the event $\left\{\sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X};t\right)<0\right\}$ is useful for proving the result of the lemma since it implies the following event inclusion

$$\left\{\mathbf{X}_i\in[\underline{t},\overline{t}]\times[\underline{t},\overline{t}]\right\}\subseteq\left\{I_n^+\neq\emptyset\right\}. \tag{E.14}$$

We observe that by Property 1 of the moment functions

$$P\left[I_n^+\neq\emptyset\right] = 1-P\left[I_n^+=\emptyset\right] \tag{E.15}$$

$$= 1-P\left[I_n^-\neq\emptyset\right] \tag{E.16}$$

$$= 1-\Pi_{i=1}^n\left(1-P\left[g\left(\mathbf{X}_i;t\right)\geq 0\;\forall t\in[\underline{t},\overline{t}]\right]\right) \tag{E.17}$$

$$\geq 1-\Pi_{i=1}^n\left(1-P\left[X_i\in[\underline{t},\overline{t}]\times[\underline{t},\overline{t}]\right]\right) \tag{E.18}$$

$$\geq 1-\Pi_{i=1}^n\left(1-c_2\left(\overline{t}-\underline{t}\right)^2\right) \tag{E.19}$$

$$= 1-\left(1-c_2\left(\overline{t}-\underline{t}\right)^2\right)^n \to 1 \tag{E.20}$$

as $n \to +\infty$. Since the lower bound $1-\left(1-c_2\left(\bar{t}-\underline{t}\right)^2\right)^n$ does not depend on $P \in \mathcal{P}_2(c_2)$, the above manipulation implies that

$$\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_2(c_2)} P\left[I_n^+ \neq \emptyset\right] \geq 1, \tag{E.21}$$

which is the desired result.

The last result concern a lower bound on the marginal densities of probabilities in $\mathcal{P}_2(c_2)$.

LEMMA E.4. *Given $c_2 \in \left(0, \frac{1}{(\bar{t}-\underline{t})^2}\right]$. Then marginal densities corresponding to probabilities $P \in \mathcal{P}_2(c_2)$ are bounded from below by $c_2(\bar{t}-\underline{t})$ on the interval $[\underline{t},\bar{t}]$.*

PROOF. The proof proceeds by the direct method. Let $f(x^A, x^B)$ denote a joint density corresponding to some $P \in \mathcal{P}_2(c_2)$. Then the marginal distribution for B is defined as

$$f_B(x^B) = \int_{-\infty}^{\infty} f(x^A, x^B)\, dx^A \tag{E.22}$$

with a similar definition for the marginal density of A. Then

$$\min_{x^B\in[\underline{t},\bar{t}]} f_B(x^B) = \min_{x^B\in[\underline{t},\bar{t}]} \int_{-\infty}^{\infty} f(x^A, x^B)\, dx^A \tag{E.23}$$

$$\geq \min_{(x^A,x^B)\in[\underline{t},\bar{t}]\times[\underline{t},\bar{t}]} \int_{\underline{t}}^{\bar{t}} f(x^A, x^B)\, dx^A + \min_{x^B\in[\underline{t},\bar{t}]} \int_{\mathbb{R}-[\underline{t},\bar{t}]} f(x^A, x^B)\, dx^A \tag{E.24}$$

$$\geq c_2(\bar{t}-\underline{t}). \tag{E.25}$$

since $\min_{x^B\in[\underline{t},\bar{t}]} \int_{\mathbb{R}-[\underline{t},\bar{t}]} f(x^A, x^B)\, dx^A \geq 0$ and $P \in \mathcal{P}_2(c_2)$. An identical argument applies to the marginal density of A. We omit it for brevity.

E.2. Existence and Uniqueness of $\tilde{\mathbf{p}}$ and $\acute{\mathbf{p}}$

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 denote the interior of this set by $\mathcal{H}_n^\circ$. Additionally, recall the feasible sets

$$\mathcal{H}^0(\mathbf{X}) = \left\{\mathbf{p}\in\mathcal{H}_n : \sum_{i=1}^n p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t\in[\underline{t},\bar{t}]\right\} \quad \text{and} \tag{E.26}$$

$$\mathcal{H}_n^0(\mathbf{X}) = \left\{\mathbf{p}\in\mathcal{H}_n : \sum_{i=1}^n p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t\in\mathcal{T}_n\right\}. \tag{E.27}$$

LEMMA E.5. *(a) On the event $\left\{\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right\}$, the random set*

$$\arg\max\left\{\sum_{i=1}^n \log\left(p_i\right); p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t\in[\underline{t},\bar{t}]\right\}$$

is nonempty and a singleton.

(b) On the event $\left\{\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right\}$, *the random set*

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

is nonempty and a singleton.

Proof. **Part 1**. The proof proceeds by verifying the conditions of Weierstrass' Theorem. The objective function is strictly concave in the probabilities. The constraint set, $\mathcal{H}^0(\mathbf{X})$, is certainly bounded. It is the infinite intersection of closed half-planes (which are convex), and since convexity and closedness are preserved under an arbitrary number of intersections, it is closed and convex. Thus, we are done whenever $\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset$.

Part 2. The proof follows identical steps as in part 1 of this lemma, except that we replace $\mathcal{H}^0(\mathbf{X})$ with $\mathcal{H}_n^0(\mathbf{X})$ and observe that the latter set is defined by a finite intersection of closed half-planes (which are convex), and since convexity and closedness are preserved under an arbitrary number of intersections, it is closed and convex.

The next couple of results indicate that for large enough n the constraint sets are non-empty with probability approaching unity, with uniformity. This event is shown to occur by constructing a strictly positive probability vector that satisfies the inequality constraints.

Lemma E.6. *Given* $c_1 \in (0,1)$ *and recall that* $\mathcal{P}_1(c_1)$ *is defined in (4.3). Then*

$$\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)} P\left[\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right] = 1.$$

Proof. The proof proceeds by the direct method.
For large n and uniformly in $\mathcal{P}_1(c_1)$, Lemma E.1 implies that the event

$$\exists i \in \{1, 2, \ldots, n\} \; g\left(\mathbf{X}_i, t\right) < 0 \quad \forall t \in [\underline{t}, \overline{t}] \tag{E.28}$$

occurs with probability approaching 1. Therefore, without loss of generality, suppose that this event holds only for $i = 1$. Then it follows that

$$\forall i \in \{2, 3, \ldots, n\}: \sup_{t\in[\underline{t},\overline{t}]} g\left(\mathbf{X}_i, t\right) \geq 0. \tag{E.29}$$

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

$$\sum_{i=1}^n p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t \in [\underline{t}, \overline{t}] \iff 1 > p_1 \geq \sum_{i=2}^n p_i \left(\frac{g\left(\mathbf{X}_i; t\right)}{-g\left(\mathbf{X}_1; t\right)}\right) \quad \forall t \in [\underline{t}, \overline{t}] \tag{E.30}$$

Therefore, a sufficient condition for the inequalities in (E.30) is

$$\sup_{t\in[\underline{t},\overline{t}]}\sum_{i=2}^{n}p_i\left(\frac{g\left(\mathbf{X}_i;t\right)}{-g\left(\mathbf{X}_1;t\right)}\right)\leq\max_{2\leq i\leq n}\left(\sup_{t\in[\underline{t},\overline{t}]}\frac{g\left(\mathbf{X}_i;t\right)}{-g\left(\mathbf{X}_1;t\right)}\right)\sum_{i=2}^{n}p_i \tag{E.31}$$

$$\leq\max_{2\leq i\leq n}\left(\frac{\sup_{t\in[\underline{t},\overline{t}]}g\left(\mathbf{X}_i;t\right)}{\inf_{t\in[\underline{t},\overline{t}]}-g\left(\mathbf{X}_1;t\right)}\right)\sum_{i=2}^{n}p_i<1. \tag{E.32}$$

It should be noted that $\inf_{t\in[\underline{t},\overline{t}]}-g\left(\mathbf{X}_1;t\right)>0$ follows directly from (E.28). On the event

$$\max_{2\leq i\leq n}\left(\frac{\sup_{t\in[\underline{t},\overline{t}]}g\left(\mathbf{X}_i;t\right)}{\inf_{t\in[\underline{t},\overline{t}]}-g\left(\mathbf{X}_1;t\right)}\right)\in[0,1],$$

any positive probability vector satisfies the inequalities (E.30). Otherwise, on the event

$$\max_{2\leq i\leq n}\left(\frac{\sup_{t\in[\underline{t},\overline{t}]}g\left(\mathbf{X}_i;t\right)}{\inf_{t\in[\underline{t},\overline{t}]}-g\left(\mathbf{X}_1;t\right)}\right)>1,$$

the inequality (E.31) is equivalent to

$$\sum_{i=2}^{n}p_i<\frac{1}{\max_{2\leq i\leq n}\left(\frac{\sup_{t\in[\underline{t},\overline{t}]}g(\mathbf{X}_i;t)}{\inf_{t\in[\underline{t},\overline{t}]}-g(\mathbf{X}_1;t)}\right)}\iff 1-\frac{1}{\max_{2\leq i\leq n}\left(\frac{\sup_{t\in[\underline{t},\overline{t}]}g(\mathbf{X}_i;t)}{\inf_{t\in[\underline{t},\overline{t}]}-g(\mathbf{X}_1;t)}\right)}<p_1. \tag{E.33}$$

Thus, for any p_1 such that

$$1-\left(\max_{2\leq i\leq n}\left(\frac{\sup_{t\in[\underline{t},\overline{t}]}g\left(\mathbf{X}_i;t\right)}{\inf_{t\in[\underline{t},\overline{t}]}-g\left(\mathbf{X}_1;t\right)}\right)\right)^{-1}<p_1<1,$$

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.

Since $\mathcal{H}^0(\mathbf{X})\subseteq\mathcal{H}_n^0(\mathbf{X})$ holds for each n, we have the following result.

COROLLARY E.1. *Given $c_1\in(0,1)$ and recall that $\mathcal{P}_1(c_1)$ is defined in (4.3). Then*

$$\lim_{n\to+\infty}\inf_{P\in\mathcal{P}_1(c_1)}P\left[\mathcal{H}_n^0(\mathbf{X})\cap\mathcal{H}_n^\circ\neq\emptyset\right]=1.$$

PROOF. We observe that $P\left[\mathcal{H}_n^0(\mathbf{X})\cap\mathcal{H}_n^\circ\neq\emptyset\right]\geq P\left[\mathcal{H}^0(\mathbf{X})\cap\mathcal{H}_n^\circ\neq\emptyset\right]$ holds for all n, which implies that

$$\inf_{P\in\mathcal{P}_1(c_1)}P\left[\mathcal{H}_n^0(\mathbf{X})\cap\mathcal{H}_n^\circ\neq\emptyset\right]\geq\inf_{P\in\mathcal{P}_1(c_1)}P\left[\mathcal{H}^0(\mathbf{X})\cap\mathcal{H}_n^\circ\neq\emptyset\right] \tag{E.34}$$

holds for each n. Finally, taking limits as $n\to+\infty$ on both sides of the above inequality and applying Lemma E.6 to the right side of the above inequality implies the desired result.

E.3. Convergence of $\{\mathcal{T}_n\}_{n\geq 1}$: A Maximal Spacings Approach

Recall that $\{\mathcal{T}_n\}_{n\geq 1}$ comprises the order statistics from the set comprises the order statistics from the sample

$$\left\{X_i^A, X_i^B\right\}_{i=1}^n \bigcap (\underline{t}, \bar{t}) \bigcup \left\{\underline{t}, \bar{t}\right\}, \tag{E.35}$$

where $t_{(0)} = \underline{t}$ and $t_{(N)} = \bar{t}$. Let $d_H\left[\mathcal{T}_n, [\underline{t}, \bar{t}]\right]$ denote the Hausdorff distance between the grid $\mathcal{T}_n$ and the interval $[\underline{t}, \bar{t}]$. This section presents a convergence result for this distance that holds uniformly over sets of probabilities of the form (4.4).

We have the following result

LEMMA E.7. *For every $\epsilon > 0$ and $c_2 \in \left(0, \frac{1}{(\bar{t}-\underline{t})^2}\right)$,*

$$\lim_{n\to+\infty} \sup_{P\in\mathcal{P}_2(c_2)} P\left[d_H\left[\mathcal{T}_n, [\underline{t}, \bar{t}]\right] > \epsilon\right] = 0, \quad \textit{and} \tag{E.36}$$

$$d_H\left[\mathcal{T}_n, [\underline{t}, \bar{t}]\right] = O_P\left(\frac{\log n}{n}\right) \quad \textit{uniformly in} \quad P \in \mathcal{P}_2(c_2), \tag{E.37}$$

where $\mathcal{P}_2(c_2)$ is defined in (4.4).

PROOF. The proof proceeds using the direct method. For large n with probability approaching unity, Lemma E.2 implies that the grid $\mathcal{T}_n = \left\{t_{(j)}\right\}_{j=0}^N$ will contain at least one element from the bivariate random sample. Observe that

$$d_H\left[\mathcal{T}_n, [\underline{t}, \bar{t}]\right] \leq \max_{j=1,\ldots,N}\left(t_{(j+1)} - t_{(j)}\right) \leq \max_{K=A,B}\max_{j=0,\ldots,N^K}\left(X_{(j+1)}^K - X_{(j)}^K\right) \tag{E.38}$$

$$\leq \sum_{K=A,B}\max_{j=0,\ldots,N^K}\left(X_{(j+1)}^K - X_{(j)}^K\right), \tag{E.39}$$

where $X_{(0)}^K = \underline{t}, X_{(N^K+1)}^K = \bar{t}$ and $\left\{X_{(j)}^K\right\}_{j=1}^{N^K}$ are the elements of $\left\{t_{(j)}\right\}_{j=0}^N$ that belong to population K for $K = A, B$.

Next, express the spacings $X_{(j+1)}^K - X_{(j)}^K$ in terms of spacings from Uniform$(0,1)$ random variables using the CDF $F_K(\cdot)$. Noting that $U_{(j)}^K = F_K\left(X_{(j)}^K\right)$ for $j = 0, \ldots, N^K$, the spacings can be expressed as

$$X_{(j+1)}^K - X_{(j)}^K = \frac{F_K^{-1}\left(U_{(j+1)}^K\right) - F_K^{-1}\left(U_{(j)}^K\right)}{U_{(j+1)}^K - U_{(j)}}\left(U_{(j+1)}^K - U_{(j)}\right) \tag{E.40}$$

$$= \frac{U_{(j+1)}^K - U_{(j)}^K}{f_K\left(F_K^{-1}\left(aU_{(j+1)}^K + (1-a)U_{(j)}^K\right)\right)} \tag{E.41}$$

where $a \in [0,1]$ and the second equality follows from an application of the Mean Value Theorem to the marginal quantile function $F_K^{-1}(\cdot)$, and $f_K(\cdot)$ is the marginal

PDF. For $P \in \mathcal{P}_2(c_2)$ Lemma E.4 shows that the marginal densities on the interval $[\underline{t}, \overline{t}]$ are bounded from below by $c_2(\overline{t} - \underline{t})$. Because $\left\{X^K_{(j)}\right\}_{j=1}^{N^K}$ are the elements of $\left\{t_{(j)}\right\}_{j=0}^{N}$, they can only take values in the interval $[\underline{t}, \overline{t}]$. Hence,

$$\frac{U^K_{(j+1)} - U^K_{(j)}}{f_K\left(F_K^{-1}\left(aU^K_{(j+1)} + (1-a)U^K_{(j)}\right)\right)} \leq \frac{U^K_{(j+1)} - U^K_{(j)}}{c_2(\overline{t} - \underline{t})} \tag{E.42}$$

So that the spacing $X^K_{(j+1)} - X^K_{(j)}$ is bounded from above by $\frac{U^K_{(j+1)} - U^K_{(j)}}{c_2(\overline{t} - \underline{t})}$.

In consequence,

$$\max_{j=0,\ldots,N^K}\left(X^K_{(j+1)} - X^K_{(j)}\right) \leq \left(c_2(\overline{t} - \underline{t})\right)^{-1} \max_{j=0,\ldots,N^K}\left(U^K_{(j+1)} - U^K_{(j)}\right) \quad K = A, B, \tag{E.43}$$

$$\leq \left(c_2(\overline{t} - \underline{t})\right)^{-1} \max_{i=1,\ldots,n}\left(U^K_{(i+1)} - U^K_{(i)}\right) \quad K = A, B. \tag{E.44}$$

where $U^K_{(i)} = F_K\left(X^K_{(i)}\right)$ for $i = 1, \ldots, n$. That is, the maximal spacing

$$\max_{j=0,\ldots,N^K}\left(X^K_{(j+1)} - X^K_{(j)}\right)$$

are bounded above by a constant times the maximal spacing of a random sample from Uniform$(0, 1)$. Combining these results yields

$$d_H\left[\mathcal{T}_n, [\underline{t}, \overline{t}]\right] \leq \left(c_2(\overline{t} - \underline{t})\right)^{-1} \sum_{K=A,B} \max_{i=1,\ldots,n}\left(U^K_{(i+1)} - U^K_{(i)}\right), \tag{E.45}$$

where the right side of this inequality depends on the the underlying probability $P \in \mathcal{P}_2(c_2)$ only through the dependence structure between the marginal distributions. Now we can apply the Theorem from Devroye (1982) on maximal uniform spacings to deduce that

$$P\left[\max_{i=1,\ldots,n}\left(U^K_{(i+1)} - U^K_{(i)}\right) \leq \frac{\log n - \log\log\log n - \log 2}{n} \quad \text{infinitely often}\right] = 1 \tag{E.46}$$

for each $K = A, B$. This result and the manipulation above imply that

$$\sup_{P\in\mathcal{P}_2(c_2)} P\left[d_H\left[\mathcal{T}_n,[\underline{t},\overline{t}]\right] > \epsilon\right] \le \sup_{P\in\mathcal{P}_2(c_2)} P\left[\sum_{K=A,B} \frac{\max_{i=1,\ldots,n}\left(U^K_{(i+1)} - U^K_{(i)}\right)}{c_2(\overline{t}-\underline{t})} > \epsilon\right] \tag{E.47}$$

$$\le \epsilon^{-1}\sum_{K=A,B} E\left[\frac{\max_{i=1,\ldots,n}\left(U^K_{(i+1)} - U^K_{(i)}\right)}{c_2(\overline{t}-\underline{t})}\right] \tag{E.48}$$

$$\le \frac{2}{c_2(\overline{t}-\underline{t})\epsilon}\left(\frac{\log n - \log\log\log n - \log 2}{n}\right), \tag{E.49}$$

where we used Markov's inequality to obtain (E.48) with expectation taken with respect to the joint distribution of a random sample of size n from Uniform$(0,1)$, which does not depend on $K = A, B$. Finally, taking limits on both sides of (E.49) as $n \to +\infty$ yields the desired result. Furthermore, (E.45) and (E.46) imply

$$d_H\left[\mathcal{T}_n,[\underline{t},\overline{t}]\right] = O_P\left(\frac{\log n}{n}\right) \quad \text{uniformly in} \quad P \in \mathcal{P}_2(c_2).$$

E.4. Upper Bound On Lower Level Problem: $s > 1$

Recall that $\acute{\mathbf{p}}$ is the solution of discretized constrained empirical likelihood problem. For $s > 1$, and for each $c_1 \in (0,1)$ and $c_2 \in \left(0, \frac{1}{(\overline{t}-\underline{t})^2}\right)$, this section presents an upper bound for the value

$$\max_{t\in[\underline{t},\overline{t}]} \sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i; t\right) \tag{E.50}$$

that holds with probability approaching unity uniformly over probabilities $P \in \mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2)$. The bound is given by

$$q_n = \left(\frac{4q'_n}{c_2(\overline{t}-\underline{t})\left[(s-1)!\right]}\right)\left(\frac{\log n}{n}\right), \quad \text{where} \tag{E.51}$$

$$q'_n = \max\left\{\sum_{\ell=1}^{s-1}\binom{s-1}{\ell}\left[\sum_{i=1}^{n}\acute{p}_i\left(\overline{t} - X^A_i\right)^{s-1-\ell} 1\left[X^A_i \le \overline{t}\right]\right], 1\right\}. \tag{E.52}$$

The result is the following.

LEMMA E.8. *Let* $s > 1$, *and for each* $c_1 \in (0,1)$ *and* $c_2 \in \left(0, \frac{1}{(\overline{t}-\underline{t})^2}\right)$

$$\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2)} P\left[\max_{t\in[\underline{t},\overline{t}]}\sum_{i=1}^{n}\acute{p}_i g\left(\mathbf{X}_i; t\right) \le q_n\right] = 1. \tag{E.53}$$

Proof. The proof proceeds by the direct method. We will first show

$$\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2)} P\left[\max_{t\in[\underline{t},\overline{t}]}\sum_{i=1}^{n}\acute{p}_i g\left(\mathbf{X}_i;t\right)\le\frac{2q_n'\max_{j=0,\ldots,N}\left(t_{(j+1)}-t_{(j)}\right)}{(s-1)!}\right]=1, \tag{E.54}$$

where q_n' is defined in (E.52). Then, using steps identical to those in the proof of Lemma E.7, we can (i) bound the maximal spacing $\max_{j=0,\ldots,N}\left(t_{(j+1)}-t_{(j)}\right)$ from above by

$$\frac{\sum_{K=A,B}\max_{i=1,\ldots,n}\left(U_{(i+1)}^{K}-U_{(i)}^{K}\right)}{c_2(\overline{t}-\underline{t})},$$

where $U_{(j)}^{K}=F_K\left(X_{(j)}^{K}\right)\ \forall j$ and $K=A,B$, and (ii) apply the Theorem in Devroye (1982) to this bound to obtain the desired result.

On the event $\left\{\mathcal{H}_n^0(\mathbf{X})\cap\mathcal{H}_n^\circ\neq\emptyset\right\}$, Part 1 of Lemma E.5 shows that the extremum $\acute{\mathbf{p}}$ exists and is unique. Given $t\in[\underline{t},\overline{t}]-\mathcal{T}_n$, $\exists j$ such that $t_{(j)}<t<t_{(j+1)}$. Then, for such a t

$$g\left(\mathbf{X}_i;t\right)\le\frac{\left(t_{(j+1)}-X_i^B\right)^{s-1}1\left[X_i^B\le t_{(j+1)}\right]}{(s-1)!}-\frac{\left(t_{(j)}-X_i^A\right)^{s-1}1\left[X_i^A\le t_{(j)}\right]}{(s-1)!} \tag{E.55}$$

$$=g\left(\mathbf{X}_i;t_{(j+1)}\right)+\frac{\left(t_{(j+1)}-X_i^A\right)^{s-1}1\left[X_i^B\le t_{(j+1)}\right]}{(s-1)!}-\frac{\left(t_{(j)}-X_i^A\right)^{s-1}1\left[X_i^A\le t_{(j)}\right]}{(s-1)!} \tag{E.56}$$

$$\le\left\{\sum_{\ell=1}^{s-1}\binom{s-1}{\ell}\left(\overline{t}-X_i^A\right)^{s-1-\ell}1\left[X_i^A\le\overline{t}\right]+1\left[t_{(j)}<X_i^A\le t_{(j+1)}\right]\right\}\left(t_{(j+1)}-t_{(j)}\right)$$
$$+\,g\left(\mathbf{X}_i;t_{(j+1)}\right). \tag{E.57}$$

where we made use of the Binomial Theorem to expand $\left(t_{(j+1)}-X_i^A\right)^{s-1}$ and $\left(t_{(j)}-X_i^A\right)^{s-1}$ in (E.56), and that $\left(t_{(j+1)}-t_{(j)}\right)^{s-1}\le\left(t_{(j+1)}-t_{(j)}\right)$ holds with n large, and with uniformity, because the maximal spacings tend to zero in probability. Therefore, multiplying by $\acute{p}_i$ and summing over i, we obtain the following bound on $\sum_{i=1}^{n}\acute{p}_i g\left(\mathbf{X}_i;t\right)$ that holds $\forall t\in[\underline{t},\overline{t}]-\mathcal{T}_n$:

$$\sum_{i=1}^{n}\acute{p}_i g\left(\mathbf{X}_i;t\right)\le\sum_{i=1}^{n}\acute{p}_i g\left(\mathbf{X}_i;t_{(j+1)}\right)+\frac{t_{(j+1)}-t_{(j)}}{(s-1)!}\sum_{i=1}^{n}\acute{p}_i 1\left[t_{(j)}<X_i^A\le t_{(j+1)}\right]$$
$$+\frac{t_{(j+1)}-t_{(j)}}{(s-1)!}\left\{\sum_{\ell=1}^{s-1}\binom{s-1}{\ell}\sum_{i=1}^{n}\acute{p}_i\left(\overline{t}-X_i^A\right)^{s-1-\ell}1\left[X_i^A\le\overline{t}\right]\right\} \tag{E.58}$$

$$\le\frac{2q_n'}{(s-1)!}\max_{j=0,\ldots,N}\left(t_{(j+1)}-t_{(j)}\right), \tag{E.59}$$

because $\max_{j=0,\ldots,N} \sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i; t_{(j)}\right) \leq 0$ and $\sum_{i=1}^{n} \acute{p}_i 1\left[t_{(j)} < X_i^A \leq t_{(j+1)}\right] \leq 1$. So what we have shown is the event inclusion

$$\left\{\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right\} \subseteq \left\{\frac{2q_n'}{(s-1)!} \max_{j=0,\ldots,N} \left(t_{(j+1)} - t_{(j)}\right)\right\}, \tag{E.60}$$

and Corollary E.1 implies the limit (E.54), because

$$\begin{aligned} 1 &= \lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)} P\left[\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right] \\ &\leq \lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2)} P\left[\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right]. \end{aligned}$$

Now by using steps identical to those in the proof of Lemma E.7 we have

$$\max_{j=0,\ldots,N} \left(t_{(j+1)} - t_{(j)}\right) \leq \left(c_2(\bar{t} - \underline{t})\right)^{-1} \sum_{K=A,B} \max_{i=1,\ldots,n} \left(U_{(i+1)}^K - U_{(i)}^K\right). \tag{E.61}$$

And an application of the Theorem of Devroye (1982) to the maximal uniform spacings

$$\max_{i=1,\ldots,n} \left(U_{(i+1)}^K - U_{(i)}^K\right) \quad K = A, B,$$

yields

$$P\left[\sum_{K=A,B} \max_{i=1,\ldots,n} \left(U_{(i+1)}^K - U_{(i)}^K\right) \leq 2\frac{\log n - \log\log\log n - \log 2}{n} \text{ i.o.}\right] = 1.$$

Combining this result with the inequality (E.61) and the event inclusion (E.60) yields the set inclusion

$$\left\{\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right\} \subseteq \left\{\max_{t\in[\underline{t},\bar{t}]} \sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i; t\right) \leq q_n\right\}, \tag{E.62}$$

where q_n is given by (E.51). Finally, an application of Corollary E.1 yields the desired result.

E.5. Construction of $\check{\mathbf{p}}$*:* $s > 1$

This subsection presents a result that establishes the feasibility of a particular construction of a probability vector $\check{\mathbf{p}} \in \mathcal{H}^0(\mathbf{X})$ from $\acute{\mathbf{p}}$. As the vector $\acute{\mathbf{p}}$ is not necessarily a member of $\mathcal{H}^0(\mathbf{X})$, the construction is such that $\check{\mathbf{p}}$ is nearby to $\acute{\mathbf{p}}$ and will be essential in proving Part 2 Theorem 4.1 using Lemma E.10. The construction uses an approach similar to that of Lemma 3 in Still (2001), but is different since we make use of Property 1 of the moment functions.

Following Still (2001), consider the construction $\check{\mathbf{p}} = \acute{\mathbf{p}} + \rho q_n \mathbf{d}$, where $\mathbf{d} \in \mathbb{R}^n$ and $\rho > 0$, and q_n is defined in (E.51) in the previous subsection. Naturally, we need to impose restrictions on the vector $\mathbf{d}$ and ρ that enures $\check{\mathbf{p}} \in \mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^{\circ}$. Let $\underline{\delta}$ and $\overline{\delta}$ be random variables such that $0 < \underline{\delta} < \acute{p}_{(1)}$ and $0 < \overline{\delta} < 1 - \acute{p}_{(n)}$, where $\acute{p}_{(1)} = \min_{i=1,\ldots,n} \acute{p}_i$ and $\acute{p}_{(n)} = \max_{i=1,\ldots,n} \acute{p}_i$. Consider the following set of vectors:

$$D_n\left(\underline{\delta}, \overline{\delta}\right) = \left\{ \mathbf{d} \in \mathbb{R}^n : \sum_{i=1}^n d_i = 0,\, \epsilon_n(\mathbf{d}) > 0,\, \underline{\delta} \leq \check{p}_i \leq \overline{\delta}\, \forall i \right\}. \tag{E.63}$$

where $\epsilon_n(\mathbf{d}) = -\sup_{t \in [\underline{t}, \overline{t}]} \sum_{i=1}^n d_i g\left(\mathbf{X}_i; t\right)$.

The following result has two parts. The first part shows that it is feasible to construct a vector $\check{\mathbf{p}} \in \mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^{\circ}$ for any choice of $\underline{\delta}$ and $\overline{\delta}$ that satisfies the conditions $0 < \underline{\delta} < \acute{p}_{(1)}$ and $0 < \overline{\delta} < 1 - \acute{p}_{(n)}$. The second part shows that with additional conditions on $\underline{\delta}$ and $\overline{\delta}$, we can construct $\check{\mathbf{p}} \in \mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^{\circ}$ which is close to the vector $\acute{\mathbf{p}}$ in a particular way so that we can use it Steps 2 and 4 of the proof for Part 2 of Theorem 4.1.

LEMMA E.9. *Suppose that the event $\{I_n^- \neq \emptyset\}$ occurs, where I_n^- is defined in (E.1). Also let the set D_n be defined as in (E.63). Then the following statements hold.*

(a) For each $\mathbf{d} \in D_n\left(\underline{\delta}, \overline{\delta}\right)$ and $\rho > \frac{1}{\epsilon_n(\mathbf{d})}$, $\check{\mathbf{p}} = \acute{\mathbf{p}} + \rho q_n \mathbf{d} \in \mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^{\circ}$.

(b) Suppose that $\mathbf{d} \in D_n\left(\underline{\delta}, \overline{\delta}\right)$. If $\rho > \frac{1}{\epsilon_n(\mathbf{d})}$, $\underline{\delta} > \acute{p}_{(1)} + \acute{p}_{(n)} + \overline{\delta} - 1$ and $\underline{\delta} = \acute{p}_{(1)}/2$, then

$$\rho q_n \left|d_i\right| \leq \frac{\acute{p}_{(1)} - \underline{\delta}}{n^2}\, \forall i \implies \left\|\rho q_n \mathbf{d}\right\| \leq \frac{\acute{p}_{(1)}}{2n^{3/2}}. \tag{E.64}$$

PROOF. The proof proceeds by the direct method.

Part 1. Firstly note that from the proofs of Lemmas E.5 and E.6 we have the following event inclusions:

$$\left\{I_n^- \neq \emptyset\right\} \subseteq \left\{\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^{\circ} \neq \emptyset\right\} \subseteq \left\{\acute{\mathbf{p}} \text{ exists and is unique}\right\}. \tag{E.65}$$

Suppose that $\mathbf{d} \in D_n$. The condition $\sum_{i=1}^n d_i = 0$ implies that $\sum_{i=1}^n \check{p}_i = 1$. Now we will show that any value of $\rho > \frac{1}{\epsilon_n(\mathbf{d})}$ yields $\sum_{i=1}^n \check{p}_i g\left(\mathbf{X}_i; t\right) < 0 \quad \forall t \in [\underline{t}, \overline{t}]$. Using Property 1 of the moment functions the d_i can be chosen so that they obey the following sign restrictions:

$$\operatorname{sign}\left(d_i\right) = \begin{cases} < 0, & \text{if } g\left(\mathbf{X}_i, t\right) \geq 0\, \forall t \in [\underline{t}, \overline{t}] \\ > 0, & \text{if } g\left(\mathbf{X}_i, t\right) < 0\, \forall t \in [\underline{t}, \overline{t}]. \end{cases} \tag{E.66}$$

These sign restrictions on $\mathbf{d}$ yield

$$\sum_{i=1}^n d_i g\left(\mathbf{X}_i; t\right) < 0 \quad \forall t \in [\underline{t}, \overline{t}]. \tag{E.67}$$

Now since $s > 1$ implies that the moment functions are continuous in the index variable t, the compactness of the interval $[\underline{t}, \overline{t}]$ and the sign restrictions imply that $\epsilon_n(\mathbf{d}) = -\sup_{t\in[\underline{t},\overline{t}]} g(\mathbf{X}_i;t) > 0$. In consequence, $\forall t \in [\underline{t}, \overline{t}]$

$$\sum_{i=1}^{n} \breve{p}_i g(\mathbf{X}_i;t) = \sum_{i=1}^{n} \acute{p}_i g(\mathbf{X}_i;t) + \rho q_n \sum_{i=1}^{n} d_i g(\mathbf{X}_i;t) \tag{E.68}$$

$$\leq q_n - \rho\epsilon_n(\mathbf{d}) q_n, \tag{E.69}$$

where $\sup_{t\in[\underline{t},\overline{t}]} \sum_{i=1}^{n} \acute{p}_i g(\mathbf{X}_i;t) \leq q_n$ follows from Lemma E.8. Thus,

$$q_n - \rho\epsilon_n(\mathbf{d}) q_n \leq 0 \iff \rho \geq \frac{1}{\epsilon_n(\mathbf{d})}. \tag{E.70}$$

Furthermore, observe that

$$\breve{p}_i \geq \underline{\delta}\ \forall i \iff d_i \geq \frac{\underline{\delta} - \acute{p}_i}{\rho q_n}\ \forall i \quad \text{and} \quad \frac{\underline{\delta} - \acute{p}_i}{\rho q_n} < 0\ \forall i \iff \underline{\delta} < \acute{p}_{(1)}. \tag{E.71}$$

$$\breve{p}_i \leq \overline{\delta}\ \forall i \iff d_i \leq \frac{1 - \overline{\delta} - \acute{p}_i}{\rho q_n}\ \forall i \quad \text{and} \quad \frac{1 - \overline{\delta} - \acute{p}_i}{\rho q_n} > 0\ \forall i \iff \overline{\delta} < 1 - \acute{p}_{(n)}. \tag{E.72}$$

Hence, the conditions above with $\underline{\delta}, \overline{\delta} > 0$ yields $\breve{\mathbf{p}} = \acute{\mathbf{p}} + \rho q_n \mathbf{d} \in \mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ$. Since $\mathbf{d} \in D_n(\underline{\delta}, \overline{\delta})$ was arbitrary, the result holds for all $\mathbf{d} \in D_n(\underline{\delta}, \overline{\delta})$, which concludes this part of the proof.

Part 2. The proof proceeds by the direct method. Consider a vector $\mathbf{d} \in D_n(\underline{\delta}, \overline{\delta})$ for which $\rho > \frac{1}{\epsilon_n(\mathbf{d})}$, $\underline{\delta} > \acute{p}_{(1)} + \acute{p}_{(n)} + \overline{\delta} - 1$ and $\underline{\delta} = \acute{p}_{(1)}/2$. If the vector also satisfies $\rho q_n |d_i| \leq \frac{\acute{p}_{(1)} - \underline{\delta}}{n^2}\ \forall i$, then

$$\rho q_n |d_i| \leq \frac{\acute{p}_{(1)} - \underline{\delta}}{n^2} = \frac{\acute{p}_{(1)}}{2n^2}\ \forall i. \tag{E.73}$$

So by squaring both sides (E.73), then summing over i, and applying the square-root yields to both sides the desired result. This concludes the proof.

The result of Lemma E.9 establishes that one can indeed construct a probability vector $\breve{\mathbf{p}}$ defined above whenever $\{I_n^- \neq \emptyset\}$ occurs. This arises from the event inclusions in (E.65). In consequence, the probability of constructing $\breve{\mathbf{p}}$ is determined by the probability of the event $\{I_n^- \neq \emptyset\}$. The next result establishes that the probability of being able to construct $\breve{\mathbf{p}}$ as in Lemma E.9 tends to unity , with uniformity.

Corollary E.2. *Let $s > 1$, and let V_n denote the event: $\forall \mathbf{d} \in D_n(\underline{\delta}, \overline{\delta})$ any value of $\rho > \frac{1}{\epsilon_n(\mathbf{d})}$ yields $\breve{\mathbf{p}} = \acute{\mathbf{p}} + \rho q_n \mathbf{d} \in \mathcal{H}^0(\mathbf{X})$, where $0 < \underline{\delta} < \acute{p}_{(1)}$ and $0 < \overline{\delta} < 1 - \acute{p}_{(n)}$. Then, for each $c_1 \in (0,1)$ and $c_2 \in \left(0, \frac{1}{(\overline{t}-\underline{t})^2}\right)$*

$$\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2)} P[V_n] = 1. \tag{E.74}$$

Proof. Using the event inclusions in (E.65), Lemmas E.1 and E.6 and Corollary E.1

$$
\begin{aligned}
1 = \lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)} P\left[I_n^- \neq \emptyset\right] &\leq \lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2)} P\left[I_n^- \neq \emptyset\right] \\
&\leq \lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2)} P\left[\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right] \\
&\leq \lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2)} P\left[\acute{\mathbf{p}} \text{ exists and is unique}\right] \\
&\leq \lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)\bigcap\mathcal{P}_2(c_2)} P\left[V_n\right],
\end{aligned}
$$

which concludes the proof.

E.6. Quadratic Growth Condition of Objective Function At $\tilde{\mathbf{p}} : s > 1$

This subsection presents a quadratic growth condition on the objective function arising in the SIP problem for the case $s > 1$. The proof of the result in this subsection uses the Karush-Kuhn-Tucker conditions for the minimization formulation of the SIP problem. It should be noted that the objective function in the SIP problem can be reformulated so as to write it as a minimization problem. In this reformulation, the objective function is given by $-\sum_{i=1}^{n}\log(p_i)$.

As the moment functions in this case are continuous in the index variable for each probability vector, we embed the constraints into the functional space $C\left([\underline{t},\overline{t}]\right)$, where $C\left([\underline{t},\overline{t}]\right)$ denotes the space of continuous functions $\gamma : [\underline{t},\overline{t}] \to \mathbb{R}$ with sup-norm. The space $C\left([\underline{t},\overline{t}]\right)$ is a Banach space and its dual is the space of finite signed measures on $\left([\underline{t},\overline{t}], \mathcal{B}\right)$, where $\mathcal{B}$ is the Borel sigma algebra of $[\underline{t},\overline{t}]$, with scalar product of $\mu \in C\left([\underline{t},\overline{t}]\right)^*$ and $\gamma \in C\left([\underline{t},\overline{t}]\right)$ given by the integral

$$\int_{[\underline{t},\overline{t}]} \gamma(t)\, d\mu(t). \tag{E.75}$$

thus, the SIP problem has the following Lagrangian function:

$$\mathcal{L} = -\sum_{i=1}^{n}\log(p_i) + \lambda\left(1 - \sum_{i=1}^{n} p_i\right) - n\int_{[\underline{t},\overline{t}]} \sum_{i=1}^{n} p_i g\left(\mathbf{X}_i; t\right)\, d\mu(t), \tag{E.76}$$

where $\lambda \in \mathbb{R}$ is the multiplier on the equality constraint.

The Karush-Kuhn-Tucker conditions are

$$-\frac{1}{p_i} = \lambda + n\int_{[\underline{t},\overline{t}]} g\left(\mathbf{X}_i; t\right)\, d\mu(t) \quad i = 1, 2, \ldots, n \tag{E.77}$$

$$\sum_{i=1}^{n} p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t \in [\underline{t},\overline{t}], \quad \sum_{i=1}^{n} p_i = 1 \tag{E.78}$$

$$\operatorname{supp}(\mu) \subseteq \Delta\left(\tilde{\mathbf{p}}\right), \tag{E.79}$$

where $\Delta(\tilde{\mathbf{p}}) = \left\{t \in [\underline{t}, \overline{t}] : \sum_{i=1}^{n} \tilde{p}_i g(\mathbf{X}_i; t) = 0\right\}$ and the vector of probabilities $\tilde{\mathbf{p}}$ denotes the solution of the SIP problem. As the equality constraint on the probability vector is linear in that vector, this fact in conjunction with Lemma E.6 implies that the *Strong Slater Condition* of Mordukhovich and Nghia (2013) holds with probability tending unity, with uniformity over sets of probabilities of the form $\mathcal{P}_1(c_1)$. In consequence, there exists a (positive) Borel measure $\mu \in C\left([\underline{t}, \overline{t}]\right)^*$ that solve the Karush-Kuhn-Tucker conditions (E.77) - (E.79) along with $\tilde{\mathbf{p}}$.

We have the following result.

LEMMA E.10. *For each $c_1 \in (0,1)$ recall that $\mathcal{P}_1(c_1)$ is defined in (4.3). Let A_n denote the event that there exists a neighborhood of $\tilde{\mathbf{p}}$, U, and a constant $K_0 > 0$ such that*

$$h(\mathbf{p}) - h(\tilde{\mathbf{p}}) \geq K_0 \|\mathbf{p} - \tilde{\mathbf{p}}\|^2 \quad \forall \mathbf{p} \in U \cap \mathcal{H}^0(\mathbf{X}), \tag{E.80}$$

where $\|\cdot\|$ is the Euclidean norm and $h(\mathbf{p}) = -\sum_{i=1}^{n} \log(p_i)$. Then

$$\lim_{n \to +\infty} \inf_{P \in \mathcal{P}_1(c_1)} P[A_n] = 1. \tag{E.81}$$

PROOF. The proof proceeds by the direct method. Given $c_1 \in (0,1)$, we observe that by Lemma E.6 and Part 1 of Lemma E.5, the extremum $\tilde{\mathbf{p}}$ exists and is unique with probability tending to unity, uniformly over probabilities in $\mathcal{P}_1(c_1)$. Thus, we can apply a second-order Taylor expansion to $h(\mathbf{p})$ in a neighborhood of $\tilde{\mathbf{p}}$. That is,

$$h(\mathbf{p}) - h(\tilde{\mathbf{p}}) = \mathbf{d}^T \nabla h(\tilde{\mathbf{p}}) + \frac{1}{2} \mathbf{d}^T \nabla^2 h(\dot{\mathbf{p}}) \mathbf{d}, \tag{E.82}$$

where $\nabla h(\cdot)$ and $\nabla^2 h(\cdot)$ are the gradient and Hessian of a function h, $\dot{\mathbf{p}} = a\tilde{\mathbf{p}} + (1-a)\mathbf{p}$ with $a \in [0,1]$, and $\mathbf{d}^T = \mathbf{p} - \tilde{\mathbf{p}}$.

Now we will show that $\mathbf{d}^T \nabla h(\tilde{\mathbf{p}}) \geq 0$ using the Karush-Kuhn-Tucker conditions (E.77) - (E.79). Observe that

$$\mathbf{d}^T \nabla h(\tilde{\mathbf{p}}) = \sum_{i=1}^{n} \frac{p - \tilde{p}_i}{\tilde{p}_i} = -\sum_{i=1}^{n} \frac{\tilde{p}_i - p_i}{\tilde{p}_i} = -n \int_{\Delta(\tilde{\mathbf{p}})} \sum_{i=1}^{n} p_i g(\mathbf{X}_i; t) \, d\mu(t). \tag{E.83}$$

For $\mathbf{p} \in \mathcal{H}^0(\mathbf{X})$, we must have $-n \int_{\Delta(\tilde{\mathbf{p}})} \sum_{i=1}^{n} p_i g(\mathbf{X}_i; t) \, d\mu(t) \geq 0$ because the Lagrange multiplier measure is non-negative. Hence, $\forall \mathbf{p} \in \mathcal{H}^0(\mathbf{X})$ that is in a neighborhood of $\tilde{\mathbf{p}}$, it follows that

$$h(\mathbf{p}) - h(\tilde{\mathbf{p}}) \geq \frac{1}{2} \mathbf{d}^T \nabla^2 h(\dot{\mathbf{p}}) \mathbf{d}. \tag{E.84}$$

Now we will construct the neighborhood U and the constant K_0. Let $\epsilon > 0$; we will show that one can consider the neighborhood

$$U = \{\mathbf{p} :\in \mathbb{R}^n : \|\mathbf{p} - \tilde{\mathbf{p}}\| \leq \epsilon\}. \tag{E.85}$$

Observe that $\dot{\mathbf{p}} \in U$ since $\|\dot{\mathbf{p}} - \tilde{\mathbf{p}}\|^2 = (1-a)^2 \|\tilde{\mathbf{p}} - \mathbf{p}\| \leq \epsilon^2$. Furthermore, note that for each $i = 1, \ldots, n$

$$\dot{p}_i^2 = (\dot{p}_i - \tilde{p}_i + \tilde{p}_i)^2 = (\dot{p}_i - \tilde{p}_i)^2 + \tilde{p}_i^2 + 2\tilde{p}_i (\dot{p}_i - \tilde{p}_i) \tag{E.86}$$

$$\leq \epsilon^2 + \tilde{p}_i + 2(1-a)\tilde{p}_i (p_i - \tilde{p}_i) \tag{E.87}$$

$$\leq \epsilon^2 + 5. \tag{E.88}$$

In consequence,

$$\frac{1}{2}\mathbf{d}^T \nabla^2 h(\dot{\mathbf{p}})\, \mathbf{d} = \frac{1}{2}\sum_{i=1}^{n} \frac{d_i^2}{\dot{p}_i^2} \geq \frac{1}{2}\frac{\sum_{i=1}^{n} d_i^2}{\epsilon^2 + 5} = \frac{\|\mathbf{p} - \tilde{\mathbf{p}}\|^2}{2\epsilon^2 + 10}. \tag{E.89}$$

Therefore, for any $\epsilon > 0$, we have that

$$h(\mathbf{p}) - h(\tilde{\mathbf{p}}) \geq \frac{\|\mathbf{p} - \tilde{\mathbf{p}}\|^2}{2\epsilon^2 + 10} \quad \forall \mathbf{p} \in U \cap \mathcal{H}^0(\mathbf{X}), \tag{E.90}$$

which means that we can select $K_0 = 1/(2\epsilon^2 + 10)$. For this reason, we have the event inclusion

$$\left\{\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right\} \subseteq A_n, \tag{E.91}$$

which implies that

$$\begin{aligned} 1 &= \lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)} P\left[\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\right] \\ &\leq \lim_{n\to+\infty} \inf_{P\in\mathcal{P}_1(c_1)} P\left[A_n\right]. \end{aligned}$$

This concludes the proof.

F. Intermediate Technical Results for Theorems 5.1 and 5.2

This section presents intermediate technical results that are used in the proofs of Theorems 5.1 and 5.2. Subsection F.1 presents a technical result that is a consequence of Definition 5.1. Section F.2 uses the results of the previous sections to develop the large-sample properties of the Lagrange multipliers under H_0, under collections of probabilities satisfying Definition 5.1. Finally, Section F.3 establishes the uniform consistency of the proposed empirical likelihood estimator of the moments from the discretized problem, under H_0.

REMARK F.1. *For each $s \in \mathbb{Z}_+$, the moment functions $\left\{\mathbf{x} \mapsto g(\mathbf{x}; t), t \in [\underline{t}, \overline{t}]\right\}$ is suitably measurable and Vapnik-Chervonenkis with envelope function*

$$s \max\left\{|\underline{t}|, |\overline{t}|\right\}^{s-1} \left(|X^B|^{s-1} + |X^A|^{s-1}\right). \tag{F.1}$$

In consequence, Condition (iv) in Assumption 3.1 implies that this set of moment functions are Glivenko-Cantelli and Donsker, both uniformly in $P \in \mathcal{P}$. These properties of the set of moment functions are established by invoking Theorems 2.8.1 and 2.8.2 in van der Vaart and Wellner (1996), and they drive the uniform asymptotic results in the present work.

F.1. A Technical Lemma for Part 4 of Theorem 5.1

LEMMA F.1. *Let* $\Delta\left(\acute{P}_n\right) = \left\{t \in \mathcal{T}_n : E_{\acute{P}_n}\left[g\left(\mathbf{X}; t\right)\right] = 0\right\}$. *For each* $s \in \mathbb{Z}_+$, $c_1 \in (0,1)$, $c_2 \in \left(0, \frac{1}{(\bar{t}-\underline{t})^2}\right)$ *and* $c_3 \in (0, +\infty)$,

$$\lim_{n\to+\infty} \sup_{P \in \mathcal{P}_0^s(c_1,c_2,c_3)\bigcap \mathcal{P}_{000}} P\left[\acute{P}_n = \hat{P}_n\right] \leq \frac{1}{2} \quad \textit{and} \tag{F.2}$$

$$\lim_{n\to+\infty} \inf_{P \in \mathcal{P}_0^s(c_1,c_2,c_3)\bigcap \mathcal{P}_{000}} P\left[\Delta\left(\acute{P}_n\right) \neq \emptyset\right] \geq \frac{1}{2}, \tag{F.3}$$

where $\mathcal{P}_0^s(c_1, c_2, c_3)$ *is given by (5.1) in Definition 5.1 and*

$$\mathcal{P}_{000} = \left\{P \in \mathcal{P}_0 : \Delta(P) \neq \emptyset\right\}.$$

PROOF. The proof proceeds by the direct method. First we prove (F.2). Let $s = 1$ and $\Phi(\cdot)$ denote the CDF of the standard normal distribution. Then for each n

$$\begin{aligned}
P\left[\acute{P}_n = \hat{P}_n\right] &= P\left[E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \leq 0 \;\forall t \in \mathcal{T}_n\right] && \text{(F.4)}\\
&= P\left[E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \leq 0 \;\forall t \in [\underline{t}, \bar{t}]\right] && \text{(F.5)}\\
&\leq P\left[E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] \leq 0 \;\forall t \in \Delta(P)\right] && \text{(F.6)}\\
&\leq P\left[E_{\hat{P}_n}\left[g\left(\mathbf{X}; t'\right)\right] \leq 0 \; t' \in \Delta(P)\right] && \text{(F.7)}\\
&= P\left[\sqrt{n}E_{\hat{P}_n}\left[g\left(\mathbf{X}; t'\right)\right] / E_P\left[g^2\left(\mathbf{X}; t'\right)\right] \leq 0 \; t' \in \Delta(P)\right] && \text{(F.8)}\\
&\leq \left| P\left[\sqrt{n}\frac{E_{\hat{P}_n}\left[g\left(\mathbf{X}; t'\right)\right]}{\sqrt{E_P\left[g^2\left(\mathbf{X}; t'\right)\right]}} \leq 0; t' \in \Delta(P)\right] - \Phi(0)\right| + \Phi(0). && \text{(F.9)}
\end{aligned}$$

We can apply the Berry-Esseen Theorem to the random variable $\left\{g\left(\mathbf{X}_i; t'\right)\right\}_{i=1}^n$ because $t' \in \Delta(P)$ and the moment functions (with $s = 1$) are uniformly bounded i.e. their range is a subset of the interval $[-2, 2]$. That is,

$$\begin{aligned}
\left| P\left[\frac{\sqrt{n}E_{\hat{P}_n}\left[g\left(\mathbf{X}; t'\right)\right]}{\sqrt{E_P\left[g^2\left(\mathbf{X}; t'\right)\right]}} \leq 0\right] - \Phi(0)\right| &\leq \sup_{u \in \mathbb{R}} \left| P\left[\frac{\sqrt{n}E_{\hat{P}_n}\left[g\left(\mathbf{X}; t'\right)\right]}{\sqrt{E_P\left[g^2\left(\mathbf{X}; t'\right)\right]}} \leq u\right] - \Phi(u)\right| && \text{(F.10)}\\
&\leq \frac{C_0 E_P\left|g\left(\mathbf{X}; t'\right)\right|^3}{\sqrt{n}\left(E_P\left[g^2\left(\mathbf{X}; t'\right)\right]\right)^{3/2}} && \text{(F.11)}\\
&\leq \frac{C_0 2^3}{\sqrt{n}c_3^{3/2}}, && \text{(F.12)}
\end{aligned}$$

where C_0 is an absolute constant. Hence,

$$\lim_{n\to+\infty} \sup_{P \in \mathcal{P}_0^s(c_1,c_2,c_3)\bigcap \mathcal{P}_{000}} P\left[\acute{P}_n = \hat{P}_n\right] \leq \lim_{n\to+\infty} \frac{C_0 2^3}{\sqrt{n}c_3^{3/2}} + \Phi(0) = \Phi(0) = \frac{1}{2}. \tag{F.13}$$

For the case $s > 1$, we follow steps identical to those above for $s = 1$, except that by Lemma E.7, the steps hold for large n as $\mathcal{T}_n$ converges with uniformity (over $\mathcal{P}_2(c_2)$) to $[\underline{t}, \overline{t}]$ at the rate $\log n/n$, which is faster than the $\sqrt{n}$-rate. Furthermore, the random variables $\{g(\mathbf{X}_i; t)\}_{i=1}^n$ for $t \in \Delta(P)$ are no longer uniformly bounded and have a third moment when $\delta \geq s - 1$ in Condition (iv) of Assumption 3.1. However, if imposing $\delta \geq s - 1$ on the parameter space $\mathcal{P}$ is undesirable, then we can use a generalized Berry-Esseen Theorem due to Feller (1968). Theorem 1 of Feller (1968) does not require the existence of third moments for the random variables $\{g(\mathbf{X}_i; t)\}_{i=1}^n$. Using the envelope function (F.1), we set the following primitives in the notation of his paper

$$\tau_k' = n^{1/e}, -\tau_k = -n^{1/e}\ \forall k = 1, \ldots, n, \quad \text{where} \quad e > 6(s-1), \tag{F.14}$$

$$c = n \sup_{P \in \mathcal{P}_0^s(c_1,c_2,c_3) \bigcap \mathcal{P}_{000}} E_P\left[\sum_{K=A,B} |X^K|^{3(s-1)} 1\left[-n^{1/e} < X^K < n^{1/e}\right]\right] \tag{F.15}$$

$$b' = n \sup_{P \in \mathcal{P}_0^s(c_1,c_2,c_3) \bigcap \mathcal{P}_{000}} E_P\left[\sum_{K=A,B} |X^K|^{2(s-1)} 1\left[|X^K| > n^{1/e}\right]\right] \tag{F.16}$$

to deduce that

$$\sup_{u \in \mathbb{R}} \left| P\left[\frac{\sqrt{n} E_{\hat{P}_n}[g(\mathbf{X}; t')]}{\sqrt{E_P[g^2(\mathbf{X}; t')]}} \leq u\right] - \Phi(u)\right| \leq 6\left(\frac{c}{n^{3/2} c_3^{3/2}} + \frac{b'}{n c_3}\right). \tag{F.17}$$

Hence,

$$\frac{c}{n^{3/2} c_3^{3/2}} = \frac{\sup_{P \in \mathcal{P}_0^s(c_1,c_2,c_3) \bigcap \mathcal{P}_{000}} E_P\left[\sum_{K=A,B} |X^K|^{3(s-1)} 1\left[-n^{1/e} < X^K < n^{1/e}\right]\right]}{\sqrt{n} c_3^{3/2}} \tag{F.18}$$

$$\leq \frac{2n^{3(s-1)/e}}{\sqrt{n} c_3^{3/2}} \tag{F.19}$$

$$\leq 2c_3^{-3/2} n^{\frac{3(s-1)}{e} - \frac{1}{2}} \to 0 \quad \text{as} \quad n \to +\infty, \tag{F.20}$$

and

$$\frac{b'}{n c_3} = \frac{\sup_{P \in \mathcal{P}_0^s(c_1,c_2,c_3) \bigcap \mathcal{P}_{000}} E_P\left[\sum_{K=A,B} |X^K|^{2(s-1)} 1\left[|X^K| > n^{1/e}\right]\right]}{c_3} \tag{F.21}$$

$$\leq \sum_{K=A,B} \frac{\sup_{P \in \mathcal{P}} E_P\left[|X^K|^{2(s-1)} 1\left[|X^K| > n^{1/e}\right]\right]}{c_3} \to 0 \quad \text{as} \quad n \to +\infty, \tag{F.22}$$

since $\sup_{P\in\mathcal{P}} E_P\left[|X^K|^{2(s-1)}\right] < +\infty$ for $K = A, B$ by Condition (iv) of Assumption 3.1. This proves the limit (F.2).

Now we prove the limit (F.3). Because of the event inclusion $\left\{\Delta\left(\acute{P}_n\right) = \emptyset\right\} \subseteq \left\{\acute{P}_n = \hat{P}_n\right\}$, it follows that

$$\lim_{n\to+\infty} \sup_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)\bigcap\mathcal{P}_{000}} P\left[\Delta\left(\acute{P}_n\right) = \emptyset\right] \leq \lim_{n\to+\infty} \sup_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)\bigcap\mathcal{P}_{000}} P\left[\acute{P}_n = \hat{P}_n\right] \leq \frac{1}{2},$$

which implies

$$\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)\bigcap\mathcal{P}_{000}} P\left[\Delta\left(\acute{P}_n\right) \neq \emptyset\right] \geq \frac{1}{2}.$$

F.2. Properties of Lagrange Multipliers under H_0

This subsection presents the properties of the Lagrange multipliers under H_0 arising in the discretized constrained empirical likelihood problem introduced in Section 4. This optimization problem has the following Lagrangian function:

$$\mathcal{L} = \sum_{i=1}^{n} \log(p_i) + \lambda\left(1 - \sum_{i=1}^{n} p_i\right) - n\sum_{t\in\mathcal{T}_N} \mu_t \sum_{i=1}^{n} p_i g(\mathbf{X}_i; t), \tag{F.23}$$

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 $t \in \mathcal{T}_n$ are the multipliers on the inequality constraints. The Karush-Kuhn-Tucker (KKT) conditions are

$$\frac{1}{p_i} = \lambda + n\sum_{t\in\mathcal{T}_n} \mu_t g(\mathbf{X}_i; t) \quad i = 1, 2, \ldots, n \tag{F.24}$$

$$\sum_{i=1}^{n} p_i g(\mathbf{X}_i; t) \leq 0 \quad \forall t \in \mathcal{T}_n, \quad \sum_{i=1}^{n} p_i = 1 \tag{F.25}$$

$$\mu_t \sum_{i=1}^{n} p_i g(\mathbf{X}_i; t) = 0 \quad \forall t \in \mathcal{T}_n. \tag{F.26}$$

In classical optimization theory, the existence of $\acute{\lambda}$ and $\acute{\mu}_t$ for $t \in \mathcal{T}_n$ 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

$$\mathcal{S}_n = \left\{\exists d \in \mathbb{R}^n : \sum_{i=1}^{n} d_i = 0 \quad \text{and} \quad \sum_{i=1}^{n} d_i g(\mathbf{X}_i; t) < 0 \; \forall t \in \Delta\left(\acute{P}_n\right)\right\} \text{ where} \tag{F.27}$$

$$\Delta\left(\acute{P}_n\right) = \left\{t \in \mathcal{T}_n : \sum_{i=1}^{n} \acute{p}_i g(\mathbf{X}_i; t) = 0\right\}. \tag{F.28}$$

The following result establishes the existence of the Lagrange multipliers with probability approaching unity, with uniformity over the set of probabilities of the form $\mathcal{P}_1(c_1)$.

LEMMA F.2 (EXISTENCE). *Given* $c_1 \in (0,1)$, *suppose that* $P_0 \in \mathcal{P}_0(c_1, c_2)$. *Then*

$$\sup_{P \in \mathcal{P}_1(c_1)} P[\mathcal{S}_n] \to 1 \quad as \quad n \to +\infty. \tag{F.29}$$

PROOF. The proof proceeds by using the direct method. Given $c_1 \in (0,1)$, for large enough n and uniformly in $\mathcal{P}_1(c_1)$ Corollary E.1 and Part 2 of E.6 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 $\mathcal{P}_1(c_1)$.

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

$$\operatorname{sign}(d_i) = \begin{cases} < 0, & \text{if } g(\mathbf{X}_i, t) \geq 0 \; \forall t \in \mathcal{T}_n \\ > 0, & \text{if } g(\mathbf{X}_i, t) < 0 \; \forall t \in \mathcal{T}_n, \end{cases} \tag{F.30}$$

Lemma E.1 implies the occurrence of the event $\{\exists i : g(\mathbf{X}_i, t) < 0 \; \forall t \in \mathcal{T}_n\}$ with probability approaching one, uniformly in $\mathcal{P}_1(c_1)$. Therefore, the above construction is asymptotically feasible. Such vectors $\mathbf{d}$ 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 $\mathcal{P}_1(c_1)$.

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

$$\acute{p}_i = \frac{1}{n}\left(\frac{1}{1 + \sum_{t \in \Delta(\acute{P}_n)} \acute{\mu}_t g(\mathbf{X}_i; t)}\right) \quad i = 1, 2, \ldots, n, \tag{F.31}$$

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

$$\Lambda_n\left(\acute{P}_n\right) = \left\{\acute{\mu}_t \; t \in \Delta(\acute{P}_n) \text{ that satisfy } (F.24) - \; (F.26)\right\}. \tag{F.32}$$

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,

$$l_w^1 = \left\{ a = (a_1, a_2, \ldots, a_w) \in \mathbb{R}^w : \sum_{j=1}^{w} |a_j| < +\infty \right\}, \tag{F.33}$$

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

Lemma F.3 (Asymptotic Bound for Lagrange Multipliers). *Let $\Delta\left(\acute{P}_n\right)$ be given by (F.28) and $\omega_n = \left|\Delta(\acute{P}_n)\right|$. For each $s \in \mathbb{Z}_+$, $c_1 \in (0,1), c_2 \in \left(0, \frac{1}{(\bar{t}-\underline{t})^2}\right)$ and $c_3 \in (0, +\infty)$,*

(i) $\lim_{n\to+\infty} \inf_{P\in\mathcal{P}_0 \bigcap \mathcal{P}_1(c_1)} P\left[\Delta\left(\acute{P}_n\right) \subseteq \Delta(P)\right] = 1.$

(ii) Denote the vector of Lagrange multipliers on the constraints binding constraints by $\boldsymbol{\acute{\mu}}$ and the $l^1_{\omega_n}$ norm of the vector $\boldsymbol{\acute{\mu}}$ by $||\boldsymbol{\acute{\mu}}||_{l^1_{\omega_n}}$. Then $\forall \epsilon > 0$

$$\lim_{n\to+\infty} \sup_{P\in\mathcal{P}_0^s(c_1,c_2,c_3)} P\left[\sup_{\boldsymbol{\acute{\mu}}\in\Lambda_n\left(\acute{P}_n\right)} ||\boldsymbol{\acute{\mu}}||_{l^1_{\omega_n}} > \epsilon\right] = 0, \quad \textit{and} \tag{F.34}$$

$$\sup_{\boldsymbol{\acute{\mu}}\in\Lambda_n\left(\acute{P}_n\right)} ||\boldsymbol{\acute{\mu}}||_{l^1_{\omega_n}} = O_P\left(n^{-1/2}\right) \tag{F.35}$$

uniformly in $\mathcal{P}_0^s(c_1, c_2, c_3)$.

Proof.

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

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

Given $c_1 \in (0,1)$, for large enough n and uniformly in $\mathcal{P}_0 \bigcap \mathcal{P}_1(c_1)$ Corollary E.1 and Part 2 of E.6 imply that $\acute{\mathbf{p}}$ exists and is unique. Consider any $t \in [\underline{t}, \bar{t}]$. For large enough n, Property 1 and the non-negativity of the Lagrange multipliers implies that

$$\sum_{i=1}^{n} p_i' g(X_i;t) \leq \frac{1}{n}\sum_{i=1}^{n} g(X_i;t) = \frac{1}{n}\sum_{i=1}^{n} g(X_i;t) - E_{P_0}\left[g(X;t)\right] + E_{P_0}\left[g(X;t)\right] \tag{F.36}$$

Now, for $t \notin \Delta(P)$, it follows that $E_P\left[g(X;t)\right] < 0$. By the Central Limit Theorem,

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

uniformly in $\mathcal{P}_0 \bigcap \mathcal{P}_1(c_1)$. Thus, for sufficiently large n, equation (F.36) simplifies to

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

This shows that $t \notin \Delta(\acute{P}_n)$ with probability approaching unity, uniformly over probabilities in $\mathcal{P}_0 \bigcap \mathcal{P}_1(c_1)$.

(ii) On the event $\left\{\Delta(\acute{P}_n)=\emptyset\right\}$, the Lagrange multipliers that solve the KKT conditions is a singleton equal to the zero vector i.e. $\Lambda_n\left(\acute{P}_n\right)=\mathbf{0}\in l^1_{\omega_n}$. Therefore, on the event $\left\{\Delta(\acute{P}_n)\neq\emptyset\right\}$, $\Lambda_n\left(\acute{P}_n\right)$ is not equal to $\mathbf{0}$. In consequence.

$$P\left[\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}>\epsilon\right]=P\left[\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}>\epsilon,\Delta(\acute{P}_n)\neq\emptyset\right]. \quad \text{(F.37)}$$

Thus, our approach in the proof will be to construct an upper bound on $\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}$ on the event $\left\{\Delta(\acute{P}_n)\neq\emptyset\right\}$ that is $o_p(1)$ uniformly in $\mathcal{P}^s_0(c_1,c_2,c_3)$.

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

$$\Delta(\acute{P}_n)=\{t_1,t_2,\ldots,t_{\omega_n}\}. \quad \text{(F.38)}$$

Therefore, the probabilities (F.31) can be expressed as

$$\acute{p}_i=\frac{1}{n}\left(1+\sum_{j=1}^{\omega_n}\acute{\mu}_j g(X_i;t_j)\right)^{-1} \quad \text{(F.39)}$$

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

$$\sum_{i=1}^{n}\acute{p}_i g(X_i;t_j)=\frac{1}{n}\sum_{i=1}^{n}\frac{g(X_i;t_j)}{1+\sum_{j=1}^{\omega_n}\acute{\mu}_j g(X_i;t_j)}=0 \quad \text{(F.40)}$$

To express the system of equations described by (F.40) in vectorised form, define the vector

$$\mathbf{g}_i=[g(X_i;t_1),\,g(X_i;t_2),\,\ldots,\,g(X_i;t_{\omega_n})]^T \quad \text{(F.41)}$$

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 (F.40) for all $t\in\Delta(\acute{P}_n)$ can be written in the following form

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{g}_i}{1+(\acute{\boldsymbol{\mu}})^T\mathbf{g}_i}=\mathbf{0}\implies\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{g}_i}{1+(\acute{\boldsymbol{\mu}})^T\mathbf{g}_i}\right)=0 \quad \text{(F.42)}$$

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 (F.42) gives

$$\boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \left(1 - \frac{\mathbf{g}_i^T \acute{\boldsymbol{\mu}}}{1+Y_i} \right) \right) = 0$$

$$\boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) = \boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{g}_i \mathbf{g}_i^T \acute{\boldsymbol{\mu}}}{1+Y_i} \right)$$

$$\boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) = \boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{g}_i \mathbf{g}_i^T ||\acute{\boldsymbol{\mu}}|| \boldsymbol{\theta}}{1+Y_i} \right)$$

$$\therefore \boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) = ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{g}_i \mathbf{g}_i^T}{1+Y_i} \right) \boldsymbol{\theta} \tag{F.43}$$

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

$$Y_{max} = \max_i \sum_{j=1}^{\omega_n} \acute{\mu}_j \left| g\left(\mathbf{X}_i; t_j\right) \right| \tag{F.44}$$

$$\leq \sum_{j=1}^{\omega_n} \acute{\mu}_j \max_i \left| g\left(\mathbf{X}_i; t_j\right) \right| \tag{F.45}$$

$$= ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} s \max \left\{ |\underline{t}|, |\overline{t}| \right\}^{s-1} \sum_{K=A,B} \max_i \left| X^K \right|^{s-1}, \tag{F.46}$$

where we used the envelope function (F.1) to bound the moment functions, uniformly in $t \in [\underline{t}, \overline{t}]$.
Let $X_{max} = s \max \left\{ |\underline{t}|, |\overline{t}| \right\}^{s-1} \sum_{K=A,B} \max_i \left| X_i^K \right|^{s-1}$, and consider

$$||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \left(\boldsymbol{\theta}^T \hat{\Sigma}_{\Delta(\acute{P}_n)} \boldsymbol{\theta} \right) = ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \left(\boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \mathbf{g}_i^T \right) \boldsymbol{\theta} \right)$$

$$\leq ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \left(\boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{g}_i \mathbf{g}_i^T}{1+Y_i} \right) \boldsymbol{\theta} \right) (1 + Y_{max})$$

$$\leq ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \left(\boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{g}_i \mathbf{g}_i^T}{1+Y_i} \right) \boldsymbol{\theta} \right) (1 + X_{max} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}})$$

$$\therefore ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \left(\boldsymbol{\theta}^T \hat{\Sigma}_{\Delta(\acute{P}_n)} \boldsymbol{\theta} \right) \leq \boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) (1 + X_{max} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}) \tag{F.47}$$

where the last line results from substituting the expression given in (F.43). Rearranging (F.47) gives

$$||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}\left[\boldsymbol{\theta}^T\hat{\Sigma}_{\omega_n}\boldsymbol{\theta}-\boldsymbol{\theta}^T\left(\frac{X_{max}}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)\right]\leq\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)\quad\forall\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right),\tag{F.48}$$

since the derivation above holds for each $\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)$. We consider the components of (F.48) to find the required asymptotic bound on $||\acute{\boldsymbol{\mu}}||$. From part (i) of this lemma, for large n we have $\Delta(\acute{P}_n)\subset\Delta(P)$. This means for large n, we have that for all $t\in\Delta(\acute{P}_n)$, $E_P\left[g(X;t_j)\right]=0$. As a result,

$$\begin{aligned}\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)&=\sum_{j=1}^{\omega_n}\theta_j\left(\frac{1}{n}\sum_{i=1}^{n}g(X_i;t_j)-E_P\left[g(X;t_j)\right]\right)\\ \left|\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)\right|&\leq\sum_{j=1}^{\omega_n}\theta_j\left|\frac{1}{n}\sum_{i=1}^{n}g(X_i;t_j)-E_P\left[g(X;t_j)\right]\right|\\ &\leq\max_j\left|\frac{1}{n}\sum_{i=1}^{n}g(X_i;t_j)-E_P\left[g(X;t_j)\right]\right|\left(\sum_{j=1}^{\omega_n}\theta_j\right)\\ &\leq\sup_{t\in[\underline{t},\overline{t}]}\left|\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t\right)-E_P\left[g\left(\mathbf{X};t\right)\right]\right|\end{aligned}\tag{F.49}$$

The last line follows from the fact that $\sum_{j=1}^{\omega_n}\theta_j=1$ by construction. The upper bound given by equation (F.49) is $o_P(1)$ uniformly in $\mathcal{P}$. This follows from the moment functions being uniformly Glivenko-Cantelli: it is a Vapnik-Chervonenkis class with square-integrable envelope function (F.1), uniformly in $\mathcal{P}$. Therefore, this upper bound is also $o_P(1)$ uniformly in $\mathcal{P}_0^s(c_1,c_2,c_3)$.

Now we focus on the large-sample behavior of the term $\boldsymbol{\theta}^T\left(\frac{X_{max}}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)$. We will show that it is also $o_P(1)$ uniformly in $\mathcal{P}_0^s(c_1,c_2,c_3)$. We have

$$\begin{aligned}\left|\boldsymbol{\theta}^T\left(\frac{X_{max}}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)\right|&\leq X_{max}\sum_{j=1}^{\omega_n}\theta_j\left|\frac{1}{n}\sum_{i=1}^{n}g(X_i;t_j)-E_P\left[g(X;t_j)\right]\right|\\ &\leq X_{max}\max_j\left|\frac{1}{n}\sum_{i=1}^{n}g(X_i;t_j)-E_P\left[g(X;t_j)\right]\right|\left(\sum_{j=1}^{\omega_n}\theta_j\right)\\ &\leq X_{max}\sup_{t\in[\underline{t},\overline{t}]}\left|\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t\right)-E_P\left[g\left(\mathbf{X};t\right)\right]\right|.\end{aligned}\tag{F.50}$$

Now we can apply Lemma 11.2 of Owen (2001) and Theorem 2.8.2 of van der Vaart and Wellner (1996) to X_{max} and $\sup_{t\in[\underline{t},\overline{t}]}\left|\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t\right)-E_P\left[g\left(\mathbf{X};t\right)\right]\right|$, respectively, to deduce that $X_{max} = o(n^{1/2})$ and

$$\sup_{t\in[\underline{t},\overline{t}]} \left| \frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t\right) - E_P\left[g\left(\mathbf{X};t\right)\right] \right| = O_P(n^{-1/2}) \text{ uniformly in } \mathcal{P}.$$

Therefore, the right side of (F.50) is $o(n^{1/2})O_P(n^{-1/2}) = o_P(1)$, uniformly in $\mathcal{P}$. In consequence, $\boldsymbol{\theta}^T\left(\frac{X_{max}}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)$ is also $o_P(1)$ uniformly in $\mathcal{P}_0^s(c_1,c_2,c_3)$.

Now, for sufficiently large n, part (i) of this lemma tells us that $\Delta(\acute{P}_n) \subset \Delta(P)$, with probability approaching unity uniformly over $\mathcal{P}_0 \bigcap \mathcal{P}_1(c_1)$. Whether or not $\Delta(P) \neq \emptyset$ the following manipulation holds. Because $\mathcal{P}_0^s(c_1,c_2,c_3) \subseteq \mathcal{P}_0 \bigcap \mathcal{P}_1(c_1)$, Definition 5.1 implies that $\boldsymbol{\theta}^T\Sigma_{\Delta(\acute{P}_n)}\boldsymbol{\theta} \geq c_3 > 0$ holds with probability tending to unity uniformly over $\mathcal{P}_0^s(c_1,c_2,c_3)$. Using this result and the bound from equation (F.49), we can rewrite (F.48) as

$$||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \leq \frac{o_P(1)}{c_3+o_P(1)} \quad \forall\acute{\boldsymbol{\mu}} \in \Lambda_n\left(\acute{P}_n\right), \quad \text{uniformly in} \quad \mathcal{P}_0^s(c_1,c_2,c_3). \tag{F.51}$$

Consequently,

$$\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \leq \frac{o_P(1)}{c_3+o_P(1)}, \quad \text{uniformly in} \quad \mathcal{P}_0^s(c_1,c_2,c_3). \tag{F.52}$$

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 $\mathcal{P}_0^s(c_1,c_2,c_3)$. 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 $\mathcal{P}_0^s(c_1,c_2,c_3)$, first note that the expression on the right side of (F.48) has this property. So that

$$\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \leq \frac{O_P(n^{-1/2})}{c_3+o_P(1)} \quad \text{uniformly in} \quad \mathcal{P}_0^s(c_1,c_2,c_3), \tag{F.53}$$

which implies

$$\sqrt{n}\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \leq \frac{O_P(1)}{c_3+o_P(1)} \quad \text{uniformly in} \quad \mathcal{P}_0^s(c_1,c_2,c_3). \tag{F.54}$$

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 $\mathcal{P}_0^s(c_1,c_2,c_3)$. Therefore, we must have that

$$\sqrt{n}\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} = O_P(1) \quad \text{uniformly in} \quad \mathcal{P}_0^s(c_1,c_2,c_3).$$

F.3. Relationship Between $E_{\hat{P}_n}\left[g\left(\mathbf{X},\cdot\right)\right]$ *and* $E_{\acute{P}_n}\left[g\left(\mathbf{X},\cdot\right)\right]$

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

PROPOSITION F.1. *For each* $s \in \mathbb{Z}_+$, $c_1 \in (0,1), c_2 \in \left(0, \frac{1}{(\bar{t}-\underline{t})^2}\right)$ *and* $c_3 \in (0,+\infty)$,

$$\sup_{t\in[\underline{t},\bar{t}]} \left| E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] - E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right]\right| = O_P(n^{-1/2}) \quad \textit{uniformly over} \quad \mathcal{P}_0^s(c_1,c_2,c_3). \tag{F.55}$$

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

$$\begin{aligned}
\left| E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] - E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right]\right| &= E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right] - E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right] && \text{(F.56)}\\
&= \sum_{i=1}^{n} \frac{1}{n} g(X_i;t) - \sum_{i=1}^{n} \acute{p}_i g(X_i;t)\\
&= \sum_{i=1}^{n} \left(\frac{1}{n} - \acute{p}_i\right) g(X_i;t)\\
&= \sum_{i=1}^{n} \frac{1}{n}\left(1 - \frac{1}{1+\sum_{j=1}^{N} \acute{\mu}_j g(X_i;t_j)}\right) g\left(\mathbf{X}_i,t\right)\\
&= \sum_{i=1}^{n} \frac{1}{n} \cdot \frac{g\left(\mathbf{X}_i,t\right)\sum_{j=1}^{N} \acute{\mu}_j g(X_i;t_j)}{1+\sum_{j=1}^{N} \acute{\mu}_j g(X_i;t_j)}\\
&= \sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i,t\right) \sum_{j=1}^{N} \acute{\mu}_j g(X_i;t_j). && \text{(F.57)}
\end{aligned}$$

Now using the envelope function (F.1), we can obtain the following upper bound on the term (F.57):

$$s^2 \max\left\{|\underline{t}|,|\bar{t}|\right\}^{2(s-1)} \sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \sum_{i=1}^{n} \acute{p}_i \left(\sum_{K=A,B} \left|X_i^K\right|^{s-1}\right)^2, \tag{F.58}$$

where $\Lambda_n\left(\acute{P}_n\right)$ is the set of Lagrange multipliers on $\Delta(\acute{P}_n)$ defined in (F.32). Part 2 of Lemma F.3 establishes

$$\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} = O_P(n^{-1/2}) \quad \text{uniformly over} \quad \mathcal{P}_0^s(c_1,c_2,c_3). \tag{F.59}$$

Thus, to deduce the desired result, we need to show that

$$\sum_{i=1}^{n} \acute{p}_i \left(\sum_{K=A,B} \left|X_i^K\right|^{s-1}\right)^2 = O_P(1) \quad \text{uniformly over} \quad \mathcal{P}_0^s(c_1,c_2,c_3). \tag{F.60}$$

For each $i = 1, \ldots, n$, we will apply the expansion $\frac{1}{1+Y_i} = 1 - \frac{Y_i}{1+Y_i}$ to

$$\acute{p}_i = \frac{1}{n}\left(1 + \sum_{j=1}^{N} \acute{\mu}_j g(X_i;t_j)\right)^{-1},$$

where $Y_i = \sum_{j=1}^{N} \acute{\mu}_j g(X_i; t_j)$, to deduce that the left side of (F.60) equals

$$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2 - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{N}\acute{\mu}_j g(X_i;t_j)}{1+\sum_{j=1}^{N}\acute{\mu}_j g(X_i;t_j)}\right)\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2. \tag{F.61}$$

Next, apply Jensen's inequality to the second term in (F.61) to obtain the following upper bound on (F.61)

$$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2 - \frac{\sum_j \acute{\mu}_j \frac{1}{n}\sum_{i=1}^{n}\left(g\left(\mathbf{X}_i,t_j\right)\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2\right)}{1+\sum_j \acute{\mu}_j \frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i,t_j\right)}. \tag{F.62}$$

By Condition (iv) of Assumption 3.1, the term $\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2$ converges in probability to $E_P\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2$, uniformly in $P\in\mathcal{P}$, which implies that it converges uniformly in $P\in\mathcal{P}_0^s(c_1,c_2,c_3)$ as $\mathcal{P}_0^s(c_1,c_2,c_3)\subseteq\mathcal{P}$; therefore, this term is $O_P(1)$ uniformly in $P\in\mathcal{P}_0^s(c_1,c_2,c_3)$. Next, we show that the second term in (F.62) is $o_P(1)$ uniformly in $P\in\mathcal{P}_0^s(c_1,c_2,c_3)$, which implies that it is $O_P(1)$ uniformly in $P\in\mathcal{P}_0^s(c_1,c_2,c_3)$.

The modulus of the second term in (F.62) is bounded above by

$$\frac{\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}\max_{i,j}\left|g\left(\mathbf{X}_i,t_j\right)\right|\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2}{\left|1+\sum_j \acute{\mu}_j\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i,t_j\right)\right|}. \tag{F.63}$$

We tackle the numerator and denominator of (F.63) separately. Using the envelope function (F.1), the numerator is bounded above by

$$s\max\left\{|\underline{t}|,|\overline{t}|\right\}^{(s-1)}\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}\left(\sum_{K=A,B}\max_i\left|X_i^K\right|^{s-1}\right)\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2. \tag{F.64}$$

By Condition (iv) of Assumption 3.1, an application of Lemma 11.2 of Owen (2001) and Part 2 of Lemma F.3 imply that $\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}\left(\sum_{K=A,B}\max_i\left|X_i^K\right|^{s-1}\right) = o_P(1)$ uniformly in $P\in\mathcal{P}_0^s(c_1,c_2,c_3)$. Furthermore, $\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2 = O_P(1)$ uniformly in $P\in\mathcal{P}_0^s(c_1,c_2,c_3)$, which implies that the term (F.64) is $o_P(1)$ uniformly in $P\in\mathcal{P}_0^s(c_1,c_2,c_3)$.

Next, we tackle the denominator. We will show that $\sum_j \acute{\mu}_j\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i,t_j\right) = o_P(1)$ uniformly in $P\in\mathcal{P}_0^s(c_1,c_2,c_3)$. Observe for large enough n and uniformly in

$P \in \mathcal{P}_0^s(c_1, c_2, c_3)$ that

$$\left|\sum_j \acute{\mu}_j \frac{1}{n}\sum_{i=1}^n g(\mathbf{X}_i, t_j)\right| \leq \sup_{\acute{\boldsymbol{\mu}} \in \Lambda_n(\acute{P}_n)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \sup_{t \in \Delta(\acute{P}_n)} \left|E_{\hat{P}_n}[g(\mathbf{X}, t)]\right| \tag{F.65}$$

$$= O_P(n^{-1/2}) \sup_{t \in \Delta(\acute{P}_n)} \left|E_{\hat{P}_n}[g(\mathbf{X}, t)]\right| \tag{F.66}$$

$$\leq O_P(n^{-1/2}) \sup_{t \in \Delta(P)} \left|E_{\hat{P}_n}[g(\mathbf{X}, t)]\right| \tag{F.67}$$

$$= O_P(n^{-1/2})O_P(n^{-1/2}) = O_P(n^{-1}) = o_P(1) \tag{F.68}$$

by Lemma F.3 and the Uniform Central Limit Theorem. This concludes the proof.

G. Technical Lemmas for Theorems 6.1 and 6.2

G.1. Theorem 6.1

This subsection presents two technical lemmas that are useful for proving Theorem 6.1. They are a consequence of the condition $P\left[\sup_{t \in [\underline{t}, \overline{t}]} g(\mathbf{X}_1; t) < 0\right] > 0$ being true. The first lemma is similar to Lemma E.1, but we now do not constrain P to satisfy H_0.

LEMMA G.1. *Suppose $P_0 \in \mathcal{P}$ and let I_n^- be given by (E.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 E.1. We show that the probability of the complement of $\{I_n^- \neq \emptyset\}$ converges to zero. This set is

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

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

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

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

$n \to +\infty$ by Condition (i) of Assumption 3.1.

The second lemma concerns the existence and uniqueness of the constrained empirical likelihood probability vector $\acute{\mathbf{p}}$. Recall that

$$\mathcal{H}_n = \left\{p_i, i = 1, \ldots, n; \sum_{i=1}^n p_i = 1, p_i \geq 0, \forall i = 1, \ldots, n\right\}$$

and that its interior is $\mathcal{H}_n^\circ$. Additionally, recall that the constraint is

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

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

LEMMA G.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 G.1 implies that the event

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

occurs with probability approaching 1, since $\mathcal{T}_n \subset [\underline{t}, \overline{t}]$ for each n. The rest of the proof proceeds using steps similar to those in the proof of Lemma E.5; therefore, we omit them for brevity.

G.2. Theorems 6.2

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). In the context of the paper, we have the triangular array

$$\left\{ \{\mathbf{X}_{i,j}, i = 1, \ldots, n\}, n = 1, 2, \ldots \right\}, \tag{G.4}$$

where for each n $\{\mathbf{X}_{i,j}, i = 1, \ldots, n\}$ is bivariate random sample form P_n that satisfies Assumption 6.1.

First, we introduce a technical lemma that shows the largest value in a sample of size n in the triangular arrays of row-wise IID random variables cannot grow to infinite as fast as $\sqrt{n}$. We establish this result, though, in the context of the paper.

LEMMA G.3. *Suppose that Assumption 6.1 holds. Then*

$$\frac{\max_{1\leq i\leq n} \sum_{K=A,B} |X_{i,n}|^{s-1}}{\sqrt{n}} = o_{P_n}(1).$$

PROOF. The proof proceeds by the direct method. We will show that

$$\forall \epsilon > 0, \lim_{n\to+\infty} P_n \left[\frac{\max_{1\leq i\leq n} \sum_{K=A,B} |X_{i,n}|^{s-1}}{\sqrt{n}} \leq \epsilon \right] = 1, \tag{G.5}$$

which implies the desired result.

Under Assumption 6.1 and $\mathcal{P}$, we have that

$$\sup_n E_{P_n}\left[\sum_{K=A,B}|X_{i,n}|^{s-1}\right]^2 \leq \sup_{P\in\mathcal{P}} E_P\left[\sum_{K=A,B}|X_i|^{s-1}\right]^2 < +\infty, \tag{G.6}$$

holds. Then, for every $\epsilon > 0$ Markov's inequality implies that

$$\lim_{n\to+\infty}\sum_{j=1}^{n} P_j\left[\left[\sum_{K=A,B}|X_{i,j}|^{s-1}\right]^2 > n\epsilon\right] < +\infty. \tag{G.7}$$

As $\sum_{j=1}^{n} P_j\left[\left[\sum_{K=A,B}|X_{i,j}|^{s-1}\right]^2 > n\epsilon\right]$ is a convergent series of non-negative terms, it follows that

$$\lim_{n\to+\infty} P_n\left[\left[\sum_{K=A,B}|X_{i,n}|^{s-1}\right]^2 > n\epsilon\right] = 0, \tag{G.8}$$

holds. In consequence, the limit of the complementary probabilities satisfies

$$\lim_{n\to+\infty} P_n\left[\frac{\sum_{K=A,B}|X_{i,n}|^{s-1}}{\sqrt{n}} \leq \epsilon\right] = 1, \tag{G.9}$$

which implies that

$$\lim_{n\to+\infty} P_n\left[\frac{\max_{1\leq i\leq n}\sum_{K=A,B}|X_{i,n}|^{s-1}}{\sqrt{n}} \leq \sqrt{\epsilon}\right] = 1, \tag{G.10}$$

holds. The limit (G.10) implies the desired result as $\epsilon > 0$ was arbitrary and because the square-root function on the positive reals is a monotonic function i.e. there is a one-to-one correspondence between $\sqrt{\epsilon}$ and ϵ.

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

LEMMA G.4. *Suppose that the conditions of Theorem 6.2 hold. Then*

(a) $\lim_{n\to+\infty} P_n\left[I_n^- \neq \emptyset\right] = 1$, *where* I_n^- *is defined in (E.1).*

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

(c) $\lim_{n\to+\infty} P_n\left[\mathcal{S}_n\right] = 1$, *where* $\mathcal{S}_n$ *is the event defined in (F.27).*

(d) $\lim_{n\to+\infty} P_n\left[E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right] \leq E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]\ \forall t\in\left[\underline{t},\overline{t}\right]\right] = 1$.

(e) $\sqrt{n}\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n(\acute{P}_n)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} = O_{P_n}(1)$.

PROOF. Under the conditions of Theorem 6.2, the steps for proving parts 1 to 4 of this lemma are identical to their counterparts in Appendix F, 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

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

The proof will follow steps similar to those of pat (i) of Lemma F.3. Proceeding by contraposition, we need to show that

$$t\notin C \implies t\notin\Delta(\acute{P}_n) \tag{G.11}$$

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

$$E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right]\leq E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]=E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]-E_{P_n}\left[g\left(\mathbf{X};t\right)\right]+E_{P_n}\left[g\left(\mathbf{X};t\right)\right]. \tag{G.12}$$

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,

$$E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]-E_{P_n}\left[g\left(\mathbf{X};t\right)\right]=o_{P_n}(1). \tag{G.13}$$

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 form above by the square-integrable envelope function (F.1) under the local alternatives. Thus for sufficiently large n, the inequality (G.12) simplifies to

$$E_{\acute{P}_n}\left[g\left(\mathbf{X};t\right)\right]\leq H(t)<0 \quad \text{as} \quad n\to+\infty. \tag{G.14}$$

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 F.3, and following identical steps to those up to the inequality (F.48), we have that

$$||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}\left[\boldsymbol{\theta}^T\hat{\Sigma}_{\omega_n}\boldsymbol{\theta}-\boldsymbol{\theta}^T\left(\frac{X_{max}}{n}\sum_{i=1}^{n}\mathbf{g}_i\right)\right]\leq\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right) \quad \forall\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right), \tag{G.15}$$

where

$$X_{max}=s\max\left\{|\underline{t}|,|\overline{t}|\right\}^{s-1}\sum_{K=A,B}\max_{1\leq i\leq n}\left|X_{i,n}^K\right|^{s-1}, \tag{G.16}$$

$$\mathbf{g}_i=\left[g(X_i;t_1),\,g(X_i;t_2),\,\ldots,\,g(X_i;t_{\omega_n})\right]^T, \tag{G.17}$$

$$\Delta(\acute{P}_n)=\{t_1,t_2,\ldots,t_{\omega_n}\} \tag{G.18}$$

and $\boldsymbol{\theta} \in \mathbb{R}_{+}^{\omega_n}$ with $||\boldsymbol{\theta}||_{l^1_{\omega_n}} = 1$. Noting that

$$\boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) = \sum_{j=1}^{\omega_n} \theta_j \left(\frac{1}{n} \sum_{i=1}^{n} g(X_i; t_j) - E_{P_n}\left[g(X; t_j)\right] \right) + \sum_{j=1}^{\omega_n} \theta_j (\delta(t_j)/\sqrt{n}) \tag{G.19}$$

$$\leq \sup_{t \in [\underline{t}, \overline{t}]} \left| E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{P_n}\left[g\left(\mathbf{X}; t\right)\right] \right| + \sup_{t \in [\underline{t}, \overline{t}]} \delta(t)/\sqrt{n} = o_{P_n}(1) \tag{G.20}$$

by the Uniform WLLN for triangular arrays of random variables that are row-wise IID and that

$$\begin{aligned} \boldsymbol{\theta}^T \left(\frac{X_{max}}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) &= \sum_{j=1}^{\omega_n} \theta_j \left(\frac{X_{max}}{n} \sum_{i=1}^{n} g(X_i; t_j) - E_{P_n}\left[g(X; t_j)\right] \right) \\ &\quad + \sum_{j=1}^{\omega_n} \theta_j (\delta(t_j)/\sqrt{n}) \\ &\leq X_{max} \sup_{t \in [\underline{t}, \overline{t}]} \left| E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{P_n}\left[g\left(\mathbf{X}; t\right)\right] \right| + \frac{X_{max}}{\sqrt{n}} \sup_{t \in [\underline{t}, \overline{t}]} \delta(t) \\ &= \frac{X_{max}}{\sqrt{n}} \sqrt{n} \sup_{t \in [\underline{t}, \overline{t}]} \left| E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{P_n}\left[g\left(\mathbf{X}; t\right)\right] \right| + \frac{X_{max}}{\sqrt{n}} \sup_{t \in [\underline{t}, \overline{t}]} \delta(t) \\ &= o_{P_n}(1) O_{P_n}(1) + o_{P_n}(1) = o_{P_n}(1) \end{aligned}$$

by Lemma G.3 and the Uniform Central Limit Theorem (i.e. Theorem 2.8.9 in van der Vaart and Wellner, 1996), we have that

$$\sup_{\acute{\boldsymbol{\mu}} \in \Lambda_n(\acute{P}_n)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \leq \frac{\boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right)}{\left[\boldsymbol{\theta}^T \hat{\Sigma}_{\omega_n} \boldsymbol{\theta} - \boldsymbol{\theta}^T \left(\frac{X_{max}}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) \right]} \tag{G.21}$$

since Property 1 and 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,

$$\sqrt{n} \sup_{\acute{\boldsymbol{\mu}} \in \Lambda_n(\acute{P}_n)} ||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}} \leq \frac{\sqrt{n} \boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right)}{\left[\boldsymbol{\theta}^T \hat{\Sigma}_{\omega_n} \boldsymbol{\theta} - \boldsymbol{\theta}^T \left(\frac{X_{max}}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) \right]}. \tag{G.22}$$

To conclude the proof, all we need to do is to show that the numerator on the right side of the inequality (G.22) is $O_{P_n}(1)$. Noting the inequality (G.20) above, we have that

$$\sqrt{n} \boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) \leq \sqrt{n} \sup_{t \in [\underline{t}, \overline{t}]} \left| E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right] - E_{P_n}\left[g\left(\mathbf{X}; t\right)\right] \right| + \sup_{t \in [\underline{t}, \overline{t}]} \delta(t), \tag{G.23}$$

where $\sqrt{n}\sup_{t\in[\underline{t},\overline{t}]}\left|E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]-E_{P_n}\left[g\left(\mathbf{X};t\right)\right]\right|=O_{P_n}(1)$ by Theorem 2.8.9 in van der Vaart and Wellner (1996), and $\sup_{t\in[\underline{t},\overline{t}]}\delta(t)$ is finite by Part (iii) of Assumption 6.1, which implies the desired result. Therefore,

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

which concludes the proof.

The next result is the counterpart of Proposition F.1 under the sequence of local alternatives demarcated by Assumption 6.1.

PROPOSITION G.1. *Suppose that Assumption 6.1 holds. Then For each* $s\in\mathbb{Z}_+$,

$$\sup_{t\in[\underline{t},\overline{t}]}\left|E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right]-E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right]\right|=O_{P_n}(n^{-1/2}). \tag{G.24}$$

PROOF. The proof follows steps identical to those in the proof of Proposition F.1 except that the limits are taken under the sequence of local alternatives. Firstly, we can follow the same steps to deduce that $\sup_{t\in[\underline{t},\overline{t}]}\left|E_{\hat{P}_n}\left[g\left(\mathbf{X},t\right)\right]-E_{\acute{P}_n}\left[g\left(\mathbf{X},t\right)\right]\right|$ is bounded above by

$$s^2\max\left\{|\underline{t}|,|\overline{t}|\right\}^{2(s-1)}\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}\sum_{i=1}^{n}\acute{p}_i\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2, \tag{G.25}$$

Then Part 5 of Lemma G.4 implies that $\sup_{\acute{\boldsymbol{\mu}}\in\Lambda_n\left(\acute{P}_n\right)}||\acute{\boldsymbol{\mu}}||_{l^1_{\omega_n}}=O_{P_n}(n^{-1/2})$, holds, which implies we need to show that

$$\sum_{i=1}^{n}\acute{p}_i\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2=O_{P_n}(1), \tag{G.26}$$

holds, in order to conclude the proof.

To show that (G.26) holds, we can implement the same decomposition for this term and apply Jensen's inequality as in the proof of Proposition F.1 to show that it is bounded above by the expression (F.62), which we repeat here for convenience:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2-\frac{\sum_j\acute{\mu}_j\frac{1}{n}\sum_{i=1}^{n}\left(g\left(\mathbf{X}_i,t_j\right)\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2\right)}{1+\sum_j\acute{\mu}_j\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i,t_j\right)}. \tag{G.27}$$

Condition (iv) of Assumption 3.1 implies that the term $\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{K=A,B}\left|X_i^K\right|^{s-1}\right)^2=O_{P_n}(1)$. Then we can tackle the denominator and numerator of the second term in (G.27) separately. For the numerator, we follow the same steps as in Proposition F.1 but use Lemma G.3 and Part 5 of Lemma G.4 instead of Lemma 11.2 of Owen (2001) and Part 2 of Lemma F.3, respectively, to deduce that it is $o_{P_n}(1)$, under the sequence of local alternatives. For the denominator, again, we follow the same steps as in Proposition F.1 except that we replace the contact set $\Delta(P)$ with the set C and use Part 5 of Lemma G.4. and the Theorem 2.8.9 of van der Vaart and Wellner (1996) instead of Part 2 of Lemma F.3 and the Uniform Central Limit Theorem, respectively, to deduce that

$$\sum_j \acute{\mu}_j \frac{1}{n} \sum_{i=1}^{n} g\left(\mathbf{X}_i, t_j\right) = o_{P_n}(1),$$

under the sequence of local alternatives. This concludes the proof.