



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

6 June 2018

# An Improved Bootstrap Test for Restricted Stochastic Dominance

Thomas M. Lok\* and Rami V. Tabri†

June 6, 2018

## Abstract

Bootstrap Testing for restricted stochastic dominance of a pre-specified order between two distributions is of interest in many areas of economics. This paper develops a new method for improving the performance of such tests that employ a moment selection procedure: tilting the empirical distribution in the moment selection procedure. We propose that the amount of tilting be chosen to maximize the empirical likelihood subject to the restrictions of the null hypothesis, which are a continuum of unconditional moment inequality conditions. 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 leading tests are noticeably less conservative than their non-modified counterparts and have improved power. Finally, an empirical example is discussed to illustrate the proposed method.

JEL Classification: C12;C13;C14

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

## 1 Introduction

Testing for stochastic dominance of one distribution function by another is a frequently encountered statistical inference topic in many areas of economics. For example, these tests are used to compare income distributions (Abadie, 2002, and Barrett and Donald, 2003), investment strategies (Linton et al., 2005), mental stress levels in health economics (Madden, 2009), water-conserving irrigation strategies (Harris and Mapp, 1986), and to determine whether collusion occurs in asymmetric first-price auctions (Aryal and Gabrielli, 2013). While stochastic ordering is ideal for representing simple and important information, for example, the income distribution in a country ‘changed’ from time  $t_1$  to time  $t_2$ , the methodological issues that arise in statistical inference are complex. As a consequence, methodological research to improve statistical inference has been, and

---

\*Sydney Children’s Hospitals Network, Sydney, New South Wales 2113, Australia.

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

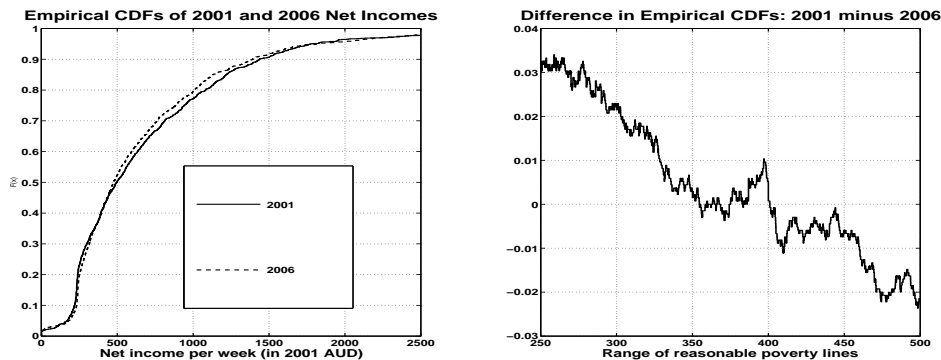


Figure 1: The left panel reports the empirical CDFs of weekly net incomes from households ‘at risk’ of retirement for the years 2001 and 2006 over the interval  $[0, 2500]$ . The right panel reports the difference between these CDFs over the range of plausible poverty lines.

continues to be, important. In this paper, we develop a new method that improves the performance of the leading existing methods of statistical inference involving stochastic dominance.

To motivate the topic, we consider a data example based on the Household, Income and Labour Dynamics in Australia (HILDA) panel survey for the periods 2001 and 2006. The left panel in Figure 1 shows that the empirical cumulative distribution functions (CDFs) of weekly net incomes in 2001 and 2006 for households ‘at risk’ of retirement (Barrett and Brzozowski, 2012). A question of fundamental importance for its policy implications is the following: *Does poverty increase at retirement?* Since there is uncertainty regarding the poverty line for income, in practice, one chooses an interval, say  $[a, b]$ , and the poverty line is assumed to lie in that interval; all statistical inferences must be carried out assuming that the poverty line is unknown and that it is somewhere in  $[a, b]$ . The foregoing substantive question on comparing the two income distributions, is usually formulated as a test of  $H_0 : F_{2001}(x) \leq F_{2006}(x), \forall x \in [a, b]$  against  $H_1 : \text{Not } H_0$ . If  $F_{2001}(x) \leq F_{2006}(x), \forall x \in [a, b]$ , we say that  $F_{2001}(\cdot)$  *stochastically dominates*  $F_{2006}(\cdot)$ , at first-order, over  $[a, b]$  (Foster and Shorrocks, 1988). For illustrative purposes, let us choose  $[a, b] = [250, 500]$  as the range of incomes of interest. The right panel in Figure 1 shows that the empirical CDFs corresponding to  $F_{2001}(\cdot)$  and  $F_{2006}(\cdot)$  cross one another; thus, there is some evidence against  $H_0$ . What we need is a sound statistical method to quantify the evidence of  $H_0$  properly.

Testing the aforementioned composite null hypothesis is challenging; the source of the challenge is that it is characterized by a continuum of inequality constraints, which complicates the evaluation of the asymptotic null distribution of conventional test statistics (e.g., the one-sided Kolmogorov-Smirnov). It turns out that bootstrapping can be used as an alternative. However, the simplest bootstrap procedure, namely, the canonical nonparametric bootstrap, is not valid in this test problem. It is invalid because the asymptotic null distributions depend discontinuously on  $F_{2001}(\cdot)$  and  $F_{2006}(\cdot)$  through the *contact set*  $\{x \in [250, 500] : F_{2001}(x) = F_{2006}(x)\}$ , the set of points in the domain of the CDFs where they coincide. This type of bootstrap failure was pointed out by Bickel and Freedman (1981) in a general context, and Andrews (2000) has demonstrated that it arises in testing problems when a parameter is on or near the boundary of its parameter space. An approach that restores the validity of the canonical nonparametric bootstrap for testing  $H_0$  uses an estimate of the contact set in the computation of the bootstrap test statistics (e.g., Linton

et al., 2010). Existing tests that follow this approach estimate the contact set using the empirical CDFs. As there is a continuum of inequality restrictions that define  $H_0$ , the discontinuous behavior of the asymptotic null distributions is quite elaborate. For this reason, an accurate estimator of the contact set is necessary to ensure good finite-sample properties for the tests. In this paper, we improve the accuracy of the contact set estimators in those tests by replacing the empirical CDFs with estimators that satisfy the restrictions of  $H_0$ , and show that this modification of those tests improves their performance.

Next, we introduce some notation to formulate the problem. Let  $P_A$  and  $P_B$  denote the marginal probability distributions from the bivariate distribution  $P$ . Let  $\text{supp}(P_K)$  denote the support of  $P_K$ , for  $K = A, B$ , and let  $[\underline{t}, \bar{t}]$  be a subset of  $\text{Interior}(\text{supp}(P_A) \cup \text{supp}(P_B))$ . We say that distribution  $P_B$  *stochastically dominates* distribution  $P_A$  at order  $s \in \mathbb{Z}_+$ , over the interval  $[\underline{t}, \bar{t}]$  if

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

A more general version of the testing problem considered in the foregoing example for the HILDA survey data is to test the null hypothesis that the continuum of inequalities in (1.1) holds.

Since the null hypothesis is composite, bootstrap approximations based on the least favorable case of the null hypothesis have been studied (e.g., Abadie, 2002, Barrett and Donald, 2003, and Horváth et al., 2006). Their approaches develop bootstrap critical values from the submodel of the null where the contact set is the interval  $[\underline{t}, \bar{t}]$ . While this approach yields valid testing procedures, it is quite conservative. Addressing this issue, several recent papers have proposed bootstrap tests whose limiting behavior mimics the discontinuity of the original test's limiting distribution (e.g., Linton et al., 2010 and Donald and Hsu, 2016). The approach of these papers is to use an estimator of the contact set in the bootstrap procedure, which yields tests that are less conservative and have higher limiting local power than their least favorable counterparts.

The usefulness of stochastic dominance tests that employ a contact set estimator rests upon accurate estimation of that set under the null hypothesis. The contact set estimators they use, are obtained without imposing the restrictions of the null hypothesis. The motivation for the method in our paper is based on our conjecture that the bootstrap tests of this null hypothesis would improve if the contact set used in computing the bootstrap test statistics was estimated after imposing the restrictions in the null hypothesis. In a way, we are claiming that the restrictions of the null hypothesis are additional information available to us, and we would expect to do better by using the additional information. This type of adjustment is a biased bootstrap technique, suggested by Hall and Presnell (1999); the stochastic dominance testing literature has not considered this procedure. In this paper, we show that our conjecture is in the right direction and our modification leads to a significant improvement for tests of stochastic dominance.

The contributions of our paper are summarized as follows:

- (a) It develops the constrained empirical likelihood estimator (Owen, 2001) of moments, where the constraints represent the restrictions of the null hypothesis (i.e., (1.1)), and proposes the following modification of stochastic dominance tests that use the contact set approach: replace the unconstrained estimator of the moments with this empirical likelihood estimator obtained under the null hypothesis;

- (b) it proves the validity of the proposed modification for the Linton et al. (2010) and Andrews and Shi (2017) tests, and that they are less conservative than their non-modified versions in large-samples; and
- (c) for a general class of  $n^{-1/2}$ -local alternatives and large enough sample sizes, it establishes the superiority of the modified tests over their non-modified versions in terms of local power, and the following criteria: *efficiency* (Pitman, 1948), *deficiency* (Hodges and Lehmann, 1970) and, *normalized deficiency* (Akahira, 1999).

The intuition behind these results is as follows. The proposed modification in (a) reformulates the bootstrap test statistics of the Linton et al. (2010) and Andrews and Shi (2017) tests in a data-dependent way that incorporates the statistical information contained in the constraints that characterize the null hypothesis, through a reduction of the parameter space. Consequently, under the null, the modified bootstrap distribution of the test statistic in either of these tests is more accurate than its non-modified counterpart as an estimator of the original test statistic’s sampling distribution, because the constraints imposed in the estimation of the contact set are correct. Under local alternative sequences, the constraints our method imposes are false for each  $n$ ; however, they are correct in the limit, which enables the modified tests to better detect  $n^{-1/2}$ -local deviations from the null.

Computing the solution of the constrained empirical likelihood optimization problem is one of the main technical contributions of this paper. 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  $t \in [\underline{t}, \bar{t}]$ —and a finite-dimensional choice variable. A common approach for solving semi-infinite programs is *discretization* (see, e.g., Shapiro, 2009). In such a method one selects a finite grid  $\mathcal{T}_n$ ,  $\mathcal{T}_n \subset [\underline{t}, \bar{t}]$ , and solves the constrained empirical likelihood optimization problem that imposes the constraints at each  $t$  in  $\mathcal{T}_n$ . We show that the difference between the proposed approximate solution and exact solution of the semi-infinite problem converges in probability to zero as the sample size tends to infinity. Further, we also show that the convergence is uniform over a collection of distributions and we characterize the collections of distributions.

The papers that employ a contact set estimator for bootstrap 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 for 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 may be restrictive for some studies. Hence, in this paper we omit a discussion of Donald and Hsu’s test. We adopt the statistical framework of LSW and defer a discussion of Andrews and Shi’s testing procedure to Appendix D, for ease of exposition.

We describe collections of population distributions on which the LSW and Andrews and Shi tests are asymptotically equivalent to their corresponding modified versions, at first-order. However, they differ in finite-samples. For this reason, we compare the tests according to the criteria described in (c) above. The comparisons using deficiency and normalized deficiency are more refined than efficiency because they discriminate tests on the basis of higher-order asymptotics (e.g., Albers, 1975). In the Supplementary Appendix, Section B illustrates the advantages of our general approach in terms of deficiency and normalized deficiency in the context of a simple exam-

ple: testing on a bivariate Gaussian mean. The example creates a new avenue for future research on the comparison of testing procedures for parameters defined by moment inequality conditions.

We report Monte Carlo simulation results in Section 6 that compare the LSW and Andrews and Shi tests with their respective modified versions. Overall, the simulation results corroborate the theoretical superior performance of our method, and that the differences in their relative performance may be substantial.

## 1.1 Related Literature

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

The literature on constrained estimation of CDFs subject to order restrictions is related to this paper. Several estimators have been proposed in that literature (e.g., Brunk et al., 1966, Lo, 1987, El-Barmi and Mukerjee, 2005, and Davidov and Herman, 2012). The procedure of Brunk et al. (1966) is closest to the estimation procedure in this paper. They developed the empirical likelihood estimator of two CDFs subject to a first-order stochastic dominance constraint, and prove its pointwise consistency when the constraints hold in the population. A key assumption in their method is that the random samples are drawn from statistically independent CDFs. By contrast, the proposed estimation procedure (i) allows for general dependence between the population CDFs, (ii) applies to stochastic dominance constraints higher than the first-order, and (iii) is uniformly consistent over a collection of CDFs when the constraints hold in the population.

Accommodating general dependence between the population CDFs is necessary in many empirical settings; for example, where income distributions are compared over time, as in the illustration described previously. Stochastic dominance orderings greater than one arise in practical situations; for example, when comparing income distributions using any poverty index from the Foster et al. (1984) class of poverty indices. Each ordering of the CDFs utilizing a poverty index from that class is in a one-to-one correspondence with their ranking according to a particular restricted stochastic dominance relation (Foster and Shorrocks, 1988). Furthermore, uniform consistency of the proposed estimator is particularly crucial for the test problem in this paper, because the asymptotic null distribution of conventional test statistics are discontinuous functions of the population CDFs.

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 that the practitioner must choose. By contrast, the constrained estimation method this paper proposes is for distributions and uses a data-driven grid of points.

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 Alvarez-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.

## 1.2 Organization of The Paper

We organize the paper as follows. Section 2 introduces the statistical framework of LSW. Section 3 introduces the constrained empirical likelihood optimization problem, the data-driven discretization method and its asymptotic properties, and the proposed contact set estimator. Section 4 presents the asymptotic null properties of the proposed contact set estimator and modified LSW test. Section 5 presents the asymptotic power properties of the modified LSW test. Section 6 reports the results of Monte Carlo simulations. Section 7 continues the empirical illustration described above, and Section 8 concludes.

## 2 Setup

### 2.1 Statistical Model, Null Hypothesis and a Property of the Moment Functions

Let  $P_0$  denote the “true” distribution of the random vector  $\mathbf{X} = [X^A, X^B]$ . For ease of exposition, let  $\mathcal{G}_s = \{\mathbf{x} \mapsto g(\mathbf{x}; t), t \in [\underline{t}, \bar{t}]\}$  denote the set of moment functions in (1.1), where

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

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

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

Let  $\mathcal{P}$  denote the set of all potential continuous distributions of  $\mathbf{X}$  that satisfies the following assumption.

**Assumption 2.1.** (i)  $[\underline{t}, \bar{t}] \subset \text{Interior}(\text{supp}(P_A) \cup \text{supp}(P_B))$ ; (ii)  $\text{supp}(P) \subseteq \mathbb{R}^2$ ; (iii)  $\{\mathbf{X}_i\}_{i=1}^n$  is i.i.d.  $P$ , and (iv)  $\sup_{P \in \mathcal{P}} E_P [|X^K|^{2((s-1) \vee 1) + \delta}] < +\infty$  for  $K = A, B$ , and for some  $\delta > 0$ .

Define  $\mathcal{P}_0 = \{P \in \mathcal{P} : E_P [g(\mathbf{X}; t)] \leq 0 \quad \forall t \in [\underline{t}, \bar{t}]\}$ . 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. Although the term ‘size’ and ‘similar’ relating to a test are quite familiar, it is convenient to recall the following asymptotic versions (for e.g., see Definition 1 of LSW).

**Definition 2.1.** Suppose that  $\Omega \subseteq \mathcal{P}_0$  is the model of the null hypothesis. (i) A test  $\tau$  with a nominal level  $\alpha$  is said to have an *asymptotically exact size* if there exists a nonempty subset  $\Omega' \subset \Omega$  such that:

$$\limsup_{n \rightarrow +\infty} \sup_{P \in \Omega} E_P \tau \leq \alpha, \quad \text{and} \tag{2.1}$$

$$\limsup_{n \rightarrow +\infty} \sup_{P \in \Omega'} |E_P \tau - \alpha| = 0. \tag{2.2}$$

(ii) When a test  $\tau$  satisfies (2.2), we say that the test is *asymptotically similar* on  $\Omega'$ .

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 among the elements of  $\{g(\mathbf{X}; t) : t \in \Delta(P)\}$ , where  $\Delta(P) = \{t \in [\underline{t}, \bar{t}] : E_P[g(\mathbf{X}; t)] = 0\}$  is the contact set with respect to  $P \in \mathcal{P}$ . The non-negative covariance structure is due to the following property of  $\mathcal{G}_s$ .

**Property 1.** For each  $s \in \mathbb{Z}_+$ , the class of functions  $\mathcal{G}_s$  satisfies the following property. For each  $\mathbf{x} \in \mathbb{R}^2$ , either  $g(\mathbf{x}; t) \leq 0 \forall t \in [\underline{t}, \bar{t}]$  or  $g(\mathbf{x}; t) \geq 0 \forall t \in [\underline{t}, \bar{t}]$ .

This property states that the sign of the moment functions  $g$  is determined solely by its first argument – the data  $\mathbf{x}$ . 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  $P \in \mathcal{P}$ ,  $\text{Cov}_P(g(\mathbf{X}; t_1), g(\mathbf{X}; t_2)) = E_P[g(\mathbf{X}; t_1)g(\mathbf{X}; t_2)] \geq 0 \forall t_1, t_2 \in \Delta(P)$ .

## 2.2 Test Statistic, Asymptotic Theory and LSW Bootstrap Procedure

The test statistic LSW use is of the Cramér-von-Mises type. In the setting of this paper it is given by  $\hat{T}_n = n \int_{\underline{t}}^{\bar{t}} \max\{E_{\hat{P}_n}[g(\mathbf{X}; t)], 0\}^2 dt$ , 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  $\mathcal{G}_s$  for each  $s \in \mathbb{Z}_+$ :

$$\hat{T}_n \xrightarrow{d} \begin{cases} \int_{\Delta(P)} \max\{\nu(t), 0\}^2 dt, & \text{if } P \in \mathcal{P}_{00}, \\ 0, & \text{if } P \in \mathcal{P}_0 - \mathcal{P}_{00}, \end{cases} \quad (2.3)$$

where  $\mathcal{P}_{00} = \{P \in \mathcal{P}_0 : \int_{\Delta(P)} dt > 0\}$ , and  $\nu(\cdot)$  is a zero-mean Gaussian process on  $[\underline{t}, \bar{t}]$  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. A consequence of the discontinuity is that it invalidates the use of the canonical nonparametric bootstrap implementation of  $T_n$  for testing  $H_0$ . For this reason, LSW propose a bootstrap testing procedure that uses a contact set estimator. Their contact set estimator is defined as

$$\hat{\Delta}_n = \{t \in [\underline{t}, \bar{t}] : |E_{\hat{P}_n}[g(\mathbf{X}; t)]| < r_n\}, \quad (2.4)$$

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

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  $n$ ,  $\{\mathbf{X}_{i,l}^*\}_{i=1}^n$ , for  $l = 1, \dots, B$ , using resampling with replacement from  $\hat{P}_n$ . Then for each bootstrap sample, compute the bootstrap test statistic as follows:

$$\hat{T}_{n,l}^* = \begin{cases} \int_{\underline{t}}^{\bar{t}} (\max\{\mathbb{G}_{n,l}, 0\})^2 dt, & \text{if } \int_{\hat{\Delta}_n} dt = 0, \\ \int_{\hat{\Delta}_n} (\max\{\mathbb{G}_{n,l}, 0\})^2 dt, & \text{if } \int_{\hat{\Delta}_n} dt > 0, \end{cases} \quad (2.5)$$



where

$$\mathbb{G}_{n,l} = \frac{1}{\sqrt{n}} \sum_{i=1}^n [g(\mathbf{X}_{i,l}^*; t) - E_{\hat{P}_n} [g(\mathbf{X}; t)]], \quad l = 1, \dots, B,$$

is the bootstrap empirical process, and  $\hat{\Delta}_n$  as defined in (2.4). After that, compute the approximate bootstrap p-value  $\hat{\Upsilon}_B = B^{-1} \sum_{l=1}^B 1[\hat{T}_{n,l}^* \geq \hat{T}_n]$ . 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 (2.4) 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 benefit of using the proposed modification is that it reformulates the LSW bootstrap test statistics in a data-dependent way that incorporates the statistical information contained in the constraints of the null hypothesis. Under the null, our approach imposes restrictions that are correct, which yields a more accurate estimator of the contact set; in consequence, the bootstrap distribution under our method is more reliable than the distribution of  $\hat{T}_{n,l}^*$  as an estimator of the sampling distribution of  $T_n$ . Under the alternative, the constraints our method imposes are false but enables the modified test to be better than the LSW test at detecting that the alternative hypothesis holds. The next section introduces the empirical likelihood estimator procedure and the contact set estimator based on it.

### 3 Empirical Likelihood and Contact Set Estimation

Consider the following constrained empirical likelihood optimization problem:

$$\tilde{\mathbf{p}} = \arg \max \left\{ \sum_{i=1}^n \log(p_i); \sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq 0 \forall t \in [\underline{t}, \bar{t}], \sum_{i=1}^n p_i = 1, p_i \geq 0 \forall i \right\}, \quad (3.1)$$

where  $\tilde{\mathbf{p}} = [\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_n] \in \mathbb{R}^n$ . 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 (3.1) is, in fact, a finite programming problem: there is a finite subset of the constraints  $\sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq 0 \forall t \in [\underline{t}, \bar{t}]$  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 (3.1). 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  $\sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq 0 \forall t \in [\underline{t}, \bar{t}]$  whose index set,  $\mathcal{T}_n = \{t_{(j)}\}_{j=0}^N$ , comprises the order statistics of  $\{X_i^A, X_i^B\}_{i=1}^n \cap (\underline{t}, \bar{t}) \cup \{\underline{t}, \bar{t}\}$ , where  $t_{(0)} = \underline{t}$  and  $t_{(N)} = \bar{t}$ . The solution of the discretized SIP problem is defined

as

$$\hat{\mathbf{p}} = \arg \max \left\{ \sum_{i=1}^n \log(p_i); \sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq 0 \forall t \in \mathcal{T}_n, \sum_{i=1}^n p_i = 1, p_i \geq 0 \forall i \right\}. \quad (3.2)$$

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  $\mathcal{P}$  where the probability of the event  $\{\tilde{\mathbf{p}} \text{ and } \hat{\mathbf{p}} \text{ exist}\}$  tends to unity exponentially fast and with uniformity. The uniformity in the convergence is so that the same minimal sample size controls the probability of that event for all elements in the particular subset of  $\mathcal{P}$ , which is particularly important because the test statistic exhibits a pointwise discontinuous limiting behavior.

**Definition 3.1.** 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  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$  define the sets of probabilities

$$\mathcal{P}_1(c_1) = \left\{ P \in \mathcal{P} : P \left[ \sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}; t) < 0 \right] \geq c_1 \right\} \quad \text{and} \quad (3.3)$$

$$\mathcal{P}_2(c_2) = \left\{ P \in \mathcal{P} : \inf_{(x^A, x^B) \in [\underline{t}, \bar{t}] \times [\underline{t}, \bar{t}]} f(x^A, x^B) \geq c_2 \right\}. \quad (3.4)$$

In Appendix C.1.1, we show that Part (i) of Assumption 2.1 implies that  $P[\sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}; t) < 0] > 0$  for each  $P \in \mathcal{P}$ . Therefore, 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  $\{\sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}; t) < 0\}$ . 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  $[\underline{t}, \bar{t}] \times [\underline{t}, \bar{t}]$ . This condition begets the uniform convergence of the sequence  $\{\mathcal{T}_n\}_{n \geq 1}$  to the interval  $[\underline{t}, \bar{t}]$  in the Hausdorff metric.

The constants,  $c_1$  and  $c_2$ , appear in the base of the exponential rate of uniform convergence of  $P[\tilde{\mathbf{p}} \text{ and } \hat{\mathbf{p}} \text{ exist}]$  to unity; see Section F.1 in Appendix F for more details. Thus, particular numerical values of these constants yield explicit rates of uniform convergence. In practice, it is important to verify that there are values of these constants that are consistent with  $P_0$ , i.e., there exists values of  $c_1 \in (0, 1)$  and  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$  such that  $P_0 \in \mathcal{P}_1(c_1)$  and/or  $P_0 \in \mathcal{P}_2(c_2)$ . In Appendices C.1.1 and C.1.2 we discuss how to verify these conditions in using statistical tests.

We have the following result.

**Theorem 3.1.** Let  $\tilde{\mathbf{p}}$  and  $\hat{\mathbf{p}}$  be given by (3.1) and (3.2), respectively. Furthermore, for each  $s \in \mathbb{Z}_+$ ,  $c_1 \in (0, 1)$  and  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$ , let  $\mathcal{P}_1(c_1)$  and  $\mathcal{P}_2(c_2)$  be given by (3.3) and (3.4), respectively, in Definition 3.1. The following statements hold.

1. Let  $s = 1$ . For each  $c_1 \in (0, 1)$ ,  $\lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_1(c_1)} P[\hat{\mathbf{p}} = \tilde{\mathbf{p}}] = 1$ .
2. Let  $s > 1$ . For each  $c_1 \in (0, 1)$ ,  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$  and for each  $\epsilon > 0$ ,

$$\lim_{n \rightarrow +\infty} \sup_{P \in \mathcal{P}_1(c_1) \cap \mathcal{P}_2(c_2)} P[\|\hat{\mathbf{p}} - \tilde{\mathbf{p}}\| > \epsilon] = 0,$$

where  $\|\cdot\|$  denotes the Euclidean metric on  $\mathbb{R}^n$ .

*Proof.* See Appendix E.1. □

Part 1 of Theorem 3.1 indicates that for  $s = 1$  and large enough  $n$ , the vector of probabilities  $\tilde{\mathbf{p}}$  and  $\hat{\mathbf{p}}$  coincide with uniformity over sets of probabilities  $\mathcal{P}_1(c)$ . Part 2 of Theorem 3.1 establishes the uniform convergence of the error between  $\tilde{\mathbf{p}}$  and  $\hat{\mathbf{p}}$  to zero over sets of probabilities  $\mathcal{P}_1(c_1) \cap \mathcal{P}_2(c_2)$ , for any  $s > 1$ . The result of Theorem 3.1 is crucial for ensuring good finite-sample properties for the modified LSW test because the (pointwise) limiting distribution of the LSW test statistic under the null hypothesis exhibits various forms of discontinuity; see (2.3).

The computational cost of  $\hat{\mathbf{p}}$  is rather low, even when  $n$  is large. For brevity, we relegate to Appendix C.1.3 a discussion of this point and show how to improve the numerical accuracy of the optimization problem (3.2).

The contact set estimator this paper proposes is defined as

$$\hat{\Delta}_n = \{t \in [\underline{t}, \bar{t}] : |E_{\hat{P}_n} [g(\mathbf{X}; t)]| < r_n\}, \quad (3.5)$$

where  $\hat{P}_n = \sum_{i=1}^n \hat{p}_i \delta_{X_i}$  and  $\hat{\mathbf{p}}$  is given by (3.2). Since  $\hat{P}_n$  is obtained without imposing the restrictions of the null hypothesis, the resulting estimator  $E_{\hat{P}_n} [g(\mathbf{X}; \cdot)]$  of  $E_P [g(\mathbf{X}; \cdot)]$  does not necessarily satisfy the restrictions of the null hypothesis. By contrast, the definition of  $\hat{P}_n$  implies  $E_{\hat{P}_n} [g(\mathbf{X}; \cdot)]$  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  $\hat{\Delta}_n$ , and the asymptotic size properties of the proposed testing procedure which uses it instead of  $\hat{\Delta}_n$  in the LSW bootstrap procedure.

## 4 Asymptotic Null Properties

### 4.1 Behavior of The Contact Set Estimator

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

**Definition 4.1.** For each  $s \in \mathbb{Z}_+$ ,  $c_1 \in (0, 1)$ ,  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$  and  $c_3 \in (0, +\infty)$ , let

$$\mathcal{P}_0^s(c_1, c_2, c_3) = \begin{cases} \{P \in \mathcal{P}_0(c_1) : \inf_{t \in \Delta(P)} E_P [g^2(\mathbf{X}; t)] \geq c_3\}, & \text{if } s = 1 \\ \{P \in \mathcal{P}_0(c_1, c_2) : \inf_{t \in \Delta(P)} E_P [g^2(\mathbf{X}; t)] \geq c_3\}, & \text{if } s > 1, \end{cases}$$

where  $\mathcal{P}_0(c_1) = \mathcal{P}_0 \cap \mathcal{P}_1(c_1)$  and  $\mathcal{P}_0(c_1, c_2) = \mathcal{P}_0 \cap \mathcal{P}_1(c_1) \cap \mathcal{P}_2(c_2)$ .

Definition 4.1 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  $\hat{\Delta}_n$  when  $s > 1$  which does not arise in the case  $s = 1$ . The defining condition  $\inf_{t \in \Delta(P)} E_P [g^2(\mathbf{X}; t)] \geq c_3$  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 (3.2), 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 G. 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 4.1.** For each  $s \in \mathbb{Z}_+$ , let the contact set estimators  $\hat{\Delta}_n$  and  $\hat{\Delta}_n$  be given by (2.4) and (3.5), respectively, and  $\mathcal{P}_{000} = \{P \in \mathcal{P}_0 : \Delta(P) \neq \emptyset\}$ . For each  $s \in \mathbb{Z}_+$ ,  $c_1 \in (0, 1)$ ,  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$  and  $c_3 \in (0, +\infty)$ , let  $\mathcal{P}_0^s(c_1, c_2, c_3)$  be given by Definition 4.1, and the following statements hold.

1.  $\lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_0^s(c_1, c_2, c_3)} P \left[ E_{\hat{P}_n} [g(\mathbf{X}; t)] \leq E_{\hat{P}_n} [g(\mathbf{X}; t)] \quad \forall t \in [\underline{t}, \bar{t}] \right] = 1.$   
 Moreover,  $\{\hat{P}_n \neq \hat{P}_n\} \implies \{E_{\hat{P}_n} [g(\mathbf{X}; t)] < E_{\hat{P}_n} [g(\mathbf{X}; t)] \quad \forall t \in [\underline{t}, \bar{t}]\}.$
2.  $\lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_0^s(c_1, c_2, c_3)} P \left[ \Delta(P) \subseteq \hat{\Delta}_n \subseteq \hat{\Delta}_n \right] = 1.$
3.  $\lim_{n \rightarrow +\infty} \sup_{P \in \mathcal{P}_0^s(c_1, c_2, c_3)} P \left[ \hat{\Delta}_n \subsetneq \hat{\Delta}_n \right] = 0.$
4.  $\lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_0^s(c_1, c_2, c_3) \cap \mathcal{P}_{000}} P \left[ \hat{\Delta}_n \subsetneq \hat{\Delta}_n \right] \geq \frac{1}{2}.$
5.  $\forall \epsilon > 0, \lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}} P \left[ \hat{\Delta}_n \subseteq \{t \in [\underline{t}, \bar{t}] : |E_P [g(\mathbf{X}; t)]| \leq (1 + \epsilon)r_n\} \right] = 1.$

*Proof.* See Appendix E.2. □

**Remark 4.1.** Part 1 of Theorem 4.1 is the consequence of the marriage between Property 1 of the moment function and constrained empirical likelihood estimation introduced in Section 3. This result is used in the proofs of Parts 2 - 4 of the same theorem.

**Remark 4.2.** Parts 2 and 3 of Theorem 4.1 imply the LSW bootstrap test statistic  $\hat{T}_n^*$ , described in (2.5), 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  $\{\hat{\Delta}_n \subsetneq \hat{\Delta}_n\}$ . Part 4 of Theorem 4.1 indicates the probability of  $\{\hat{\Delta}_n \subsetneq \hat{\Delta}_n\}$  in large samples, with uniformity. This result is derived from Lemma G.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 2.1 is feasible.

**Remark 4.3.** The set  $\{t \in [\underline{t}, \bar{t}] : |E_P [g(\mathbf{X}; t)]| \leq (1 + \epsilon)r_n\}$  is an enlargement of  $\Delta(P)$  that shrinks to it as the sample size tends to infinity because  $\{(1 + \epsilon)r_n\}_{n \geq 1}$  is a null sequence for each  $\epsilon > 0$ . Therefore, Parts 2 and 5 of Theorem 4.1 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.

## 4.2 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  $\nu$  in (2.3) with respect to  $r$ -enlargements of the contact set  $B(r) = \{t \in [\underline{t}, \bar{t}] : |E_P[g(\mathbf{X}; t)]| \leq r\}$ . For ease of exposition, this condition is given by Definition C.1 in Appendix C.2.1 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 4.2.** (i) For each  $\epsilon > 0$ ,  $s \in \mathbb{Z}_+$ ,  $c_1 \in (0, 1)$ ,  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$  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  $\nu$  in (2.3) is *regular* on  $B_n$  in the sense of Definition C.1, for each  $n \geq 1$ , where

$$B_n = \begin{cases} B((1 - \epsilon)r_n), & \text{if } \int_{B((1+\epsilon)r_n)} dt > 0, \text{ and} \\ [\underline{t}, \bar{t}], & \text{if } \int_{B((1+\epsilon)r_n)} dt = 0; \end{cases} \quad (4.1)$$

(ii) Given  $\xi_n \rightarrow 0$ , 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  $n > 1/\epsilon$ ,  $\nu$  in (2.3) is *regular* on  $B(n^{-1/2}\xi_n)$  in the sense of Definition C.1,

$$\int_{B((1-\epsilon)r_n)} dt > 0 \quad \text{and} \quad \int_{B((1+\epsilon)r_n) - B(n^{-1/2}\xi_n)} dt \leq \xi_n. \quad (4.2)$$

Denote by  $\hat{\Upsilon}_B = B^{-1} \sum_{l=1}^B 1[\hat{T}_{n,l}^* \geq \hat{T}_n]$  the bootstrap p-value that uses the bootstrap test statistics  $\{\hat{T}_{n,l}^*\}_{l=1}^B$  computed as in (2.5) of the LSW procedure, but with  $\hat{\Delta}_n$  replaced by  $\hat{\Delta}_n$ .

**Theorem 4.2.** 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  $\epsilon > 0$ ,  $s \in \mathbb{Z}_+$ ,  $c_1 \in (0, 1)$ ,  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$  and  $c_3 \in (0, +\infty)$ ,  $\hat{\Upsilon}_B - \Upsilon_B \xrightarrow{P} 0$  conditional on  $\mathcal{A}_n$  uniformly in  $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$ .

*Proof.* See Appendix E.3. □

Theorem 4.2 establishes the asymptotic equivalence of the bootstrap p-values  $\hat{\Upsilon}_B$  and  $\Upsilon_B$ , uniformly over  $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$ . For each  $\epsilon > 0$ ,  $s \in \mathbb{Z}_+$ ,  $c_1 \in (0, 1)$ ,  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$  and  $c_3 \in (0, +\infty)$ , the LSW test has asymptotically exact size, in the sense of Definition 2.1, 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  $\epsilon > 0$ ,  $s \in \mathbb{Z}_+$ ,  $c_1 \in (0, 1)$ ,  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$  and  $c_3 \in (0, +\infty)$ ,

$$(i) \limsup_{n \rightarrow +\infty} \sup_{P \in \mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)} P[\hat{\Upsilon}_\infty \leq \alpha] \leq \alpha, \text{ and}$$

$$(ii) \limsup_{n \rightarrow +\infty} \sup_{P \in \mathcal{P}_{00}^s(c_1, c_2, c_3, \epsilon, \{\xi_n\})} |P[\hat{\Upsilon}_\infty \leq \alpha] - \alpha| = 0, \text{ for each decreasing sequence } \xi_n \rightarrow 0.$$

Therefore, Theorem 4.2 implies that these properties hold with  $\hat{\Upsilon}_\infty$  replaced by  $\Upsilon_\infty$ .

**Remark 4.4.** The subsets of  $\mathcal{P}_0$  that establish the validity of the testing procedure this paper proposes (i.e., Definition 4.1) 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 4.1 are closer to the LSW counterparts. For ease of exposition, Appendix C.2.2 elaborates on this point, and Appendices C.1.2 and C.2.3 discusses the choice of the constants in practice.

## 5 Asymptotic Power Properties

### 5.1 Test Consistency

This section establishes consistency of the proposed testing procedure.

**Theorem 5.1.** Given  $s \in \mathbb{Z}_+$ , suppose  $P \in \mathcal{P} - \mathcal{P}_0$  is such that  $\int_{\underline{t}}^{\bar{t}} \max \{E_P [g(\mathbf{X}; t)], 0\}^2 dt > 0$ . Then,  $\lim_{n \rightarrow +\infty} P[\hat{\Upsilon}_B \leq \alpha] = 1$ .

*Proof.* See Appendix E.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.

### 5.2 Power Against Local Alternatives

This section compares 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 4.2 as the model of the null hypothesis, where the limit of the local alternatives (in some topology on  $\mathcal{P}$ ), satisfies  $\lim_{n \rightarrow +\infty} P_n \in \mathcal{P}_{00}$ .

To that end, we introduce the following notation. Consider a sequence of probabilities under the alternative hypothesis,  $\{P_n\}_{n \geq 1} \subset \mathcal{P} - \mathcal{P}_0$ , such that

$$E_{P_n} [g(\mathbf{X}; t)] = H(t) + \delta(t)/\sqrt{n} \text{ and } \sigma_{P_n}^2(t) = E_{P_n} [g^2(\mathbf{X}; t)] - (E_{P_n} [g(\mathbf{X}; t)])^2, \quad (5.1)$$

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

**Assumption 5.1.** (i)  $\int_C dt > 0$ , where  $C = \{t \in [\underline{t}, \bar{t}] : H(t) = 0\}$ .  
(ii)  $\sup_{t \in [\underline{t}, \bar{t}]} H(t) \leq 0$ . (iii)  $\int_C \max \{\delta(t), 0\}^2 dt > 0$ . (iv)  $\inf_{t \in [\underline{t}, \bar{t}], n \in \mathbb{N}} \sigma_{P_n}^2(t) > 0$ .

Except for Part (iv), Assumption 5.1 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.

Given  $s \in \mathbb{Z}_+$ , define the collection

$$\mathbb{H} = \bigcup \left\{ \mathcal{P}_0^s(c_1, c_2, c_3, \epsilon) : [c_1, c_2, c_3, \epsilon] \in (0, 1) \times (0, (\bar{t} - \underline{t})^{-2}) \times \mathbb{R}_{++}^2 \right\}, \quad (5.2)$$

where  $\mathbb{R}_{++}^2 = (0, +\infty) \times (0, +\infty)$ . The set  $\mathbb{H}$  collects all submodels of  $\mathcal{P}_0$  that are characterized by Part (i) of Definition 4.2. Additionally, for each  $\Omega \in \mathbb{H}$ , define  $\mathcal{P}_1^s(\Omega)$  as the following set of local alternatives

$$\mathcal{P}_1^s(\Omega) = \left\{ \{P_n\}_{n \geq 1} \subset \mathcal{P} - \mathcal{P}_0 : \text{conditions (5.1) and Assumption 5.1 hold,} \right. \\ \left. \text{and } \lim_{n \rightarrow +\infty} P_n \in \mathcal{P}_{00} \cap \Omega \right\}. \quad (5.3)$$

### 5.2.1 Behavior of Contact Set Estimator Under Local Alternatives

We first present a result that characterizes the behavior of the contact set estimators under the collections of local alternative sequences (5.3).

**Theorem 5.2.** Let  $s \in \mathbb{Z}_+$  and  $\mathbb{H}$  be given by (5.2). For each  $\Omega \in \mathbb{H}$ , let  $\mathcal{P}_1^s(\Omega)$  be given by (5.3). Then, for each  $\Omega \in \mathbb{H}$ , the following limits hold for each  $\{P_n\}_{n \geq 1} \in \mathcal{P}_1^s(\Omega)$ :

1.  $\lim_{n \rightarrow +\infty} P_n [E_{\hat{P}_n} [g(\mathbf{X}; t)] \leq E_{\hat{P}_n} [g(\mathbf{X}; t)] \quad \forall t \in [\underline{t}, \bar{t}] = 1.$
2.  $\lim_{n \rightarrow +\infty} P_n [C \subseteq \hat{\Delta}_n \subseteq \hat{\Delta}_n] = 1.$
3.  $\lim_{n \rightarrow +\infty} P_n [\hat{\Delta}_n \subsetneq \hat{\Delta}_n] = 0.$
4.  $\lim_{n \rightarrow +\infty} P_n [\hat{\Delta}_n \subseteq \{t \in [\underline{t}, \bar{t}] : |E_{P_n} [g(\mathbf{X}; t)]| \leq (1 + \epsilon)r_n\}] = 1 \quad \forall \epsilon > 0.$

*Proof.* See Appendix E.5. □

The results of Theorem 5.2 are similar to those of Theorem 4.1 but under local alternatives. Parts 2 and 3 of Theorem 5.2 imply that  $\hat{T}_n^* \leq \hat{T}_n^*$  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.

### 5.2.2 Comparison of Local Asymptotic Power Functions

Next, we compare the local asymptotic power functions of the LSW test and its modified counterpart. We have the following corollary to Theorem 5.2.

**Corollary 5.1.** Let  $s \in \mathbb{Z}_+$  and suppose the conditions of Theorem 5.2 hold. Then, for each  $\Omega \in \mathbb{H}$ ,

1.  $\lim_{n \rightarrow +\infty} P_n [\hat{\Upsilon}_\infty \leq \alpha] = \lim_{n \rightarrow +\infty} P_n [\hat{\Upsilon}_\infty \leq \alpha] \quad \forall \{P_n\}_{n \geq 1} \in \mathcal{P}_1^s(\Omega).$

2. and large enough  $n$ ,

$$P_n \left[ \hat{\Upsilon}_\infty \leq \alpha \right] \geq P_n \left[ \hat{\Upsilon}_\infty \leq \alpha \right] \quad \forall \{P_n\}_{n \geq 1} \in \mathcal{P}_1^s(\Omega). \quad (5.4)$$

Furthermore, the inequalities in (5.4) are strict for each  $\{P_n\}_{n \geq 1} \in \mathcal{P}_1^s(\Omega)$  such that  $\exists N \in \mathbb{Z}_+$ , for which  $P_n[\hat{\Delta}_n \subsetneq \hat{\Delta}_n] > 0 \forall n \geq N$ .

*Proof.* See Appendix E.6. □

Part 1 of Corollary 5.1 establishes the equality of the limiting local power functions of the two testing procedures in testing problems where the null models are demarcated by Part (i) of Definition 4.2, and with a corresponding set of local alternative sequences defined by (5.3). In conjunction with Theorem 4.2, Part 1 of Corollary 5.1 indicates that the testing procedures are asymptotically equivalent. Part 2 of Corollary 5.1 shows the modified test has local power no less than that of the LSW test in large samples, and strictly higher local power along sequences of local alternatives for which the probability of the proposed contact set estimator being proper subset of its LSW counterpart, is positive.

The consequence of Part 2 of Corollary 5.1 is that the testing procedures may differ in large-samples. This occurs along local alternatives for which  $P_n[\hat{\Delta}_n \subsetneq \hat{\Delta}_n] > 0$  holds in large-samples. Thus, a comparison of the tests that stops at the result of Part 1 of Corollary 5.1, is misleading. The result of Part 2 of Corollary 5.1 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.

### 5.2.3 Efficiency, Deficiency, and Normalized Deficiency

This section studies the relative asymptotic performance of the tests using their minimal required sample sizes for achieving a predesignated magnitude of local power. For every sample size  $n$ , define

$$k_n = \min\{N \in \mathbb{Z}_+ : P_m[\hat{\Upsilon}_\infty \leq \alpha] \geq P_n[\hat{\Upsilon}_\infty \leq \alpha] \forall m \geq N\}. \quad (5.5)$$

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.

Consider the following collections of local alternatives. Given  $s \in \mathbb{Z}_+$ , for each  $\Omega \in \mathbb{H}$ , define

$$\mathcal{P}_{1,+}^s(\Omega) = \left\{ \{P_n\}_{n \geq 1} \in \mathcal{P}_1^s(\Omega) : \exists N \in \mathbb{Z}_+ \text{ such that } P_n[\hat{\Delta}_n \subsetneq \hat{\Delta}_n] > 0 \forall n \geq N, \text{ and } P_n[\hat{\Upsilon}_\infty \leq \alpha] \nearrow p_\infty \in (0, 1) \text{ as } n \rightarrow +\infty \right\}, \quad \text{and} \quad (5.6)$$

$$\mathcal{P}_{1,-}^s(\Omega) = \left\{ \{P_n\}_{n \geq 1} \in \mathcal{P}_1^s(\Omega) : \exists N \in \mathbb{Z}_+ \text{ such that } P_n[\hat{\Delta}_n \subsetneq \hat{\Delta}_n] > 0 \forall n \geq N, \text{ and } P_n[\hat{\Upsilon}_\infty \leq \alpha] \searrow p_\infty \in (0, 1) \text{ as } n \rightarrow +\infty \right\}. \quad (5.7)$$

Local alternatives in these collections must satisfy (5.4) with strict inequality when  $n$  is large enough. These collections of local alternatives distinguish the cases in which  $k_n$  exists and does not exist, in large samples.

We have the following result.



**Theorem 5.3.** Let  $s \in \mathbb{Z}_+$  and suppose the conditions of Theorem 5.2 hold. Furthermore, let  $k_n$  be defined as in (5.5). Then, for each  $\Omega \in \mathbb{H}$ ,

1.  $k_n$  exists for large enough  $n$  and  $k_n > n$ ,  $\forall \{P_n\}_{n \geq 1} \in \mathcal{P}_{1,+}^s(\Omega)$ , where  $\mathcal{P}_{1,+}^s(\Omega)$  is given by (5.6).
2.  $k_n$  does not exist for  $n$  large enough,  $\forall \{P_n\}_{n \geq 1} \in \mathcal{P}_{1,-}^s(\Omega)$ , where  $\mathcal{P}_{1,-}^s(\Omega)$  is given by (5.7).

*Proof.* See Appendix E.7. □

Theorem 5.3 distinguishes the cases when  $k_n$  exists and does not exist, in large samples. In light of the result in Part 2 of Corollary 5.1, Part 1 of Theorem 5.3 indicates the local power functions  $n \mapsto P_n[\hat{Y}_\infty \leq \alpha]$  and  $n \mapsto P_n[\hat{Y}_\infty \leq \alpha]$ , 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  $k_n > n$ . By contrast, Part 2 of Theorem 5.3 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  $k_n$ .

Next, we characterize the large sample behavior of efficiency, deficiency and normalized deficiency of the modified test relative to the LSW test. As already mentioned, these criteria are functions of  $k_n$ ; therefore, for each  $\Omega \in \mathbb{H}$  the comparison of the tests are based on the local alternatives given by  $\mathcal{P}_{1,+}^s(\Omega)$  in (5.6). Finite-sample efficiency and deficiency are defined as

$$e_n = k_n/n \quad \text{and} \quad d_n = k_n - n, \quad (5.8)$$

respectively, and they are related through the following equation  $e_n = 1 + d_n/n$ , for each  $n$ . Intuitively, these criteria quantify the ‘price’ of using the LSW test instead of its modified version, at sample size  $n$ .

Historically, the literature has focused on deriving lower bounds on

$$\liminf_{n \rightarrow +\infty} e_n \quad \text{and} \quad \liminf_{n \rightarrow +\infty} d_n,$$

that do not depend on the underlying population, because they compactly summarize useful information on the relative asymptotic performance of statistical procedures, when these limit infima exist. See, for example, Theorem 1 of Hodges and Lehmann (1956), which provides a lower bound on the asymptotic efficiency of Wilcoxon test relative to the t-test for testing against shift of a continuous distribution. In this paper, we derive lower bounds for these limit infima (when they exist) that hold for a collection of local alternatives, because they indicate how low asymptotic efficiency and deficiency of the LSW test, relative to its modified version, can actually drop.

More recently, Akahira (1999) provides examples of statistical problems in which asymptotic efficiency and deficiency fail to discriminate particular statistical procedures. This contingency occurs when the procedures in those problems satisfy

$$\lim_{n \rightarrow +\infty} e_n = 1 \quad \text{and} \quad \lim_{n \rightarrow +\infty} d_n = +\infty. \quad (5.9)$$

See, for example, Albers (1975) for results of this sort who compares linear rank tests and parametric tests for the symmetry problem. For such situations, Akahira (1999) also proposes that one

compares statistical procedures using  $q_n$ -normalized deficiency. This criterion arises from the fact that the limits in (5.9) imply  $k_n$  has the following expansion

$$k_n = n + \ell q_n + o(q_n), \quad (5.10)$$

where  $\ell > 0$ , and  $\{q_n\}_{n \geq 1}$  is a sequence of positive numbers such that  $\lim_{n \rightarrow +\infty} q_n = +\infty$  and  $q_n = o(n)$ . From this expansion of  $k_n$ , Akahira suggests using the limit  $\lim_{n \rightarrow +\infty} (d_n/q_n) = \ell$  to quantify the relative performance of statistical procedures. Appendix B demonstrates numerically that the limits in (5.9) also arise in testing inequality restrictions on a bivariate Gaussian mean, when comparing appropriate versions of the LSW test and its modified counterpart for that testing problem. Moreover, we find that  $q_n = \sqrt{n}$  is the appropriate normalizing sequence and  $\ell \approx 19$ ; see Appendix B for more details. Consequently, it is reasonable to expect that asymptotic efficiency and deficiency will not necessarily separate the LSW test and its modified version for every sequence of local alternatives in  $\mathcal{P}_{1,+}^s(\Omega)$ , for each  $\Omega \in \mathbb{H}$ . Hence, for each  $\Omega \in \mathbb{H}$ , we characterize sub-collections of  $\mathcal{P}_{1,+}^s(\Omega)$  where the limits (5.9) hold, and compare the tests using  $q_n$ -normalized deficiency on these sub-collections.

The conditions of Part 1 of Theorem 5.3 are sufficient for the existence of  $k_n$ , which is because they imply that  $P_n[\hat{Y}_\infty \leq \alpha] < P_{k_n}[\hat{Y}_\infty \leq \alpha]$  holds for large enough  $n$ . We exploit this relationship between the local power functions to classify local alternatives according to the different limiting behaviors of  $e_n$  and  $d_n$ . The quantity that is key in doing so is

$$\psi_n = \frac{\left( P_{k_n}[\hat{Y}_\infty \leq \alpha] - P_n[\hat{Y}_\infty \leq \alpha] \right)}{\left( P_n[\hat{Y}_\infty \leq \alpha] - b_n \right)}, \quad (5.11)$$

where  $b_n$  is the intercept of the linear function that passes through the points

$$\left( n, P_n[\hat{Y}_\infty \leq \alpha] \right) \quad \text{and} \quad \left( k_n, P_{k_n}[\hat{Y}_\infty \leq \alpha] \right).$$

It is useful because  $d_n = n\psi_n$  holds whenever  $k_n$  exists, and it describes the relative concavity of the local power functions  $n \mapsto P_n[\hat{Y}_\infty \leq \alpha]$  and  $n \mapsto P_{k_n}[\hat{Y}_\infty \leq \alpha]$ .

We have the following result.

**Corollary 5.2.** Let  $s \in \mathbb{Z}_+$ , and suppose the conditions of Theorem 5.2 hold. Furthermore, for each  $n$ , let  $k_n$  be defined as in (5.5) and  $\psi_n$  by (5.11). Additionally, for each  $\Omega \in \mathbb{H}$ , let  $\mathcal{P}_{1,+}^s(\Omega)$  be given by (5.6). Then, for each  $\Omega \in \mathbb{H}$ ,

1.  $e_n > 1$  and  $d_n > 0$  for large  $n$ , and hence,  $\liminf_{n \rightarrow +\infty} e_n \geq 1$  and  $\liminf_{n \rightarrow +\infty} d_n \geq 0$ ,  $\forall \{P_n\}_{n \geq 1} \in \mathcal{P}_{1,+}^s(\Omega)$ .
2.  $\liminf_{n \rightarrow +\infty} e_n > 1$ ,  $\forall \{P_n\}_{n \geq 1} \in \mathcal{P}_{1,+}^s(\Omega)$  such that  $\liminf_{n \rightarrow +\infty} \psi_n > 0$ .
3.  $\lim_{n \rightarrow +\infty} e_n = 1$ ,  $\forall \{P_n\}_{n \geq 1} \in \mathcal{P}_{1,+}^s(\Omega)$  such that  $\limsup_{n \rightarrow +\infty} \psi_n = 0$ .
4.  $\lim_{n \rightarrow +\infty} e_n = 1$ ,  $\lim_{n \rightarrow +\infty} d_n = +\infty$ , and  $\lim_{n \rightarrow +\infty} (d_n/q_n) = \ell > 0$  where  $k_n = n + \ell q_n + o(q_n)$  for large enough  $n$ ,  $\forall \{P_n\}_{n \geq 1} \in \mathcal{P}_{1,+}^s(\Omega)$  such that  $\limsup_{n \rightarrow +\infty} \psi_n = 0$  and  $\liminf_{n \rightarrow +\infty} n\psi_n = +\infty$ ,

*Proof.* See Appendix E.8. □

Overall, the results of Corollary 5.2 provide strong evidence for the superior performance of the proposed testing procedure over the LSW test. The large sample and asymptotic bounds in Part 1 of Corollary 5.2 indicate that in large enough sample sizes the modified LSW test is superior to the LSW test according to efficiency and deficiency, for each sequence of local alternatives that satisfies Assumption 5.1 and the conditions of Part 1 of Theorem 5.3. Parts 2 and 3 give conditions under which the asymptotic bound on efficiency in Part 1 holds with strictly inequality and equality, respectively. Part 4 covers the case where asymptotic efficiency and deficiency fail to separate the tests. It establishes that, in large-samples,  $d_n \approx \ell_{q_n}$ . Thus, in this circumstance, the price for using the LSW test is increasing with the sample size at the rate  $\ell_{q_n}$ . This point is consistent with the fact that when comparing the tests along the sequences of local alternatives, the testing problems become progressively harder as more observations become available.

## 6 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. In contrast to the LSW test, the AS test uses a one-sided Kolmogorov-Smirnov test statistic in which the sample moments are Studentized. Furthermore, they Studentize the sample moments that arise in the construction of their contact set estimator. See Appendix D for a detailed discussion of the AS test and how the results in this paper on the comparison between the LSW test and its modified version carries over to the AS test and the modified AS test.

The experimental setup is the same as the one in Section 5 of LSW who focus on testing for first-order stochastic dominance. The simulation results indicate the modified versions of the tests are noticeably less conservative and have higher power in comparison to their non-modified counterparts. 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  $t \in [\underline{t}, \bar{t}]$ . 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  $t \in [\underline{t}, \bar{t}]$ , 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  $\bar{t} = 0.95$ . Finally, the number of Monte Carlo replications was set to 10000 in each simulation experiment, and the number of bootstrap replications was 499.

### 6.1 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

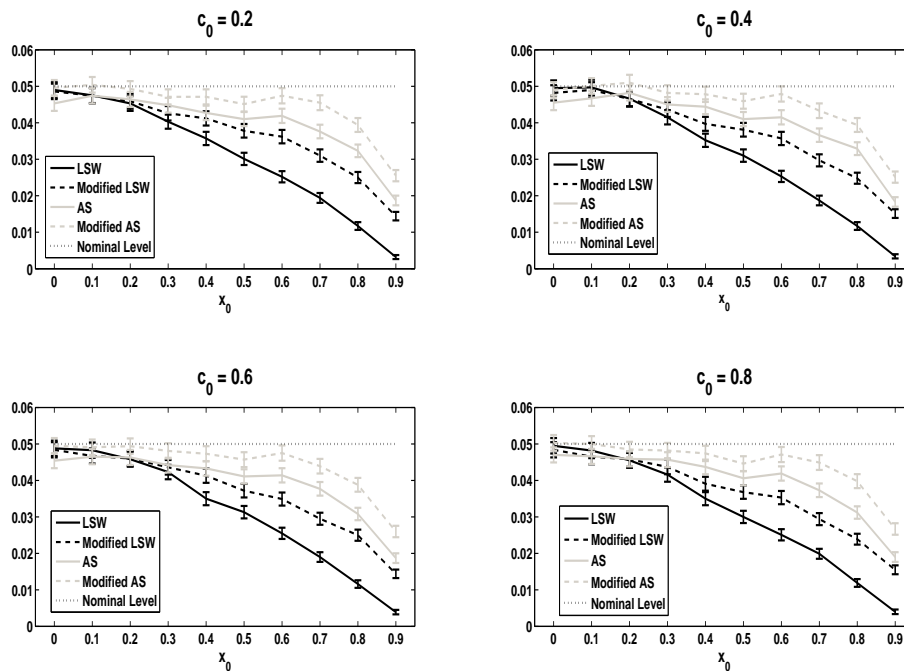


Figure 2: The empirical rejection probabilities under the null.

“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  $x_0 \in \{0, 0.1, 0.2, \dots, 0.9\}$  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 2 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.

## 6.2 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 config-

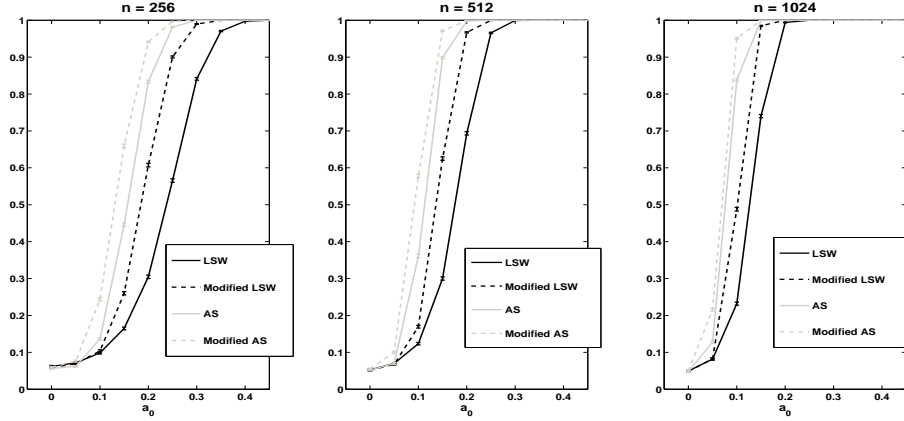


Figure 3: The empirical rejection probabilities under the alternative: directions of non-dominance.

uration 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  $a_0 \in \{0, 0.05, 0.1, 0.15, 0.2, \dots, 0.45\}$ . 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 3 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.

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.

## 7 Empirical Illustration

This section continues the discussion of the empirical illustration presented in Section 1. As an important and substantial change in income can occur at retirement, a central question for both academic researchers and policy makers is whether poverty increases at retirement. Recall that

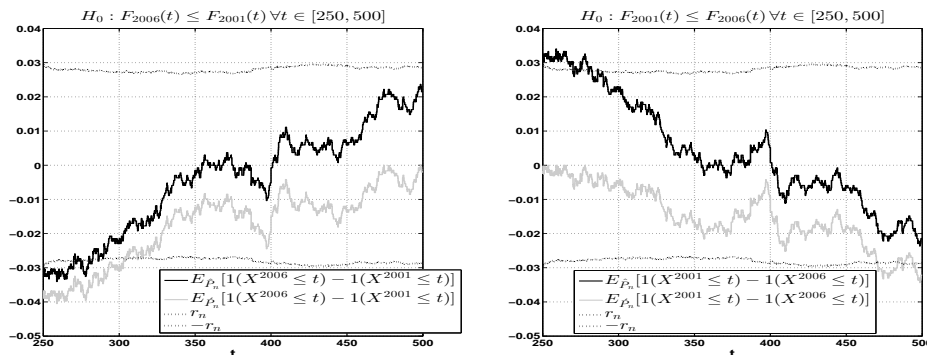


Figure 4: The left panel reports  $E_{\hat{P}_n}[g(\mathbf{X}, \cdot)]$  and  $E_{\hat{P}_n}[g(\mathbf{X}, \cdot)]$ , and the values of the localization parameter  $r_n(\cdot)$  for the test problem (7.1). The right panel reports the same quantities but for the test problem (7.2).

the left panel in Figure 1 depicts the empirical CDFs of weekly net incomes (in 2001 Australian dollars) from the HILDA panel survey for the years 2001 and 2006 for households ‘at risk’ of retiring. The data consists of  $n = 1351$  pairs on household net incomes.

In this illustration, the range of poverty lines is the interval  $[250, 500]$ . Thus, we focus on the following first-order restricted stochastic dominance test problems,

$$H_0 : F_{2006}(t) \leq F_{2001}(t) \forall t \in [250, 500] \quad \text{vs.} \quad H_1 : \text{Not } H_0, \quad \text{and} \quad (7.1)$$

$$H_0 : F_{2001}(t) \leq F_{2006}(t) \forall t \in [250, 500] \quad \text{vs.} \quad H_1 : \text{Not } H_0. \quad (7.2)$$

In terms of the notation of this paper, the test problem (7.1) corresponds to the setup in which populations  $X^A$  and  $X^B$  are  $X^{2001}$  and  $X^{2006}$ , respectively,  $\underline{t} = 250$ ,  $\bar{t} = 500$ , and  $s = 1$ . For the test problem (7.2), the setup is the same as that for the former problem except that the roles of  $X^A$  and  $X^B$  are now reversed. We apply the LSW and AS tests along with their modified versions to these test problems.

Figure 4 reports the sample-analog and constrained empirical likelihood estimates of the differences between CDFs for the two test problems. We set  $r_n(\cdot)$  as in the Monte Carlo experiments. As can be seen in the left panel of Figure 4, the event  $\{\hat{\Delta}_n \subsetneq \hat{\Delta}_n\}$  occurs. Thus, we expect the bootstrap p-values of the modified tests to be stochastically smaller than their respective non-modified counterparts, conditional on the data, for the test problem (7.1). In contrast to test problem (7.1), the right panel in Figure 4 indicates that this event does not occur for the test problem (7.2). Consequently, the bootstrap statistics of the modified versions are not necessarily stochastically larger than their respective modified counterparts, conditional on the data.

Table 1 reports the bootstrap p-values for the tests. The number of bootstrap simulations was set to  $B = 999$ , where the same random numbers were used in the implementation of the tests, so as to properly compare the outputs from the modified and non-modified tests. With the nominal level set to  $\alpha = 0.05$ , all of the tests do not reject  $H_0$  in (7.1), while all of them, except the LSW test, reject  $H_0$  in (7.2). Therefore, all of the tests, except the LSW test, indicate poverty—as measured by the *headcount ratio*—decreased from 2001 to 2006 in Australia for households ‘at risk’ of retiring.

Table 1: Bootstrap P-values

Procedure	Test problem (7.1)	Test problem (7.2)
AS	0.0761	0.01
Modified AS	0.0731	0.004
LSW	0.2142	0.0551
Modified LSW	0.1972	0.0470

## 8 Conclusion

While the main body of our paper focuses exclusively on the LSW test, the method we propose extends to the testing framework of AS. Appendix D outlines these results for the AS test.

The modification 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 we propose simplifies to solving a finite program: constrained empirical likelihood subject to a finite number of inequality constraints. A detailed theoretical analysis of this modification of their procedure goes beyond the intended scope of this paper. However, in Appendix B.5, we illustrate using a simple numerical example that our proposed modification can improve the finite-sample properties under the null hypothesis of such tests, which points to the potential usefulness of our method in other moment inequality testing problems.

The constrained empirical likelihood estimation methodology we developed for restricted stochastic dominance restrictions extends to other types of shape constraints that can be imposed through tilting the empirical distribution function. Examples of shape constraints that our methodology extends to are monotonicity and modality of a density function (e.g., Hall and Huang, 2001, 2002), and convexity (or concavity) of a regression function (e.g., Du et al., 2013). The key is that the shape constraints are linear in the probabilities,  $p_1, \dots, p_n$ .

## 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 is especially grateful for feedback from Mervyn Silvapulle. He also thanks Garry Barrett, Kevin Schnepel, Russell Toth, Andrew Wait, Peter Exterkate, Yi Sun, 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.
- Alvarez-Esteban, P. C., E. del Barrio, J. A. Cuesta-Albertos, and C. Matran (2017, 08). Models for the assessment of treatment improvement: The ideal and the feasible. *Statist. Sci.* 32(3), 469–485.
- 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). Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space. *Econometrica* 68(2), 399–405.
- Andrews, D. W. K. and G. Soares (2010). Inference for Parameters Defined by Moment Inequalities using Generalized Moment Selection. *Econometrica* 78(1), 119–157.
- Aryal, G. and F. M. Gabrielli (2013). Testing for collusion in asymmetric first-price auctions. *International Journal of Industrial Organization* 31, 26–35.
- Barrett, G. and M. Brzozowski (2012). Food expenditure and involuntary retirement: Resolving the retirement-consumption puzzle. *American Journal of Agricultural Economics* 94, 945–955.
- 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.
- Bickel, P. J. and D. A. Freedman (1981, 11). Some asymptotic theory for the bootstrap. *Ann. Statist.* 9(6), 1196–1217.
- Billingsley, P. (1995). *Probability and Measure* (Third ed.). New York: John Wiley & Sons.



- Brunk, H. D., W. E. Franck, D. L. Hanson, and R. V. Hogg (1966). Maximum likelihood estimation of the distributions of two stochastically ordered random variables. *Journal of the American Statistical Association* 61(316), 1067–1080.
- 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.
- Davidov, O. and A. Herman (2012). Ordinal dominance curve based inference for stochastically ordered distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74(5), 825–847.
- 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.
- El-Barmi, H. and H. Mukerjee (2005). Inferences under a stochastic ordering constraint: The k-sample case. *Journal of the American Statistical Association* 100(469), 252–261.
- Feller, W. (1968, Sep). On the berry-esseen theorem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 10(3), 261–268.
- Foster, J., J. Greer, and E. Thorbecke (1984). A class of decomposable poverty measures. *Econometrica* 52(3), pp. 761–766.
- 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.

- Harris, T. R. and H. P. Mapp (1986). A stochastic dominance comparison of water-conserving irrigation strategies. *American Journal of Agricultural Economics* 68(2), 298–305.
- Hodges, J. L. and E. L. Lehmann (1956). The efficiency of some nonparametric competitors of the  $t$ -test. *The Annals of Mathematical Statistics* 27(2), 324–335.
- 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., E. Maasoumi, and Y.-J. Whang (2005). Consistent testing for stochastic dominance under general sampling schemes. *The Review of Economic Studies* 72(3), 735–765.
- Linton, O., K. Song, and Y.-J. Whang (2010). An Improved Bootstrap Test for Stochastic Dominance. *Journal Of Econometrics* 154, 186–202.
- Lo, S. H. (1987). Estimation of Distribution Functions Under Order Restrictions. *Statistical Decisions* 5, 251–262.
- Madden, D. (2009). Mental stress in ireland, 1994-2000: a stochastic dominance approach. *Health Economics* 18(10), 1202–1217.
- 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.
- Shapiro, A. (2009). Semi-Infinite Programming, Duality, Discretization, and Optimality Conditions. *Optimization* 8(2), 133–161.
- Silvapulle, M. J. and P. K. Sen (2011). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. Wiley.
- 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 presents a motivating simple example of the method this paper proposes.
- Appendix C covers further discussions of points raised in the paper.
- Appendix D 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 E presents the proofs the results in the paper: Theorems 3.1, 4.1, 4.2, 5.1, 5.2, and Corollaries 5.1, 5.2.
- Appendix F presents technical lemmas used in the proof of Theorem 3.1.
- Appendix G presents technical lemmas used in the proofs of Theorems 4.1 and 4.2.
- Appendix H presents technical lemmas used in the proofs of Theorems 5.1 and 5.2.

## B Illustrative Example

This section illustrates the advantages of our general approach in the context of testing on a bivariate Gaussian mean. The hypothesis testing problem is an example of what Silvapulle and Sen (2011) refer to as a Type B testing problem in Section 3.2 of their book.

### B.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  $\Omega$  is the  $2 \times 2$  correlation matrix with correlation coefficient  $\rho$ . 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 B.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  $\rho_0$  as positive to mimic the inherent correlational structure between 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

$$H_0 : \mu_1 \leq 0 \text{ and } \mu_2 \leq 0 \quad \text{versus} \quad H_1 : \text{either } \mu_1 > 0 \text{ or } \mu_2 > 0 \text{ or both.} \quad (\text{B.1})$$

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

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  $j = 1, 2$ . Let  $F_{\mu, \Omega_0}(\cdot, \cdot)$  denote the cumulative distribution function (CDF) of  $N(\mu, \Omega_0)$ , where  $\Omega_0$  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:

$$\lim_{n \rightarrow +\infty} F_{\sqrt{n}\mu, \Omega_0}(u, u) = \begin{cases} F_{\mathbf{0}, \Omega_0}(u, u) & \text{if } \mu_1 = \mu_2 = 0 \\ \Phi(u) & \text{if } \mu_1 = 0, \mu_2 < 0 \\ \Phi(u) & \text{if } \mu_1 < 0, \mu_2 = 0, \\ 1, & \text{if } \mu_1, \mu_2 < 0, \end{cases} \quad \forall u \in \mathbb{R}, \quad (\text{B.2})$$

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.

## B.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  $\hat{\tau}_n = 1[T_n > \hat{c}]$ , 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)  $r_n \rightarrow 0$  and (ii)  $\sqrt{n}r_n \rightarrow +\infty$ , as  $n \rightarrow +\infty$ . The GMS function is the vector  $\hat{\varphi} = [\hat{\varphi}_1, \hat{\varphi}_2]$  whose components are defined as follows

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

The GMS critical value is defined as

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

where  $\alpha$  is the given nominal level.

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

$$\tilde{\mu} = \arg \max \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{X}_i - \mu)' \Omega^{-1} (\mathbf{X}_i - \mu); \mu \in \mathbb{R}_-^2 \right\}, \quad (\text{B.3})$$

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 (B.1). Using the Karush-Kuhn-Tucker conditions for the constrained optimization problem in (B.3), it is a straightforward task to deduce that

$$\tilde{\mu} = [\tilde{\mu}_1, \tilde{\mu}_2] = \begin{cases} [\hat{\mu}_1, \hat{\mu}_2] & \text{if } \hat{\mu}_1, \hat{\mu}_2 \leq 0 \\ [\hat{\mu}_1 - \rho_0 \hat{\mu}_2, 0] & \text{if } \hat{\mu}_1 \leq 0, \hat{\mu}_2 > 0 \\ [0, \hat{\mu}_2 - \rho_0 \hat{\mu}_1] & \text{if } \hat{\mu}_1 > 0, \hat{\mu}_2 \leq 0 \\ [0, 0] & \text{if } \hat{\mu}_1, \hat{\mu}_2 > 0. \end{cases} \quad (\text{B.4})$$

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

### B.3 Relative Performance of Tests Under The Null and $\sqrt{n}$ Local Alternatives

Setting the nominal level  $\alpha = 0.05$  and the localization parameter  $r_n = \sqrt{\log(n)/n}$ , the following numerical results establish the test's exact sizes are approximately equal to 0.05 for sample sizes  $n = 100, 101, 102, \dots, 10000$ . 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.

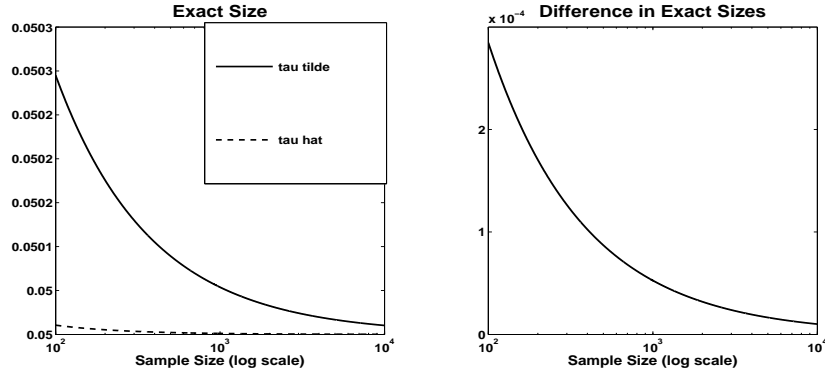


Figure 5: The first panel reports the exact sizes of the tests  $\hat{\tau}_n$  and  $\tilde{\tau}_n$ . The second panel reports the difference  $\sup_{P \in \mathcal{M}_0} E_P \hat{\tau}_n - \sup_{P \in \mathcal{M}_0} E_P \tilde{\tau}_n$ .

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 reasonable to compare the tests' rejection probabilities along local alternatives for sample sizes  $n \geq 100$  without adjusting them.

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

$$H_n : \mu_{1n} = -0.1 \quad \text{and} \quad \mu_{2n} = 2/\sqrt{n}, \quad \forall n. \quad (\text{B.5})$$

Figure 6 reports the local powers of the tests for  $n = 100, 101, 102, \dots, 7000$ . These numerical results indicate that the local powers satisfy

$$\lim_{n \rightarrow +\infty} E_{P_n} \hat{\tau}_n = \lim_{n \rightarrow +\infty} E_{P_n} \tilde{\tau}_n = 1 - \Phi(c - 2) \approx 0.6388,$$

where  $c = \Phi^{-1}(1 - \alpha)$  is the common probability limit of  $\hat{c}$  and  $\tilde{c}$ , and  $E_{P_n} \hat{\tau}_n < E_{P_n} \tilde{\tau}_n \forall n$ .

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  $+\infty$ ,

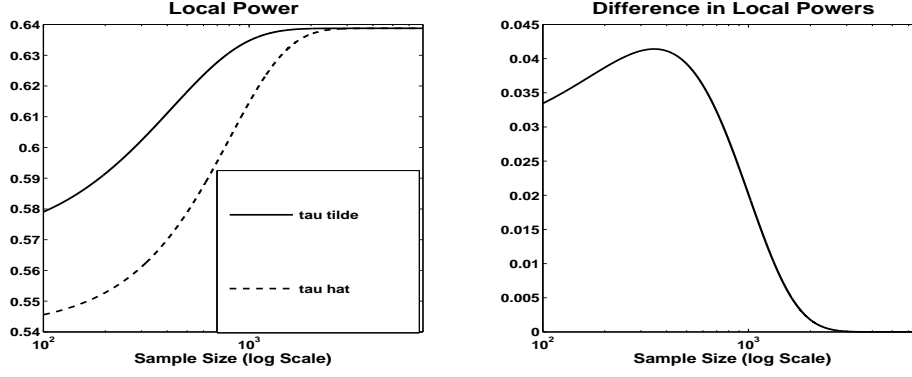


Figure 6: The left panel plots the local powers  $E_{P_n} \hat{\tau}_n$  and  $E_{P_n} \tilde{\tau}_n$ . The right panel plots the difference  $E_{P_n} \tilde{\tau}_n - E_{P_n} \hat{\tau}_n$ .

Table 2: Relative Performance Criteria

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

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 2 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 7 reports the numerical values of the criteria described Table 2. 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

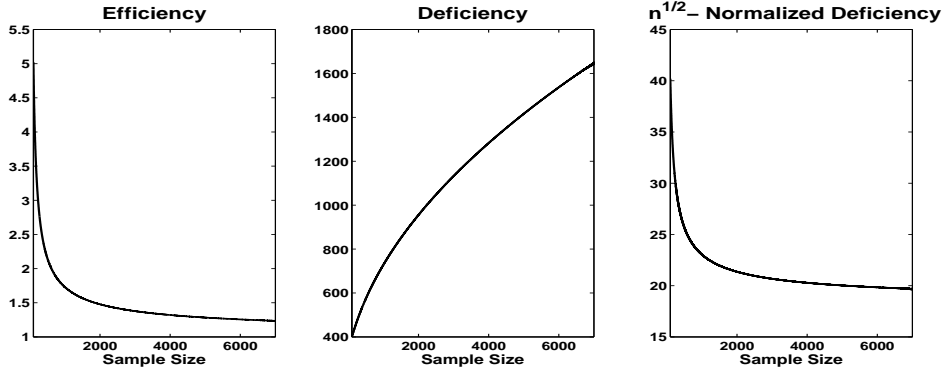


Figure 7: The panels, from left to right, plot the efficiency  $k_n/n$ , deficiency  $d_n$ , and normalized deficiency  $d_n/\sqrt{n}$  of  $\hat{\tau}_n$  with respect to  $\tilde{\tau}_n$ , respectively.

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  $\lim_{n \rightarrow +\infty} (d_n/\sqrt{n}) \approx 19$ .

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  $n \rightarrow +\infty$ . This result is not unusual because the local alternatives (B.5) 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.

Now consider local alternatives  $\{P_n\}_n \subset \mathcal{M} - \mathcal{M}_0$  where  $P_n$  satisfies the hypothesis

$$H_n : \mu_{1n} = -2/\sqrt{n} \quad \text{and} \quad \mu_{2n} = 2/\sqrt{n}, \quad \forall n = 100, 101, \dots, 50000. \quad (\text{B.6})$$

Figure 8 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, and that the local powers are decreasing with the sample size  $n$ ; therefore,  $k_n$  does not exist. 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 2.

## B.4 Synthesis

The marriage of constrained estimation and Condition (iii) of Assumption B.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.4), this marriage yields

$$\tilde{\mu}_j \leq \hat{\mu}_j \quad \text{for} \quad j = 1, 2, \quad (\text{B.7})$$

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 (B.5). 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

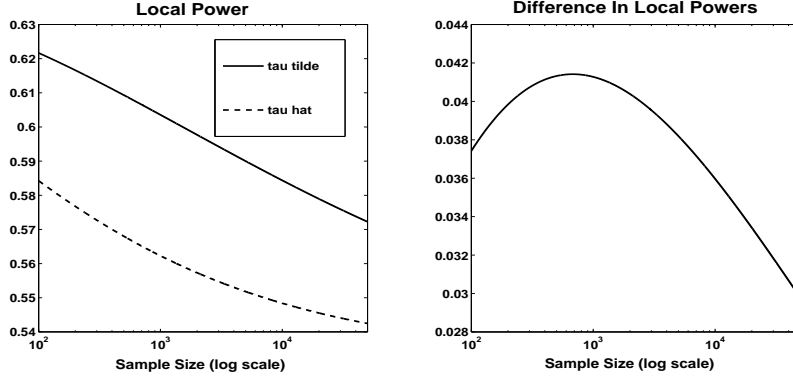


Figure 8: The left panel plots the local powers  $E_{P_n} \hat{\tau}_n$  and  $E_{P_n} \tilde{\tau}_n$ . The right panel plots the difference  $E_{P_n} \tilde{\tau}_n - E_{P_n} \hat{\tau}_n$ .

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  $\hat{\tau}_n$  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.

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 (B.2), and can have a profound effect on the relative performance of the Linton et al. (2010) and Andrews and Shi (2017) tests and their respective modified versions.

### B.5 Negative Correlation: $\rho < 0$ .

This section illustrates the finite-sample behavior of the tests,  $\hat{\tau}_n$  and  $\tilde{\tau}_n$ , under  $H_0$ , when the correlation coefficient in Assumption B.1 is negative. Note that negative correlations between estimators of the moments in stochastic dominance constraints do not arise; see Property 1 in Section 2. However, negative correlations can arise in other types of moment inequality models (e.g., Andrews and Soares, 2010). Thus, the purpose of this section is to point at the potential of the proposed modification to better control the type I error rate in finite-samples in comparison to its plain-vanilla counterpart in such models.

Figure 9 depicts the exact sizes of the tests  $\hat{\tau}_n$  and  $\tilde{\tau}_n$  when  $\rho = -0.9$ , where the nominal level is  $\alpha = 0.05$ . The exact size of  $\tilde{\tau}_n$  is 0.05 for sample sizes  $n \geq 100$ , while its counterpart for  $\hat{\tau}_n$  declines very slowly to the nominal level. In consequence, the test  $\hat{\tau}_n$  over-rejects under  $H_0$  for  $n \geq 100$ . While the magnitude of the over-rejection by  $\hat{\tau}_n$  is not large, this may not be the case when there is a larger number of inequality constraints. Furthermore, the test  $\tilde{\tau}_n$  does not require a size-correction. These features of the illustration renders our method as potentially useful in other moment inequality models as it circumvents the use of size-correction tuning-parameters because it can better control the Type I error rate in finite-samples.



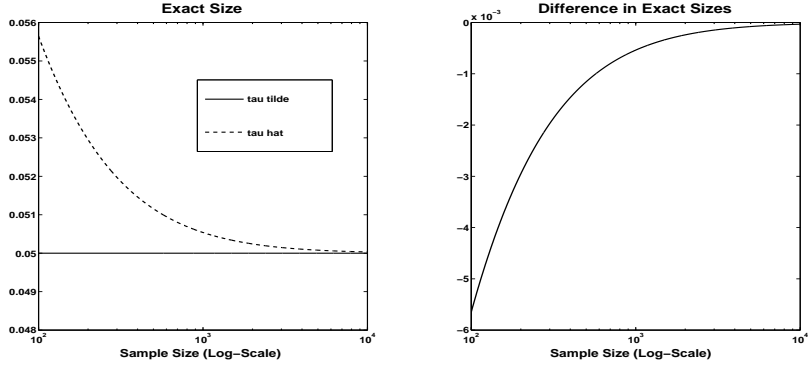


Figure 9: The first panel reports the exact sizes of the tests  $\hat{\tau}_n$  and  $\tilde{\tau}_n$ . The second panel reports the difference  $\sup_{P \in \mathcal{M}_0} E_P \hat{\tau}_n - \sup_{P \in \mathcal{M}_0} E_P \tilde{\tau}_n$ .

## C Further Discussion

### C.1 Section 3

#### C.1.1 Definition 3.1

Definition 3.1 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  $\{\sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}; t) < 0\}$ . It is given by

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

Hence, this event is a tail event, and Part (i) of Assumption 2.1 implies

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

Consequently, the set  $\mathcal{P}_1(c_1)$  excludes distributions that place probability less than  $c_1$  on the respective tail events in (C.1) 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  $[\underline{t}, \bar{t}]$ , is uniform. These subsets exclude distributions whose joint densities are arbitrary close to zero on the box  $[\underline{t}, \bar{t}] \times [\underline{t}, \bar{t}]$ . 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  $[\underline{t}, \bar{t}]$ . For ease of exposition, these intermediate technical results are relegated to Appendix F.

### C.1.2 Verifying The Conditions of $\mathcal{P}_1(c_1)$ and $\mathcal{P}_2(c_2)$ .

The conditions of Definition 3.1 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

$$H_0 : P_0 \left[ \sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}; t) < 0 \right] \geq c_1 \quad \text{versus} \quad H_1 : P_0 \left[ \sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}; t) < 0 \right] < c_1. \quad (\text{C.2})$$

A t-test based on the statistic  $\frac{1}{n} \sum_{i=1}^n 1 [\sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}_i; t) < 0]$  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  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$ , the testing problem is

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

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.

### C.1.3 Computational Aspects

The data-driven discretization scheme, introduced in Section 3, 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 (3.2) are linear in the choice variables  $p_1, \dots, p_n$ . 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  $\hat{\mathbf{p}}$ , given by (3.2), satisfies the following property

$$E_{\hat{P}_n} [g(\mathbf{X}; t)] \leq E_{\hat{P}_n} [g(\mathbf{X}; t)] \quad \forall t \in [\underline{t}, \bar{t}], \quad (\text{C.3})$$

when it exists. Therefore, the inequalities (C.3) can be used to increase the numerical accuracy and speed of computation in the optimization problem (3.2) by replacing the constraints  $\sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq 0 \quad \forall t \in \mathcal{T}_n$ , with the following:

$$\sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq 0 \quad \forall t \in \mathcal{T}_n \cap \{t \in [\underline{t}, \bar{t}] : E_{\hat{P}_n} [g(\mathbf{X}; t)] \geq 0\} \quad \text{and} \quad (\text{C.4})$$

$$\sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq E_{\hat{P}_n} [g(\mathbf{X}; t)] \quad \forall t \in \mathcal{T}_n \cap \{t \in [\underline{t}, \bar{t}] : E_{\hat{P}_n} [g(\mathbf{X}; t)] < 0\}, \quad (\text{C.5})$$

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

## C.2 Section 4

### C.2.1 Regularity On The Asymptotic Gaussian Process $\nu$

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  $\nu$  in (2.3) which is given by Definition 2 of their paper. In the context of the present work, this condition is the following.

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

$$P \left[ \int_A \max \{ \nu(t), 0 \}^2 dt < \bar{\epsilon} \right] < 1 - \alpha \quad (\text{C.6})$$

and for any  $c > 0$ ,

$$\limsup_{\eta \downarrow 0} \sup_{P \in \mathcal{P}_0} P \left[ \left| \int_A \max \{ \nu(t), 0 \}^2 dt - c \right| \leq \eta \right] = 0. \quad (\text{C.7})$$

Condition (C.6) 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  $\alpha$ . To understand this condition, note that the bootstrap empirical process in the definition of the bootstrap statistic  $\hat{T}_n^*$  converges to the Gaussian process  $\nu$  at the  $\sqrt{n}$  rate; therefore, the behavior of the sequence  $\{r_n\}_n$  implies the asymptotic behavior of  $\hat{T}_n^*$  (conditional upon the data) is equivalent to  $\int_{A_n} \max \{ \nu(t), 0 \}^2 dt$ , where  $A_n = \{t \in [t, \bar{t}] : |E_P[g(\mathbf{X}; t)]| \leq r_n\}$ . Condition (C.7) 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  $\mathcal{P}_0$ .

### C.2.2 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 4.2. 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 4.2. These sets are subsets of their LSW counterparts that can be included in the empirical likelihood framework described in Sections 3, 4.1 and 4.2. The sets of probabilities included in the LSW framework that yield asymptotically correct size are: for each  $\epsilon > 0$

$$\mathcal{P}_0(\epsilon) = \{P \in \mathcal{P}_0 : \nu \text{ in (2.3) is regular on } B_n \forall n \geq 1\},$$

where  $B_n$  is defined in (4.1), in Part (i) of Definition 4.2. And the sets of probabilities in their framework that yield asymptotic similarity are: for each  $\epsilon > 0$  and decreasing sequence  $\xi_n \rightarrow 0$ ,

$$\begin{aligned} \mathcal{P}_0(\epsilon, \{\xi_n\}) = & \{P \in \mathcal{P}_0(\epsilon) : \nu \text{ in (2.3) is regular on } B(n^{-1/2}\xi_n), \\ & \text{and (4.2) holds } \forall n \geq 1/\epsilon\}. \end{aligned}$$

A natural question to raise at this point is by how much the set of probabilities in Parts (i) and (ii) of Definition 4.2 are more restrictive than their LSW counterparts. That is, given  $\epsilon > 0$  and a decreasing sequence  $\xi_n \rightarrow 0$ , how large are the set differences

$$\mathcal{P}_0(\epsilon) - \mathcal{P}_0^s(c_1, c_2, c_3, \epsilon) \quad \text{and} \quad \mathcal{P}_0(\epsilon, \{\xi_n\}) - \mathcal{P}_0^s(c_1, c_2, c_3, \epsilon, \{\xi_n\}), \quad (\text{C.8})$$

where  $c_1 \in (0, 1)$ ,  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$  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 (C.8) tend to the empty set as  $c_j \rightarrow 0$  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 (C.8) can be non-negligible. In practice, the appropriate values for these constants can be determined via statistical tests; Appendices C.1.2 and C.2.3 elaborate on this point.

### C.2.3 Verifying The Regularity Condition in Definitions 4.1

The first step in setting up the model of the null hypothesis, defined in Part (i) of Definition 4.2, 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 C.1.2 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

$$H_0 : \inf_{t \in \Delta(P_0)} E_{P_0} [g^2(\mathbf{X}; t)] \geq c_3 \quad \text{versus} \quad H_1 : \inf_{t \in \Delta(P_0)} E_{P_0} [g^2(\mathbf{X}; t)] < c_3. \quad (\text{C.9})$$

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.

## D 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 2. 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

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

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

respectively.

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

## D.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 6 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

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

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

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.

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

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

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

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

2. Generate  $B$  bootstrap samples  $\{\mathbf{X}_{i,l}^*\}_{i=1}^n$  for  $l = 1, \dots, B$  using the ECDF on the data.
3. For each bootstrap sample, compute  $\frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_{i,l}^*; t)$ , and  $\hat{\sigma}_l^2(t)$  just as  $\hat{\sigma}^2(t)$  is computed but with the bootstrap sample in place of the original sample.
4. For each bootstrap sample, compute the bootstrap test statistic  $\hat{T}_{n,l}^*$  as  $\hat{T}_n$  is computed in (D.3) but with  $\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i; t) \right)$  replaced by  $\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_{i,l}^*; t) - \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i; t) - \bar{\varphi}_n(t) \right)$  and with  $\hat{\sigma}^2(t)$  replaced by  $\hat{\sigma}_l^2(t)$ .
5. Take the bootstrap GMS critical value  $c_{n,1-\alpha}$  to be the  $1 - \alpha + \eta$  sample quantile of the bootstrap test statistics  $\{\hat{T}_{n,l}^*, l = 1, \dots, B\}$  plus  $\eta$ , 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  $\hat{\tau}^{AS} = 1[\hat{T}_n > c_{n,1-\alpha}]$ .

## D.2 Model of The Null Hypothesis

This section describes the subsets of  $\mathcal{P}_0$  on which  $\hat{\tau}^{AS}$  satisfies (2.1) in Definition 2.1. Define the scaled covariance kernel function

$$h_{2P}(t_1, t_2) = \frac{\text{Cov}_P(g(\mathbf{X}; t_1), g(\mathbf{X}; t_2))}{\sqrt{\text{VAR}_P(g(\mathbf{X}; t_1)) \text{VAR}_P(g(\mathbf{X}; t_2))}} \quad t_1, t_2 \in [\underline{t}, \bar{t}], \quad (\text{D.7})$$

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

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

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

$$d(h_2^{(1)}, h_2^{(2)}) = \sup_{t_1, t_2 \in [\underline{t}, \bar{t}]} \left| h_2^{(1)}(t_1, t_2) - h_2^{(2)}(t_1, t_2) \right|. \quad (\text{D.9})$$

According to Theorem 5.1 of AS, the subsets of  $\mathcal{P}_0$  on which  $\hat{\tau}^{AS}$  satisfies (2.1) are of the form

$$\{P \in \mathcal{P}_0 : h_{2P} \in \mathcal{C}_{cpt}\}, \quad (\text{D.10})$$

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

$$\limsup_{n \rightarrow +\infty} \sup_{P \in \{P \in \mathcal{P}_0 : h_{2P} \in \mathcal{C}_{cpt}\}} E_P \hat{\tau}^{AS} \leq \alpha, \quad (\text{D.11})$$

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

## D.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 (D.5) with the constrained empirical likelihood estimator described in Section 3. Let  $\hat{\varphi}_n(t)$  be constructed in the same way as  $\bar{\varphi}_n(t)$ , but with  $\sum_{i=1}^n \hat{p}_i g(\mathbf{X}_i; t)$  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

$$\hat{\Delta}_n^{AS} = \{t \in [\underline{t}, \bar{t}] : \bar{\varphi}_n(t) = 0\} \quad \text{and} \quad \hat{\Delta}_n^{AS} = \{t \in [\underline{t}, \bar{t}] : \hat{\varphi}_n(t) = 0\}, \quad (\text{D.12})$$

respectively.

Property 1 of the moment functions implies that

$$[\underline{t}, \bar{t}] - \hat{\Delta}_n^{AS} \subseteq [\underline{t}, \bar{t}] - \hat{\Delta}_n^{AS} \quad (\text{D.13})$$

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

$$\hat{\varphi}_n(t) = \bar{\varphi}_n(t) \quad \text{for } t \in \hat{\Delta}_n^{AS}, \quad (\text{D.14})$$

$$\hat{\varphi}_n(t) > \bar{\varphi}_n(t) \quad \text{for } t \in \hat{\Delta}_n^{AS} - \hat{\Delta}_n^{AS}, \quad \text{and} \quad (\text{D.15})$$

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

when  $\hat{p}$  exists and is characterisable in terms of Lagrange multipliers. Thus, conditional upon the sample, the inequalities (D.14) - (D.16) yield  $\hat{T}_{n,s}^* \geq \hat{T}_{n,s}^*$ , where  $\hat{T}_{n,s}^*$  is computed in exactly the same way as  $\hat{T}_{n,s}^*$  but with  $\hat{\varphi}_n(\cdot)$  in place of  $\bar{\varphi}_n(\cdot)$ . Moreover, if  $\hat{\Delta}_n^{\text{AS}} \subsetneq \hat{\Delta}_n^{\text{AS}}$  and  $\hat{\Delta}_n^{\text{AS}} \neq \emptyset$ , then  $\hat{T}_{n,s}^* > \hat{T}_{n,s}^*$  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.

## D.4 Asymptotic Equivalence Under The Null Hypothesis

Appendix D.2 describes subsets of  $\mathcal{P}_0$  on which the AS test is asymptotically level  $\alpha$ . These subsets differ from the subsets on which the LSW test is asymptotically level  $\alpha$ . 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  $s \in \mathbb{Z}_+$ ,  $c_1 \in (0, 1)$ ,  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$ ,  $c_3 \in (0, +\infty)$  and  $\mathcal{C}_{cpt}$  compact subset of  $\mathcal{C}$ , these subsets are

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

where  $\mathcal{P}_0^s(c_1, c_2, c_3)$  is given by Definition 3.1.

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

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

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

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

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

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

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

## D.5 Relative Behavior Under Local Alternatives

Given  $s \in \mathbb{Z}_+$ ,  $c_1 \in (0, 1)$ ,  $c_2 \in (0, (\bar{t} - \underline{t})^{-2})$ ,  $c_3 \in (0, +\infty)$  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  $\lim_{n \rightarrow +\infty} P_n \in \partial \mathcal{P}_0^s(c_1, c_2, c_3, \mathcal{C}_{cpt})$ , 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  $\hat{\tau}^{\text{AS}}$ . It was shown in Appendix D.3 that the bootstrap critical value from  $\hat{\tau}^{\text{AS}}$  is never larger than that from  $\hat{\tau}^{\text{AS}}$ , 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

$$E_{P_n} \hat{\tau}^{\text{AS}} \geq E_{P_n} \hat{\tau}^{\text{AS}}, \quad (\text{D.22})$$

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

$$P_n \left[ \hat{\Delta}_n^{\text{AS}} \subsetneq \hat{\Delta}_n^{\text{AS}} \text{ and } \hat{\Delta}_n^{\text{AS}} \neq \emptyset \right] > 0, \quad (\text{D.23})$$

for large enough  $n$ .

In consequence, results analogous to Theorem 5.3 and Corollary 5.2 for the LSW test and its modified version hold for comparison of between  $\hat{\tau}^{\text{AS}}$  and  $\hat{\tau}^{\text{AS}}$ . 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 Assumption 5.1), which is due to the set inclusion  $\hat{\Delta}_n^{\text{AS}} \subset \hat{\Delta}_n^{\text{AS}}$  holding for large enough  $n$ .

## E Proofs of Results

### E.1 Theorem 3.1

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

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

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

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

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

**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:

$$g(X_i; t) = 1 [X_i^B \leq t] - 1 [X_i^A \leq t] \quad t \in [\underline{t}, \bar{t}]. \quad (\text{E.5})$$

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 F.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  $\forall t \in [\underline{t}, \bar{t}] - \mathcal{T}_n \exists j \in \{1, 2, \dots, N\}$  such that

$$t_{(j)} \leq t \leq t_{(j+1)}. \quad (\text{E.6})$$

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

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



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

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

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

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 F.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  $\check{\mathbf{p}}$ , respectively. The constraint set equality we just proved implies the equality  $\tilde{\mathbf{p}} = \check{\mathbf{p}}$ . 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 F.6 establishes this result:

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

because  $\{\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\} \subseteq \{\tilde{\mathbf{p}} \text{ exists and is unique}\}$ , 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 F.

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

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

That is, on this event, we construct the upper bound on  $\|\check{\mathbf{p}} - \tilde{\mathbf{p}}\|$ . The result then follows directly since Lemma F.6 in Appendix F.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 F.2, F.5 and F.6, respectively.

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

**Step 2.** By Lemma F.9 in Appendix F.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  $\check{\mathbf{p}}$ . The consequence of this step is that by the triangular inequality we can conclude

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

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

$$\|\check{\mathbf{p}} - \tilde{\mathbf{p}}\|^2 \leq \frac{h(\check{\mathbf{p}}) - h(\tilde{\mathbf{p}})}{K_0} = \frac{h(\check{\mathbf{p}}) - h(\check{\mathbf{p}})}{K_0} + \frac{h(\check{\mathbf{p}}) - h(\tilde{\mathbf{p}})}{K_0}, \quad (\text{E.11})$$

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

$$\|\check{\mathbf{p}} - \tilde{\mathbf{p}}\|^2 \leq \frac{h(\check{\mathbf{p}}) - h(\hat{\mathbf{p}})}{K_0}. \quad (\text{E.12})$$

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

$$\|\hat{\mathbf{p}} - \tilde{\mathbf{p}}\| \leq \left[ \frac{h(\check{\mathbf{p}}) - h(\hat{\mathbf{p}})}{K_0} \right]^{1/2} + \frac{\hat{p}_{(1)}}{2n^{3/2}}. \quad (\text{E.13})$$

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 (E.13) converge to zero in probability, with uniformity over sets of probabilities  $\mathcal{P}_1(c_1) \cap \mathcal{P}_2(c_2)$ . Observe that by the inequality  $\log(1+y) \leq y \forall y > -1$

$$h(\check{\mathbf{p}}) - h(\hat{\mathbf{p}}) = \sum_{i=1}^n \log \left( 1 + \frac{\hat{p}_i - \check{p}_i}{\check{p}_i} \right) \leq \sum_{i=1}^n \frac{\hat{p}_i - \check{p}_i}{\check{p}_i}. \quad (\text{E.14})$$

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

$$\sum_{i=1}^n \frac{\hat{p}_i - \check{p}_i}{\check{p}_i} \leq \sqrt{\sum_{i=1}^n \frac{1}{\check{p}_i^{3/2}}} \|\hat{\mathbf{p}} - \check{\mathbf{p}}\| \leq \frac{2\sqrt{n}}{\hat{p}_{(1)}} \|\hat{\mathbf{p}} - \check{\mathbf{p}}\| \leq 1/n. \quad (\text{E.15})$$

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

$$\|\hat{\mathbf{p}} - \tilde{\mathbf{p}}\| \leq (K_0 n)^{-1/2} + n^{-3/2}, \quad (\text{E.16})$$

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

Therefore, the steps described above yields the following event inclusion

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

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

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

$$\leq \lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_1(c_1) \cap \mathcal{P}_2(c_2)} P \left[ \|\hat{\mathbf{p}} - \tilde{\mathbf{p}}\| \leq (K_0 n)^{-1/2} + n^{-3/2} \right], \quad (\text{E.19})$$

which implies the desired result and concludes the proof.  $\square$

## E.2 Theorem 4.1

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

$$\Delta(\hat{P}_n) = \{t_1, t_2, \dots, t_{\omega_n}\}. \quad (\text{E.20})$$

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

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

Therefore,

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

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

$$g(\mathbf{x}; t') g(\mathbf{x}; t) \geq 0 \quad \forall (t, \mathbf{x}) \in [\underline{t}, \bar{t}] \times \text{supp}(P). \quad (\text{E.23})$$

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

**Part 2.** First, we prove the probability of the event  $\{\hat{\Delta}_n \subseteq \hat{\Delta}_n\}$  (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  $t \in \hat{\Delta}_n$ , then  $\{-r_n < E_{\hat{P}_n} [g(\mathbf{X}, t)]\}$  occurs. Consequently, part 1 of the theorem implies the probability of the event  $\{-r_n < E_{\hat{P}_n} [g(\mathbf{X}, t)]\}$ , 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  $\{r_n > E_{\hat{P}_n} [g(\mathbf{X}, t)]\}$ , for each  $t \in \hat{\Delta}_n$ , tends to unity with uniformity.

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

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

where (E.24) follows by Proposition G.1. Next we show that the probability of the event

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

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

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

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

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

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  $\sqrt{n}r_n \rightarrow +\infty$ , it follows that

$$\sup_{P \in \mathcal{P}_0^s(c_1, c_2, c_3)} P[\sqrt{n}E_{\hat{P}_n} [g(\mathbf{X}, t)] \in [\sqrt{n}r_n, O_P(1) + \sqrt{n}r_n]] \rightarrow 0. \quad (\text{E.26})$$

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

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

$$\sqrt{n}E_{\hat{P}_n} [g(\mathbf{X}, t)] = \sqrt{n} (E_{\hat{P}_n} [g(\mathbf{X}, t)] - E_P [g(\mathbf{X}, t)]) + \sqrt{n}E_P [g(\mathbf{X}, t)] \quad (\text{E.27})$$

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

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  $\sqrt{nr_n} \rightarrow +\infty$ , implies that the probability (E.25) tends to zero with uniformity. Therefore,  $\inf_{P \in \mathcal{P}_0^s(c_1, c_2, c_3)} P [E_{\hat{P}_n} [g(\mathbf{X}, t)] < r_n] \rightarrow 1$ , which concludes the proof of this part of the theorem.

Now we turn our focus to the event  $\{\Delta(P) \subseteq \hat{\Delta}_n\}$ . Let  $t \in \Delta(P)$  and consider the event  $\{-r_n < E_{\hat{P}_n} [g(\mathbf{X}, t)] < r_n\}$ . By adding and subtracting the term  $E_{\hat{P}_n} [g(\mathbf{X}, t)]$  from  $E_{\hat{P}_n} [g(\mathbf{X}, t)]$ , this event is equal to the event

$$\{-r_n < E_{\hat{P}_n} [g(\mathbf{X}, t)] - E_{\hat{P}_n} [g(\mathbf{X}, t)] + E_{\hat{P}_n} [g(\mathbf{X}, t)] < r_n\}.$$

Now we can apply Proposition G.1 to the difference  $E_{\hat{P}_n} [g(\mathbf{X}, t)] - E_{\hat{P}_n} [g(\mathbf{X}, t)]$  to deduce that this event is also equal to the event

$$\{-\sqrt{nr_n} < O_P(1) + \sqrt{n}E_{\hat{P}_n} [g(\mathbf{X}, t)] < \sqrt{nr_n}\}, \quad (\text{E.29})$$

uniformly over  $\mathcal{P}_0^s(c_1, c_2, c_3)$ . 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 (E.29) is equal to

$$\{-\sqrt{nr_n} \leq O_P(1) \leq \sqrt{nr_n}\}, \quad (\text{E.30})$$

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

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

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

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

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

diverges to  $+\infty$  as  $n \rightarrow +\infty$ , 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.,

$$\lim_{n \rightarrow +\infty} \sup_{P \in \mathcal{P}_0^s(c_1, c_2, c_3)} P [\exists t \in \Delta_n; O_P(1) < \sqrt{nr_n} - O_P(1) - \sqrt{n}E_P [g(\mathbf{X}, t)] \leq 0] = 0.$$

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

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

holds for each  $n$ .

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

$$\{\hat{\Delta}_n \subsetneq \dot{\Delta}_n\} = \{\dot{P}_n \neq \hat{P}_n\} \cap \{\exists t \in \hat{\Delta}_n; |E_{\dot{P}_n}[g(\mathbf{X}; t)]| \geq r_n\} \cap \{\dot{\Delta}_n \subseteq \hat{\Delta}_n\}. \quad (\text{E.31})$$

Noting that  $\{\forall t \in \hat{\Delta}_n; |E_{\dot{P}_n}[g(\mathbf{X}; t)]| < r_n\} = \{\hat{\Delta}_n \subseteq \dot{\Delta}_n\}$ , the probability of the complement of (E.31) is

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

which equals

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

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

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

because Lemma G.1 establishes that  $\lim_{n \rightarrow +\infty} \sup_{P \in \mathcal{P}_0^s(c_1, c_2, c_3) \cap \mathcal{P}_{000}} P \left[ \dot{P}_n = \hat{P}_n \right] \leq 1/2$ , and by Part 2 of Theorem 4.1

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

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

### E.3 Theorem 4.2

*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

$$\gamma_n^*(t) = \left( \max \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n [g(\mathbf{X}_{i,l}^*; t) - E_{\hat{P}_n}[g(\mathbf{X}; t)]] , 0 \right\} \right)^2, \quad (\text{E.32})$$

then consider the following,

$$\left| \hat{T}_{n,l}^* - \dot{T}_{n,l}^* \right| = \begin{cases} \int_{[\underline{t}, \bar{t}] - \hat{\Delta}_n} \gamma_n^*(t) dt & \text{if } \int_{\hat{\Delta}_n} dt > 0, \int_{\hat{\Delta}_n} dt = 0 \\ \int_{[\underline{t}, \bar{t}] - \hat{\Delta}_n} \gamma_n^*(t) dt & \text{if } \int_{\hat{\Delta}_n} dt = 0, \int_{\hat{\Delta}_n} dt > 0 \\ \int_{\hat{\Delta}_n \ominus \hat{\Delta}_n} \gamma_n^*(t) dt & \text{if } \int_{\hat{\Delta}_n} dt > 0, \int_{\hat{\Delta}_n} dt > 0 \\ 0 & \text{if } \int_{\hat{\Delta}_n} dt = 0, \int_{\hat{\Delta}_n} dt = 0, \end{cases} \quad (\text{E.33})$$

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

$$\left| \hat{T}_{n,l}^* - \dot{T}_{n,l}^* \right| \leq \begin{cases} \left( \sup_{t \in [\underline{t}, \bar{t}]} \gamma_n^*(t) \right) \int_{[\underline{t}, \bar{t}] - \hat{\Delta}_n} dt & \text{if } \int_{\hat{\Delta}_n} dt > 0, \int_{\hat{\Delta}_n} dt = 0 \\ \left( \sup_{t \in [\underline{t}, \bar{t}]} \gamma_n^*(t) \right) \int_{[\underline{t}, \bar{t}] - \hat{\Delta}_n} dt & \text{if } \int_{\hat{\Delta}_n} dt = 0, \int_{\hat{\Delta}_n} dt > 0 \\ \left( \sup_{t \in [\underline{t}, \bar{t}]} \gamma_n^*(t) \right) \int_{\hat{\Delta}_n \ominus \hat{\Delta}_n} dt & \text{if } \int_{\hat{\Delta}_n} dt > 0, \int_{\hat{\Delta}_n} dt > 0, \\ 0 & \text{if } \int_{\hat{\Delta}_n} dt = 0, \int_{\hat{\Delta}_n} dt = 0. \end{cases} \quad (\text{E.34})$$

To prove the result we need to prove that  $\left( \sup_{t \in [\underline{t}, \bar{t}]} \gamma_n^*(t) \right)$  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 4.1 to the integrals in (E.34). Since the set of moment functions  $\{\mathbf{x} \mapsto g(\mathbf{x}, t), t \in [\underline{t}, \bar{t}]\}$  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  $\left( \sup_{t \in [\underline{t}, \bar{t}]} \gamma_n^*(t) \right)$  yields the desired result.

Parts 3 and 2 of Theorem 4.1 imply that  $\hat{\Delta}_n$  and  $\dot{\Delta}_n$  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  $\dot{\Delta}_n \neq \emptyset, \hat{\Delta}_n \neq \emptyset$  with probability tending to one, uniformly in  $\mathcal{P}_0^s(c_1, c_2, c_3, \epsilon)$ . Applying Part 2 of Theorem 4.1 to this case in (E.34) implies  $\hat{\Delta}_n \ominus \dot{\Delta}_n = \hat{\Delta}_n - \dot{\Delta}_n$  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 4.1

$$\left( \sup_{t \in [\underline{t}, \bar{t}]} \gamma_n^*(t) \right) \int_{\hat{\Delta}_n \ominus \dot{\Delta}_n} dt \xrightarrow{P} 0 \quad (\text{E.35})$$

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

## E.4 Theorem 5.1

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

$$\hat{T}_n = n \int_{\underline{t}}^{\bar{t}} \max \{ E_{\hat{P}_n} [g(\mathbf{X}; t)], 0 \}^2 dt, \quad (\text{E.36})$$

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  $\hat{\Upsilon}_B$  converges to zero in probability under  $H_1$ .

Corollary F.1 and Part 2 of Lemma F.5 implies the existence and uniqueness of the solution from the constrained empirical likelihood problem,  $\hat{\mathbf{p}}$ , to be an event with probability converging to unity. Hence, the contact set  $\hat{\Delta}_n$  exists with probability converging to unity. Since the bootstrap test statistic  $\hat{T}_n^*$  is bounded above by

$$\int_{\underline{t}}^{\bar{t}} \left( \max \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n [g(\mathbf{X}_i^*; t) - E_{\hat{P}_n} [g(\mathbf{X}; t)]] , 0 \right\} \right)^2 dt, \quad (\text{E.37})$$

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

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

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

## E.5 Theorem 5.2

*Proof. Part 1.* The proof follows steps identical to those in Part 1 of Theorem 4.1 except that we are taking limits under the local alternatives, which is based on Lemma H.4. We omit the details for brevity.

**Part 2.** The proof proceeds by the direct method. We first focus on proving  $\lim_{n \rightarrow +\infty} P_n [C \subseteq \hat{\Delta}_n] = 1$ . Lemma H.4 implies that the random set  $\hat{\Delta}_n$  exists for large enough  $n$ , with probability approaching unity under the local alternatives. Consider  $t \in [\underline{t}, \bar{t}]$  such that  $H(t) = 0$  and the event  $\{-r_n \leq E_{\hat{P}_n} [g(\mathbf{X}; t)] \leq r_n\}$ . Upon adding and subtracting  $E_{\hat{P}_n} [g(\mathbf{X}, t)]$  and  $E_{P_n} [g(\mathbf{X}, t)]$  to  $E_{\hat{P}_n} [g(\mathbf{X}; t)]$ , and multiplying through  $\sqrt{n}/\sigma_n(t)$  on all sides of the inequalities in this event, it is equal to

$$\left\{ \left| \frac{A(t)}{\sigma_n(t)} + \frac{B(t)}{\sigma_n(t)} + \frac{\delta(t)}{\sigma_n(t)} \right| \leq \frac{\sqrt{n}r_n}{\sigma_n(t)} \right\}, \quad (\text{E.38})$$

where

$$\begin{aligned} A(t) &= \sqrt{n} (E_{\hat{P}_n} [g(\mathbf{X}; t)] - E_{\hat{P}_n} [g(\mathbf{X}, t)]) \quad \text{and} \\ B(t) &= \sqrt{n} (E_{\hat{P}_n} [g(\mathbf{X}; t)] - E_{P_n} [g(\mathbf{X}, t)]). \end{aligned}$$

Now we derive the large sample behaviors of each of the elements in the event given by (E.38), then combine them to deduce the main result. Noting that,

$$A(t) = -\sqrt{n} \sum_{t' \in \Delta(\hat{P}_n)} \hat{\mu}_{t'} \sum_{i=1}^n \hat{p}_i g(\mathbf{X}_i; t') g(\mathbf{X}_i; t), \quad (\text{E.39})$$

we can follow steps identical to those in Proposition G.1 to deduce that

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

under the local alternatives, and we omit them for brevity. In consequence, Part 5 of Lemma H.4 implies the right side of (E.39) is  $O_{P_n}(1)$ . We also have that

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{n}r_n}{\sigma_n(t)} \rightarrow +\infty$$

because  $\sigma_n(\cdot)$  is uniformly bounded by Part(iv) of Assumption 5.1 and Assumption 2.1, and because  $\lim_{n \rightarrow +\infty} \sqrt{n}r_n = +\infty$ . Furthermore, and

$$\frac{B(t)}{\sigma_n(t)} = O_{P_n}(1) \quad \text{as } n \rightarrow +\infty$$

by the Lindeberg-Feller Central Limit Theorem for IID triangular arrays. Finally, because  $\delta(t)$  is a uniformly bounded function by Part (iii) of Assumption 5.1, it follows that

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

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

Thus, to prove the result we need to show that  $\lim_{n \rightarrow +\infty} P_n \left[ E_{\hat{P}_n} [g(\mathbf{X}; t)] < r_n \right] = 1$ .

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

$$E_{\hat{P}_n} [g(\mathbf{X}; t)] = E_{\hat{P}_n} [g(\mathbf{X}; t)] - E_{P_n} [g(\mathbf{X}; t)] + E_{P_n} [g(\mathbf{X}; t)], \quad \text{and} \quad (\text{E.40})$$

$$\sqrt{n} \left( E_{\hat{P}_n} [g(\mathbf{X}; t)] - E_{P_n} [g(\mathbf{X}; t)] \right) / \sigma_{P_n}(t) = O_{P_n}(1) \quad (\text{E.41})$$

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

$$\lim_{n \rightarrow +\infty} P_n \left[ O_{P_n}(1) + \frac{\sqrt{n}}{\sigma_{P_n}(t)} E_{P_n} [g(\mathbf{X}; t)] \geq \sqrt{n}r_n / \sigma_{P_n}(t) \right]. \quad (\text{E.42})$$

As  $\sqrt{n}r_n / \sigma_{P_n}(t) \rightarrow +\infty$ , for  $t \in C$ , the term  $\sqrt{n}E_{P_n} [g(\mathbf{X}; t)] / \sigma_{P_n}(t) = \delta(t) / \sigma_{P_n}(t)$ , is uniformly bounded with uniformity over  $n$ ; hence, the limit (E.42) is equal to zero. Furthermore, for  $t \in [\underline{t}, \bar{t}] - C$ , the term

$$\sqrt{n}E_{P_n} [g(\mathbf{X}; t)] / \sigma_{P_n}(t) = (\sqrt{n}H(t) + \delta(t)) / \sigma_{P_n}(t) \rightarrow -\infty \quad (\text{E.43})$$

as  $H(t) < 0$  and because  $\delta(t) / \sigma_{P_n}(t)$  is uniformly bounded in  $n$  and  $t$  by the conditions of Assumption 5.1. Therefore, the limit (E.42) is equal to zero for such  $t$  since the divergence to  $-\infty$  in (E.43) 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 4.1, except that the limits are taken under the sequence of local alternatives. We have



that

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

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

$$\sqrt{n}r_n - O_{P_n}(1) - \sqrt{n}E_{P_n}[g(\mathbf{X}, t)] = \sqrt{n}r_n - O_{P_n}(1) - \delta(t) \rightarrow +\infty \quad (\text{E.44})$$

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

$$\sqrt{n}r_n - O_{P_n}(1) - \sqrt{n}E_{P_n}[g(\mathbf{X}, t)] = \sqrt{n}r_n - O_{P_n}(1) - \sqrt{n}H(t) - \delta(t) \rightarrow +\infty \quad (\text{E.45})$$

as  $n \rightarrow +\infty$ , because  $\sqrt{n}r_n \rightarrow +\infty$  and  $-\sqrt{n}H(t) \rightarrow +\infty$  as  $n \rightarrow +\infty$  and  $\delta(t)$  is finite. In consequence, the event (E.44) tends to the empty set under the sequence of local alternatives, which implies that

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

and concludes the proof.

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

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

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

$$\lim_{n \rightarrow +\infty} P_n \left[ \sup_{t \in [\underline{t}, \bar{t}]} |E_{P_n}[g(\mathbf{X}; t)] - E_{\hat{P}_n}[g(\mathbf{X}; t)]| \leq \epsilon r_n \right] = 1, \quad (\text{E.46})$$

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

$$\sqrt{n} [E_{P_n}[g(\mathbf{X}; t)] - E_{\hat{P}_n}[g(\mathbf{X}; t)]] \quad (\text{E.47})$$

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

$$\lim_{n \rightarrow +\infty} P_n \left[ \sup_{t \in [\underline{t}, \bar{t}]} \sqrt{n} |E_{P_n}[g(\mathbf{X}; t)] - E_{\hat{P}_n}[g(\mathbf{X}; t)]| > \epsilon \sqrt{n}r_n \right] = 0, \quad (\text{E.48})$$

holds. This limit result implies the limit (E.46) since it is the limit of the probabilities of the complementary event under the sequence of local alternatives. This concludes the proof.  $\square$

## E.6 Corollary 5.1

*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  $B \rightarrow +\infty$ . They are

$$\hat{\Upsilon}_\infty = P^* \left[ \hat{T}_n^* \geq \hat{T}_n \right] \quad \text{and} \quad \hat{\Upsilon}_\infty = P^* \left[ \hat{T}_n^* \geq \hat{T}_n \right], \quad (\text{E.49})$$

which are conditional on  $\mathcal{A}_n$ . By Theorem 5.2, the bootstrap test statistics  $\hat{T}_n^*$  and  $\hat{T}_n$  converge in distribution conditional on  $\mathcal{A}_n$  to  $\int_{\hat{\Delta}_n} \max \{\nu(t), 0\}^2 dt$  and  $\int_{\hat{\Delta}_n} \max \{\nu(t), 0\}^2 dt$ , respectively, under the sequence of local alternatives. Furthermore, Theorem 5.2 implies that asymptotically

$$\int_{\hat{\Delta}_n} \max \{\nu(t), 0\}^2 dt \leq \int_{\hat{\Delta}_n} \max \{\nu(t), 0\}^2 dt \quad (\text{E.50})$$

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

$$\hat{\Upsilon}_\infty = P^* \left[ \int_{\hat{\Delta}_n} \max \{\nu(t), 0\}^2 dt \geq \hat{T}_n \right] \leq P^* \left[ \int_{\hat{\Delta}_n} \max \{\nu(t), 0\}^2 dt \geq \hat{T}_n \right] = \hat{\Upsilon}_\infty \quad (\text{E.51})$$

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

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

Furthermore, Part 4 of Theorem 5.2 implies that the contact set estimators  $\hat{\Delta}_n$  and  $\hat{\Delta}_n$  converge to the set  $C$ ; therefore, we must have the equality

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

Finally, on the event  $\{\hat{\Delta}_n \subsetneq \hat{\Delta}_n\}$ , the inequalities (E.50) and (E.51) hold strictly. As for large  $n$  the probability  $P_n[\hat{\Delta}_n \subsetneq \hat{\Delta}_n] > 0$ , these inequalities hold strictly with positive probability, which implies that

$$P_n \left[ \left( \left\{ \hat{\Upsilon}_\infty \leq \alpha \right\} - \left\{ \hat{\Upsilon}_\infty \leq \alpha \right\} \right) \cap \left\{ \hat{\Delta}_n \subsetneq \hat{\Delta}_n \right\} \right] > 0. \quad (\text{E.52})$$

□

## E.7 Theorem 5.3

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

**Part 1.** By Part 2 of Corollary 5.1, the condition  $P_n \left[ \hat{\Delta}_n \subsetneq \hat{\Delta}_n \right] > 0$  for large  $n$ , implies that  $P_n \left[ \hat{\Upsilon}_\infty \leq \alpha \right] > P_n \left[ \hat{\Upsilon}_\infty \leq \alpha \right]$  holds for large  $n$ . In conjunction with this implication, the

condition  $P_n \left[ \hat{Y}_\infty \leq \alpha \right] \nearrow p_\infty \in (0, 1)$  as  $n \rightarrow +\infty$ , further implies  $P_n \left[ \check{Y}_\infty \leq \alpha \right] \nearrow p_\infty$  as  $n \rightarrow +\infty$ . In consequence,

$$p_\infty > P_n \left[ \check{Y}_\infty \leq \alpha \right] > P_n \left[ \hat{Y}_\infty \leq \alpha \right] \quad \text{for large } n. \quad (\text{E.53})$$

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

$$P_m \left[ \hat{Y}_\infty \leq \alpha \right] > P_n \left[ \check{Y}_\infty \leq \alpha \right] \quad \forall m \geq N_{\gamma_n}. \quad (\text{E.54})$$

Hence,

$$\left\{ N \in \mathbb{Z}_+ : P_m \left[ \hat{Y}_\infty \leq \alpha \right] \geq P_n \left[ \check{Y}_\infty \leq \alpha \right] \quad \forall m \geq N \right\} \neq \emptyset, \quad (\text{E.55})$$

as  $N_{\gamma_n}$  is an element of it. Moreover, the set in (E.55) is bounded from below by the integer 1; therefore, the infimum

$$k'_n = \inf \left\{ N \in \mathbb{Z}_+ : P_m \left[ \hat{Y}_\infty \leq \alpha \right] \geq P_n \left[ \check{Y}_\infty \leq \alpha \right] \quad \forall m \geq N \right\}, \quad (\text{E.56})$$

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 5.1, the condition  $P_n \left[ \hat{\Delta}_n \subsetneq \hat{\Delta}_n \right] > 0$  for large  $n$ , implies that  $P_n \left[ \check{Y}_\infty \leq \alpha \right] > P_n \left[ \hat{Y}_\infty \leq \alpha \right]$  holds for large  $n$ . In conjunction with this implication, the condition  $P_n \left[ \check{Y}_\infty \leq \alpha \right] \searrow p_\infty \in (0, 1)$  as  $n \rightarrow +\infty$ , further implies  $P_n \left[ \hat{Y}_\infty \leq \alpha \right] \searrow p_\infty$  as  $n \rightarrow +\infty$ . In consequence,

$$P_n \left[ \check{Y}_\infty \leq \alpha \right] > P_n \left[ \hat{Y}_\infty \leq \alpha \right] > p_\infty \quad \text{for large } n. \quad (\text{E.57})$$

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

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

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

$$P_m \left[ \hat{Y}_\infty \leq \alpha \right] < P_n \left[ \check{Y}_\infty \leq \alpha \right] \quad \forall m \geq N_{\gamma_n}. \quad (\text{E.58})$$

Hence,

$$\left\{ N \in \mathbb{Z}_+ : P_m \left[ \hat{Y}_\infty \leq \alpha \right] \geq P_n \left[ \check{Y}_\infty \leq \alpha \right] \quad \forall m \geq N \right\} = \emptyset, \quad (\text{E.59})$$

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

## E.8 Corollary 5.2

*Proof. Part 1.* The proof proceeds by the direct method. Given  $\Omega \in \mathbb{H}$  and  $\{P_n\}_{n \geq 1} \in \mathcal{P}_{1,+}^s(\Omega)$ , under the conditions of the Corollary, when  $n$  is large enough, Part 1 of Theorem 5.3 establishes the existence of  $k_n$ , and therefore,  $P_{k_n} [\hat{Y}_\infty \leq \alpha]$  exists. The strict inequality,  $k_n > n$ , holds, since  $P_n [\hat{Y}_\infty \leq \alpha] < P_n [\hat{Y}_\infty \leq \alpha]$ ; hence,  $e_n = k_n/n > 1$  and  $d_n = k_n - n > 0$ , when  $n$  is large enough. In consequence,  $\liminf_{n \rightarrow +\infty} (e_n) \geq 1$  and  $\liminf_{n \rightarrow +\infty} d_n \geq 0$ , for each  $\{P_n\}_{n \geq 1} \in \mathcal{P}_{1,+}^s(\Omega)$ . As  $\Omega \in \mathbb{H}$  was arbitrary, the above manipulations hold for each  $\Omega \in \mathbb{H}$ , and concludes the proof.

**Part 2.** The proof proceeds by the direct method. Given  $\Omega \in \mathbb{H}$  and  $\{P_n\}_{n \geq 1} \in \mathcal{P}_{1,+}^s(\Omega)$  such that  $\liminf_{n \rightarrow +\infty} \psi_n > 0$ . The proof makes use of the following linear function of  $m \in \mathbb{R}_+$  that passes through the points  $(n, P_n[\hat{Y}_\infty \leq \alpha])$  and  $(k_n, P_{k_n}[\hat{Y}_\infty \leq \alpha])$ :

$$Y = b_n + \left( \frac{P_{k_n} [\hat{Y}_\infty \leq \alpha] - P_n [\hat{Y}_\infty \leq \alpha]}{d_n} \right) m, \quad (\text{E.60})$$

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

$$b_n = P_n [\hat{Y}_\infty \leq \alpha] - \left( \frac{P_{k_n} [\hat{Y}_\infty \leq \alpha] - P_n [\hat{Y}_\infty \leq \alpha]}{d_n} \right) n, \quad \text{or} \quad (\text{E.61})$$

$$b_n = P_{k_n} [\hat{Y}_\infty \leq \alpha] - \left( \frac{P_{k_n} [\hat{Y}_\infty \leq \alpha] - P_n [\hat{Y}_\infty \leq \alpha]}{d_n} \right) k_n. \quad (\text{E.62})$$

Solving for  $d_n$  in equation (E.61) yields

$$d_n = \left( \frac{P_{k_n} [\hat{Y}_\infty \leq \alpha] - P_n [\hat{Y}_\infty \leq \alpha]}{P_n [\hat{Y}_\infty \leq \alpha] - b_n} \right) n = \psi_n n. \quad (\text{E.63})$$

Noting that  $e_n = 1 + d_n/n$ , we have that  $e_n = 1 + \psi_n$ . In consequence,  $\liminf_{n \rightarrow +\infty} (e_n) \geq 1 + \liminf_{n \rightarrow +\infty} \psi_n$ , which is greater than one by hypothesis. As  $\Omega \in \mathbb{H}$  was arbitrary, the above manipulations hold for each  $\Omega \in \mathbb{H}$  with local alternatives in  $\mathcal{P}_{1,+}^s(\Omega)$  that satisfy  $\liminf_{n \rightarrow +\infty} \psi_n > 0$ . This concludes the proof.

**Part 3.** The proof follows steps identical to those in **Part 2** to deduce that  $e_n = 1 + \psi_n$ , except that now we apply limit supremum to  $e_n$  and use the condition  $\liminf_{n \rightarrow +\infty} \psi_n = 0$ . That is,  $\limsup_{n \rightarrow +\infty} (e_n) \leq 1 + \limsup_{n \rightarrow +\infty} \psi_n = 1$ . As **Part 1** of this corollary establishes  $\liminf_{n \rightarrow +\infty} (e_n) \geq 1$  for each  $\{P_n\}_{n \geq 1} \in \mathcal{P}_{1,+}^s(\Omega)$ , it follows that

$$1 = \limsup_{n \rightarrow +\infty} e_n = \liminf_{n \rightarrow +\infty} e_n = \liminf_{n \rightarrow +\infty} e_n = \lim_{n \rightarrow +\infty} e_n, \quad (\text{E.64})$$

for each  $\{P_n\}_{n \geq 1} \in \mathcal{P}_{1,+}^s(\Omega)$  that satisfies  $\liminf_{n \rightarrow +\infty} \psi_n = 0$ . Finally, the above manipulation holds for any  $\Omega \in \mathbb{H}$ , which concludes the proof.

**Part 4.** Given  $\Omega \in \mathbb{H}$ , the proof follows steps identical to those in **Part 3** to establish that  $\lim_{n \rightarrow +\infty} e_n = 1$ . To establish that  $\lim_{n \rightarrow +\infty} d_n = +\infty$ , we use the extra condition

$$\liminf_{n \rightarrow +\infty} n\psi_n = +\infty$$

on the sequences  $\{P_n\}_{n \geq 1} \in \mathcal{P}_{1,+}^s(\Omega)$ . Noting that we established in the proof of **Part 2**, we have  $d_n = n\psi_n$  when  $n$  is large enough. So this extra condition implies  $\liminf_{n \rightarrow +\infty} d_n = +\infty$ , which, in turn, implies  $\lim_{n \rightarrow +\infty} d_n = +\infty$ . In consequence, on such sequences  $\{P_n\}_{n \geq 1}$ ,  $k_n$  admits the following expansion:  $k_n = n + \ell q_n + o(q_n)$ , when  $n$  is large enough, where  $q_n = o(n)$ ,  $\lim_{n \rightarrow +\infty} q_n = +\infty$  and  $\ell > 0$ . The result follows from re-writing the expansion as

$$d_n/q_n = \ell + o(1). \quad (\text{E.65})$$

This concludes the proof. □

## F Intermediate Technical Results for Theorem 3.1

This section presents intermediate technical results that are used in the proofs of Theorem 3.1.

### F.1 Consequences of Definition 3.1

Define the sets

$$I_n^- = \{i \in \{1, \dots, n\} : g(\mathbf{X}_i; t) < 0 \forall t \in [\underline{t}, \bar{t}]\}, \quad (\text{F.1})$$

$$I_n^+ = \{i \in \{1, \dots, n\} : g(\mathbf{X}_i; t) \geq 0 \forall t \in [\underline{t}, \bar{t}]\} \quad \text{and} \quad (\text{F.2})$$

$$I_n = \{(i, K) \in \{1, \dots, n\} \times \{A, B\} : X_i^K \in [\underline{t}, \bar{t}]\}. \quad (\text{F.3})$$

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  $[\underline{t}, \bar{t}]$ .

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

**Lemma F.1.** Given  $c_1 \in (0, 1)$  and recall that  $\mathcal{P}_1(c_1)$  is defined in (3.3). Then

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

*Proof.* The proof proceeds by the direct method. We observe that

$$\inf_{P \in \mathcal{P}_1(c_1)} P [I_n^- \neq \emptyset] = 1 - \sup_{P \in \mathcal{P}_1(c_1)} P [I_n^- = \emptyset]. \quad (\text{F.4})$$

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

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

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

$$\sup_{P \in \mathcal{P}_1(c_1)} P [I_n^- = \emptyset] = \sup_{P \in \mathcal{P}_1(c_1)} \left( P \left[ \sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}_1; t) \geq 0 \right] \right)^n \quad (\text{F.5})$$

$$= \sup_{P \in \mathcal{P}_1(c_1)} \left( 1 - P \left[ \sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}_1; t) < 0 \right] \right)^n \quad (\text{F.6})$$

$$\leq (1 - c_1)^n \rightarrow 0 \quad (\text{F.7})$$

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

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

**Lemma F.2.** Given  $c_2 \in \left(0, \frac{1}{(\bar{t} - \underline{t})^2}\right)$  and recall that  $\mathcal{P}_2(c_2)$  is defined in (3.4). Then

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

*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

$$\inf_{P \in \mathcal{P}_2(c_2)} P [I_n \neq \emptyset] = 1 - \sup_{P \in \mathcal{P}_2(c_2)} P [I_n = \emptyset]. \quad (\text{F.8})$$

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

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

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

$$\sup_{P \in \mathcal{P}_2(c_2)} P [I_n = \emptyset] = \sup_{P \in \mathcal{P}_2(c_2)} (P [X_i^A, X_i^B \notin [\underline{t}, \bar{t}]])^n \quad (\text{F.9})$$

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

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

$$\leq (1 - c_2 (\bar{t} - \underline{t})^2)^n \rightarrow 0 \quad (\text{F.12})$$

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

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

**Lemma F.3.** Given  $c_2 \in \left(0, \frac{1}{(\bar{t} - \underline{t})^2}\right)$  and recall that  $\mathcal{P}_2(c_2)$  is defined in (3.4). Then

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

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

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

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

$$\{\mathbf{X}_i \in [\underline{t}, \bar{t}] \times [\underline{t}, \bar{t}]\} \subseteq \{I_n^+ \neq \emptyset\}. \quad (\text{F.14})$$

We observe that by Property 1 of the moment functions

$$P[I_n^+ \neq \emptyset] = 1 - P[I_n^+ = \emptyset] \quad (\text{F.15})$$

$$= 1 - P[I_n^- \neq \emptyset] \quad (\text{F.16})$$

$$= 1 - \prod_{i=1}^n (1 - P[g(\mathbf{X}_i; t) \geq 0 \forall t \in [\underline{t}, \bar{t}]]) \quad (\text{F.17})$$

$$\geq 1 - \prod_{i=1}^n (1 - P[X_i \in [\underline{t}, \bar{t}] \times [\underline{t}, \bar{t}]]) \quad (\text{F.18})$$

$$\geq 1 - \prod_{i=1}^n (1 - c_2 (\bar{t} - \underline{t})^2) \quad (\text{F.19})$$

$$= 1 - (1 - c_2 (\bar{t} - \underline{t})^2)^n \rightarrow 1 \quad (\text{F.20})$$

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

$$\lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_2(c_2)} P[I_n^+ \neq \emptyset] \geq 1, \quad (\text{F.21})$$

which is the desired result.  $\square$

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

**Lemma F.4.** Given  $c_2 \in (0, \frac{1}{(\bar{t} - \underline{t})^2}]$ . 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

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

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

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

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

$$\geq c_2(\bar{t} - \underline{t}). \quad (\text{F.25})$$

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.  $\square$

## F.2 Existence and Uniqueness of $\tilde{\mathbf{p}}$ and $\hat{\mathbf{p}}$

Recall that  $\mathcal{H}_n = \{p_i, i = 1, \dots, n; \sum_{i=1}^n p_i = 1, p_i \geq 0, \forall i = 1, \dots, n\}$ , and denote the interior of this set by  $\mathcal{H}_n^\circ$ . Additionally, recall the feasible sets

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

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

**Lemma F.5.** 1. On the event  $\{\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\}$ , the random set

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

is nonempty and a singleton.

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

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

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.  $\square$

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 F.6.** Given  $c_1 \in (0, 1)$  and recall that  $\mathcal{P}_1(c_1)$  is defined in (3.3). Then

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

*Proof.* The proof proceeds by the direct method.

For large  $n$  and uniformly in  $\mathcal{P}_1(c_1)$ , Lemma F.1 implies that the event

$$\exists i \in \{1, 2, \dots, n\} \quad g(\mathbf{X}_i, t) < 0 \quad \forall t \in [\underline{t}, \bar{t}] \quad (\text{F.28})$$



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

$$\forall i \in \{2, 3, \dots, n\} : \sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}_i; t) \geq 0. \quad (\text{F.29})$$

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

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

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

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

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

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

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

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

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

the inequality (F.31) is equivalent to

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

Thus, for any  $p_1$  such that

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

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

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

**Corollary F.1.** Given  $c_1 \in (0, 1)$  and recall that  $\mathcal{P}_1(c_1)$  is defined in (3.3). Then

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

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

$$\inf_{P \in \mathcal{P}_1(c_1)} P[\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^c \neq \emptyset] \geq \inf_{P \in \mathcal{P}_1(c_1)} P[\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^c \neq \emptyset] \quad (\text{F.34})$$

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

### F.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

$$\{X_i^A, X_i^B\}_{i=1}^n \cap (\underline{t}, \bar{t}) \cup \{\underline{t}, \bar{t}\}, \quad (\text{F.35})$$

where  $t_{(0)} = \underline{t}$  and  $t_{(N)} = \bar{t}$ . Let  $d_H [\mathcal{T}_n, [\underline{t}, \bar{t}]]$  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 (3.4).

We have the following result

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

$$\lim_{n \rightarrow +\infty} \sup_{P \in \mathcal{P}_2(c_2)} P [d_H [\mathcal{T}_n, [\underline{t}, \bar{t}]] > \epsilon] = 0, \quad \text{and} \quad (\text{F.36})$$

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

where  $\mathcal{P}_2(c_2)$  is defined in (3.4).

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

$$d_H [\mathcal{T}_n, [\underline{t}, \bar{t}]] \leq \max_{j=1, \dots, N} (t_{(j+1)} - t_{(j)}) \leq \max_{K=A, B} \max_{j=0, \dots, N^K} (X_{(j+1)}^K - X_{(j)}^K) \quad (\text{F.38})$$

$$\leq \sum_{K=A, B} \max_{j=0, \dots, N^K} (X_{(j+1)}^K - X_{(j)}^K), \quad (\text{F.39})$$

where  $X_{(0)}^K = \underline{t}$ ,  $X_{(N^K+1)}^K = \bar{t}$  and  $\{X_{(j)}^K\}_{j=1}^{N^K}$  are the elements of  $\{t_{(j)}\}_{j=0}^N$  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  $U_{(j)}^K = F_K(X_{(j)}^K)$  for  $j = 0, \dots, N^K$ , the spacings can be expressed as

$$X_{(j+1)}^K - X_{(j)}^K = \frac{F_K^{-1}(U_{(j+1)}^K) - F_K^{-1}(U_{(j)}^K)}{U_{(j+1)}^K - U_{(j)}^K} (U_{(j+1)}^K - U_{(j)}^K) \quad (\text{F.40})$$

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

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 F.4

shows that the marginal densities on the interval  $[\underline{t}, \bar{t}]$  are bounded from below by  $c_2(\bar{t} - \underline{t})$ . Because  $\{X_{(j)}^K\}_{j=1}^{N^K}$  are the elements of  $\{t_{(j)}\}_{j=0}^N$ , they can only take values in the interval  $[\underline{t}, \bar{t}]$ . Hence,

$$\frac{U_{(j+1)}^K - U_{(j)}^K}{f_K \left( F_K^{-1} \left( aU_{(j+1)}^K + (1-a)U_{(j)}^K \right) \right)} \leq \frac{U_{(j+1)}^K - U_{(j)}^K}{c_2(\bar{t} - \underline{t})} \quad (\text{F.42})$$

So that the spacing  $X_{(j+1)}^K - X_{(j)}^K$  is bounded from above by  $\frac{U_{(j+1)}^K - U_{(j)}^K}{c_2(\bar{t} - \underline{t})}$ .

In consequence,

$$\max_{j=0, \dots, N^K} (X_{(j+1)}^K - X_{(j)}^K) \leq (c_2(\bar{t} - \underline{t}))^{-1} \max_{j=0, \dots, N^K} (U_{(j+1)}^K - U_{(j)}^K) \quad K = A, B, \quad (\text{F.43})$$

$$\leq (c_2(\bar{t} - \underline{t}))^{-1} \max_{i=1, \dots, n} (U_{(i+1)}^K - U_{(i)}^K) \quad K = A, B. \quad (\text{F.44})$$

where  $U_{(i)}^K = F_K \left( X_{(i)}^K \right)$  for  $i = 1, \dots, n$ . That is, the maximal spacings

$$\max_{j=0, \dots, N^K} (X_{(j+1)}^K - X_{(j)}^K) \quad K = A, B,$$

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

$$d_H [\mathcal{T}_n, [\underline{t}, \bar{t}]] \leq (c_2(\bar{t} - \underline{t}))^{-1} \sum_{K=A, B} \max_{i=1, \dots, n} (U_{(i+1)}^K - U_{(i)}^K), \quad (\text{F.45})$$

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

$$P \left[ \max_{i=1, \dots, n} (U_{(i+1)}^K - U_{(i)}^K) \leq \frac{\log n - \log \log \log n - \log 2}{n} \text{ infinitely often} \right] = 1 \quad (\text{F.46})$$

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

$$\sup_{P \in \mathcal{P}_2(c_2)} P [d_H [\mathcal{T}_n, [\underline{t}, \bar{t}]] > \epsilon] \leq \sup_{P \in \mathcal{P}_2(c_2)} P \left[ \sum_{K=A, B} \frac{\max_{i=1, \dots, n} (U_{(i+1)}^K - U_{(i)}^K)}{c_2(\bar{t} - \underline{t})} > \epsilon \right] \quad (\text{F.47})$$

$$\leq \epsilon^{-1} \sum_{K=A, B} E \left[ \frac{\max_{i=1, \dots, n} (U_{(i+1)}^K - U_{(i)}^K)}{c_2(\bar{t} - \underline{t})} \right] \quad (\text{F.48})$$

$$\leq \frac{2}{c_2(\bar{t} - \underline{t})\epsilon} \left( \frac{\log n - \log \log \log n - \log 2}{n} \right), \quad (\text{F.49})$$

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

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

□

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

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

$$\max_{t \in [\underline{t}, \bar{t}]} \sum_{i=1}^n \hat{p}_i g(\mathbf{X}_i; t) \quad (\text{F.50})$$

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

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

$$q'_n = \max \left\{ \sum_{\ell=1}^{s-1} \binom{s-1}{\ell} \left[ \sum_{i=1}^n \hat{p}_i (\bar{t} - X_i^A)^{s-1-\ell} 1 [X_i^A \leq \bar{t}] \right], 1 \right\}. \quad (\text{F.52})$$

The result is the following.

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

$$\lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_1(c_1) \cap \mathcal{P}_2(c_2)} P \left[ \max_{t \in [\underline{t}, \bar{t}]} \sum_{i=1}^n \hat{p}_i g(\mathbf{X}_i; t) \leq q_n \right] = 1. \quad (\text{F.53})$$

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

$$\lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_1(c_1) \cap \mathcal{P}_2(c_2)} P \left[ \max_{t \in [\underline{t}, \bar{t}]} \sum_{i=1}^n \hat{p}_i g(\mathbf{X}_i; t) \leq \frac{2q'_n \max_{j=0, \dots, N} (t_{(j+1)} - t_{(j)})}{(s-1)!} \right] = 1, \quad (\text{F.54})$$

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

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

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

On the event  $\{\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^{\circ} \neq \emptyset\}$ , Part 1 of Lemma F.5 shows that the extremum  $\hat{\mathbf{p}}$  exists and

is unique. Given  $t \in [\underline{t}, \bar{t}] - \mathcal{T}_n$ ,  $\exists j$  such that  $t_{(j)} < t < t_{(j+1)}$ . Then, for such a  $t$

$$g(\mathbf{X}_i; t) \leq \frac{(t_{(j+1)} - X_i^B)^{s-1} \mathbf{1}[X_i^B \leq t_{(j+1)}]}{(s-1)!} - \frac{(t_{(j)} - X_i^A)^{s-1} \mathbf{1}[X_i^A \leq t_{(j)}]}{(s-1)!} \quad (\text{F.55})$$

$$\begin{aligned} &= g(\mathbf{X}_i; t_{(j+1)}) + \frac{(t_{(j+1)} - X_i^A)^{s-1} \mathbf{1}[X_i^B \leq t_{(j+1)}]}{(s-1)!} \\ &\quad - \frac{(t_{(j)} - X_i^A)^{s-1} \mathbf{1}[X_i^A \leq t_{(j)}]}{(s-1)!} \end{aligned} \quad (\text{F.56})$$

$$\begin{aligned} &\leq \sum_{\ell=1}^{s-1} \binom{s-1}{\ell} (\bar{t} - X_i^A)^{s-1-\ell} \mathbf{1}[X_i^A \leq \bar{t}] (t_{(j+1)} - t_{(j)}) \\ &\quad + \mathbf{1}[t_{(j)} < X_i^A \leq t_{(j+1)}] (t_{(j+1)} - t_{(j)}) + g(\mathbf{X}_i; t_{(j+1)}). \end{aligned} \quad (\text{F.57})$$

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

$$\begin{aligned} \sum_{i=1}^n \dot{p}_i g(\mathbf{X}_i; t) &\leq \sum_{i=1}^n \dot{p}_i g(\mathbf{X}_i; t_{(j+1)}) + \frac{t_{(j+1)} - t_{(j)}}{(s-1)!} \sum_{i=1}^n \dot{p}_i \mathbf{1}[t_{(j)} < X_i^A \leq t_{(j+1)}] \\ &\quad + \frac{t_{(j+1)} - t_{(j)}}{(s-1)!} \left\{ \sum_{\ell=1}^{s-1} \binom{s-1}{\ell} \sum_{i=1}^n \dot{p}_i (\bar{t} - X_i^A)^{s-1-\ell} \mathbf{1}[X_i^A \leq \bar{t}] \right\} \end{aligned} \quad (\text{F.58})$$

$$\leq \frac{2q'_n}{(s-1)!} \max_{j=0, \dots, N} (t_{(j+1)} - t_{(j)}), \quad (\text{F.59})$$

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

$$\{\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^c \neq \emptyset\} \subseteq \left\{ \frac{2q'_n}{(s-1)!} \max_{j=0, \dots, N} (t_{(j+1)} - t_{(j)}) \right\}, \quad (\text{F.60})$$

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

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

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

$$\max_{j=0, \dots, N} (t_{(j+1)} - t_{(j)}) \leq (c_2(\bar{t} - \underline{t}))^{-1} \sum_{K=A, B} \max_{i=1, \dots, n} (U_{(i+1)}^K - U_{(i)}^K). \quad (\text{F.61})$$

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

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

yields

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

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

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

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

## F.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  $\dot{\mathbf{p}}$ . As the vector  $\dot{\mathbf{p}}$  is not necessarily a member of  $\mathcal{H}^0(\mathbf{X})$ , the construction is such that  $\check{\mathbf{p}}$  is nearby to  $\dot{\mathbf{p}}$  and will be essential in proving Part 2 Theorem 3.1 using Lemma F.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  $\check{\mathbf{p}} = \dot{\mathbf{p}} + \rho q_n \mathbf{d}$ , where  $\mathbf{d} \in \mathbb{R}^n$  and  $\rho > 0$ , and  $q_n$  is defined in (F.51) in the previous subsection. Naturally, we need to impose restrictions on the vector  $\mathbf{d}$  and  $\rho$  that ensures  $\check{\mathbf{p}} \in \mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ$ . Let  $\underline{\delta}$  and  $\bar{\delta}$  satisfy  $0 < \underline{\delta} < \dot{p}_{(1)}$  and  $0 < \bar{\delta} < 1 - \dot{p}_{(n)}$ , where  $\dot{p}_{(1)} = \min_{i=1,\dots,n} \dot{p}_i$  and  $\dot{p}_{(n)} = \max_{i=1,\dots,n} \dot{p}_i$ . Consider the following set of vectors:

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

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

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  $\bar{\delta}$  that satisfies the conditions  $0 < \underline{\delta} < \dot{p}_{(1)}$  and  $0 < \bar{\delta} < 1 - \dot{p}_{(n)}$ . The second part shows that with additional conditions on  $\underline{\delta}$  and  $\bar{\delta}$ , we can construct  $\check{\mathbf{p}} \in \mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ$  which is close to the vector  $\dot{\mathbf{p}}$  in a particular way so that we can use it Steps 2 and 4 of the proof for Part 2 of Theorem 3.1.

**Lemma F.9.** Suppose that the event  $\{I_n^- \neq \emptyset\}$  occurs, where  $I_n^-$  is given by (F.1). Furthermore, let  $\underline{\delta}$  and  $\bar{\delta}$  satisfy  $0 < \underline{\delta} < \dot{p}_{(1)}$  and  $0 < \bar{\delta} < 1 - \dot{p}_{(n)}$ . Also let the set  $D_n(\underline{\delta}, \bar{\delta})$  be defined as in (F.63). Then the following statements hold.

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

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

$$\rho q_n |d_i| \leq \frac{\dot{p}_{(1)} - \underline{\delta}}{n^2} \forall i \implies \|\rho q_n \mathbf{d}\| \leq \frac{\dot{p}_{(1)}}{2n^{3/2}}. \quad (\text{F.64})$$

*Proof.* The proof proceeds by the direct method.

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

$$\{I_n^- \neq \emptyset\} \subseteq \{\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^c \neq \emptyset\} \subseteq \{\check{\mathbf{p}} \text{ exists and is unique}\}. \quad (\text{F.65})$$

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  $\sum_{i=1}^n \check{p}_i g(\mathbf{X}_i; t) < 0 \quad \forall t \in [\underline{t}, \bar{t}]$ . Using Property 1 of the moment functions the  $d_i$  can be chosen so that they obey the following sign restrictions:

$$\text{sign}(d_i) = \begin{cases} < 0, & \text{if } g(\mathbf{X}_i, t) \geq 0 \forall t \in [\underline{t}, \bar{t}] \\ > 0, & \text{if } g(\mathbf{X}_i, t) < 0 \forall t \in [\underline{t}, \bar{t}]. \end{cases} \quad (\text{F.66})$$

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

$$\sum_{i=1}^n d_i g(\mathbf{X}_i; t) < 0 \quad \forall t \in [\underline{t}, \bar{t}]. \quad (\text{F.67})$$

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

In consequence,  $\forall t \in [\underline{t}, \bar{t}]$

$$\sum_{i=1}^n \check{p}_i g(\mathbf{X}_i; t) = \sum_{i=1}^n \dot{p}_i g(\mathbf{X}_i; t) + \rho q_n \sum_{i=1}^n d_i g(\mathbf{X}_i; t) \quad (\text{F.68})$$

$$\leq q_n - \rho \epsilon_n(\mathbf{d}) q_n, \quad (\text{F.69})$$

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

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

Furthermore, observe that

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

$$\check{p}_i \leq \bar{\delta} \forall i \iff d_i \leq \frac{1 - \bar{\delta} - \dot{p}_i}{\rho q_n} \forall i \quad \text{and} \quad \frac{1 - \bar{\delta} - \dot{p}_i}{\rho q_n} > 0 \forall i \iff \bar{\delta} < 1 - \dot{p}_{(n)}. \quad (\text{F.72})$$

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

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

$$\rho q_n |d_i| \leq \frac{\dot{p}_{(1)} - \underline{\delta}}{n^2} = \frac{\dot{p}_{(1)}}{2n^2} \forall i. \quad (\text{F.73})$$

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

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

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

$$\lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_1(c_1) \cap \mathcal{P}_2(c_2)} P[V_n] = 1. \quad (\text{F.74})$$

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

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

which concludes the proof.  $\square$

## F.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  $C([\underline{t}, \bar{t}])$ , where  $C([\underline{t}, \bar{t}])$  denotes the space of continuous functions  $\gamma : [\underline{t}, \bar{t}] \rightarrow \mathbb{R}$  with sup-norm. The space  $C([\underline{t}, \bar{t}])$  is a Banach space



and its dual is the space of finite signed measures on  $([\underline{t}, \bar{t}], \mathcal{B})$ , where  $\mathcal{B}$  is the Borel sigma algebra of  $[\underline{t}, \bar{t}]$ , with scalar product of  $\mu \in C([\underline{t}, \bar{t}])^*$  and  $\gamma \in C([\underline{t}, \bar{t}])$  given by the integral

$$\int_{[\underline{t}, \bar{t}]} \gamma(t) d\mu(t). \quad (\text{F.75})$$

thus, the SIP problem has the following Lagrangian function:

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

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

The Karush-Kuhn-Tucker conditions are

$$-\frac{1}{p_i} = \lambda + n \int_{[\underline{t}, \bar{t}]} g(\mathbf{X}_i; t) d\mu(t) \quad i = 1, 2, \dots, n \quad (\text{F.77})$$

$$\sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq 0 \quad \forall t \in [\underline{t}, \bar{t}], \quad \sum_{i=1}^n p_i = 1 \quad (\text{F.78})$$

$$\text{supp}(\mu) \subseteq \Delta(\tilde{\mathbf{p}}), \quad (\text{F.79})$$

where  $\Delta(\tilde{\mathbf{p}}) = \{t \in [\underline{t}, \bar{t}] : \sum_{i=1}^n \tilde{p}_i g(\mathbf{X}_i; t) = 0\}$  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 F.6 implies that the *Strong Slater Condition* of Mordukhovich and Nghia (2013) holds with probability tending to unity, with uniformity over sets of probabilities of the form  $\mathcal{P}_1(c_1)$ . In consequence, there exists a (positive) Borel measure  $\mu \in C([\underline{t}, \bar{t}])^*$  that solve the Karush-Kuhn-Tucker conditions (F.77) - (F.79) along with  $\tilde{\mathbf{p}}$ .

We have the following result.

**Lemma F.10.** For each  $c_1 \in (0, 1)$  recall that  $\mathcal{P}_1(c_1)$  is defined in (3.3). Let  $A_n$  denote the event that there exists a neighborhood of  $\tilde{\mathbf{p}}$ ,  $U$ , and a constant  $K_0 > 0$  such that

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

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

$$\lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_1(c_1)} P[A_n] = 1. \quad (\text{F.81})$$

*Proof.* The proof proceeds by the direct method. Given  $c_1 \in (0, 1)$ , we observe that by Lemma F.6 and Part 1 of Lemma F.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,

$$h(\mathbf{p}) - h(\tilde{\mathbf{p}}) = \mathbf{d}^T \nabla h(\tilde{\mathbf{p}}) + \frac{1}{2} \mathbf{d}^T \nabla^2 h(\tilde{\mathbf{p}}) \mathbf{d}, \quad (\text{F.82})$$

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 (F.77) - (F.79). Observe that

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

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

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

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

$$U = \{\mathbf{p} : \in \mathbb{R}^n : \|\mathbf{p} - \tilde{\mathbf{p}}\| \leq \epsilon\}. \quad (\text{F.85})$$

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  $i = 1, \dots, n$

$$\dot{p}_i^2 = (\dot{p}_i - \tilde{p}_i + \tilde{p}_i)^2 = (\dot{p}_i - \tilde{p}_i)^2 + \tilde{p}_i^2 + 2\tilde{p}_i(\dot{p}_i - \tilde{p}_i) \quad (\text{F.86})$$

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

$$\leq \epsilon^2 + 5. \quad (\text{F.88})$$

In consequence,

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

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

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

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

$$\{\mathcal{H}^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\} \subseteq A_n, \quad (\text{F.91})$$

which implies that

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

This concludes the proof.  $\square$

## G Intermediate Technical Results for Theorems 4.1 and 4.2

This section presents intermediate technical results that are used in the proofs of Theorems 4.1 and 4.2. Subsection G.1 presents a technical result that is a consequence of Definition 4.1. Section G.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 4.1. Finally, Section G.3 establishes the uniform consistency of the proposed empirical likelihood estimator of the moments from the discretized problem, under  $H_0$ .

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

$$s \max \{|\underline{t}|, |\bar{t}|\}^{s-1} (|X^B|^{s-1} + |X^A|^{s-1}). \quad (\text{G.1})$$

In consequence, Condition (iv) in Assumption 2.1 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.

### G.1 A Technical Lemma for Part 4 of Theorem 4.1

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

$$\lim_{n \rightarrow +\infty} \sup_{P \in \mathcal{P}_0^s(c_1, c_2, c_3) \cap \mathcal{P}_{000}} P \left[ \hat{P}_n = \hat{P}_n \right] \leq \frac{1}{2} \quad \text{and} \quad (\text{G.2})$$

$$\lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_0^s(c_1, c_2, c_3) \cap \mathcal{P}_{000}} P \left[ \Delta(\hat{P}_n) \neq \emptyset \right] \geq \frac{1}{2}, \quad (\text{G.3})$$

where  $\mathcal{P}_0^s(c_1, c_2, c_3)$  is defined in Definition 4.1 and

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

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

$$P \left[ \hat{P}_n = \hat{P}_n \right] = P \left[ E_{\hat{P}_n} [g(\mathbf{X}; t)] \leq 0 \forall t \in \mathcal{T}_n \right] \quad (\text{G.4})$$

$$= P \left[ E_{\hat{P}_n} [g(\mathbf{X}; t)] \leq 0 \forall t \in [\underline{t}, \bar{t}] \right] \quad (\text{G.5})$$

$$\leq P \left[ E_{\hat{P}_n} [g(\mathbf{X}; t)] \leq 0 \forall t \in \Delta(P) \right] \quad (\text{G.6})$$

$$\leq P \left[ E_{\hat{P}_n} [g(\mathbf{X}; t')] \leq 0 \forall t' \in \Delta(P) \right] \quad (\text{G.7})$$

$$= P \left[ \sqrt{n} E_{\hat{P}_n} [g(\mathbf{X}; t')] / E_P [g^2(\mathbf{X}; t')] \leq 0 \forall t' \in \Delta(P) \right] \quad (\text{G.8})$$

$$\leq \left| P \left[ \sqrt{n} \frac{E_{\hat{P}_n} [g(\mathbf{X}; t')]}{\sqrt{E_P [g^2(\mathbf{X}; t')]} } \leq 0; t' \in \Delta(P) \right] - \Phi(0) \right| + \Phi(0). \quad (\text{G.9})$$

We can apply the Berry-Esseen Theorem to the random variable  $\{g(\mathbf{X}_i; t')\}_{i=1}^n$  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,

$$\left| P \left[ \frac{\sqrt{n} E_{\hat{P}_n} [g(\mathbf{X}; t')]}{\sqrt{E_P [g^2(\mathbf{X}; t')]} } \leq 0 \right] - \Phi(0) \right| \leq \sup_{u \in \mathbb{R}} \left| P \left[ \frac{\sqrt{n} E_{\hat{P}_n} [g(\mathbf{X}; t')]}{\sqrt{E_P [g^2(\mathbf{X}; t')]} } \leq u \right] - \Phi(u) \right| \quad (\text{G.10})$$

$$\leq \frac{C_0 E_P |g(\mathbf{X}; t')|^3}{\sqrt{n} (E_P [g^2(\mathbf{X}; t')])^{3/2}} \quad (\text{G.11})$$

$$\leq \frac{C_0 2^3}{\sqrt{n} c_3^{3/2}}, \quad (\text{G.12})$$

where  $C_0$  is an absolute constant. Hence,

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

For the case  $s > 1$ , we follow steps identical to those above for  $s = 1$ , except that by Lemma F.7, the steps hold for large  $n$  as  $\mathcal{T}_n$  converges with uniformity (over  $\mathcal{P}_2(c_2)$ ) to  $[\underline{t}, \bar{t}]$  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 2.1. 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 (G.1), we set the following primitives in the notation of his paper

$$\tau'_k = n^{1/e}, -\tau_k = -n^{1/e} \quad \forall k = 1, \dots, n, \quad \text{where } e > 6(s-1), \quad (\text{G.14})$$

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

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

to deduce that

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

Hence,

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

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

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

and

$$\frac{b'}{nc_3} = \frac{\sup_{P \in \mathcal{P}_0^s(c_1, c_2, c_3) \cap \mathcal{P}_{000}} E_P \left[ \sum_{K=A, B} |X^K|^{2(s-1)} 1_{[|X^K| > n^{1/e}]}\right]}{c_3} \quad (\text{G.21})$$

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

since  $\sup_{P \in \mathcal{P}} E_P [ |X^K|^{2(s-1)} ] < +\infty$  for  $K = A, B$  by Condition (iv) of Assumption 2.1. This proves the limit (G.2).

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

$$\begin{aligned} \lim_{n \rightarrow +\infty} \sup_{P \in \mathcal{P}_0^s(c_1, c_2, c_3) \cap \mathcal{P}_{000}} P \left[ \Delta \left( \dot{P}_n \right) = \emptyset \right] &\leq \lim_{n \rightarrow +\infty} \sup_{P \in \mathcal{P}_0^s(c_1, c_2, c_3) \cap \mathcal{P}_{000}} P \left[ \dot{P}_n = \hat{P}_n \right] \\ &\leq \frac{1}{2}, \end{aligned}$$

which implies

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

□

## G.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 3. This optimization problem has the following Lagrangian function:

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

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

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

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

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

In classical optimization theory, the existence of  $\hat{\lambda}$  and  $\hat{\mu}_t$  for  $t \in \mathcal{T}_n$  that solve KKT conditions along with  $\hat{\mathbf{p}}$  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

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

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

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

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

*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 F.1 and Part 2 of F.6 imply that  $\hat{\mathbf{p}}$  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

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

Lemma F.1 implies the occurrence of the event  $\{\exists i : g(\mathbf{X}_i, t) < 0 \quad \forall t \in \mathcal{T}_n\}$  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  $\mathcal{S}_n$  converges to one uniformly in  $\mathcal{P}_1(c_1)$ .  $\square$

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

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

where  $\{\hat{\mu}_t = 0, \forall t \in \mathcal{T}_{N(n)} - \Delta(\hat{P}_n)\}$  and  $\{\hat{\mu}_t \geq 0, \forall t \in \Delta(\hat{P}_n)\}$ . 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

$$\Lambda_n(\hat{P}_n) = \left\{ \hat{\mu}_t \quad t \in \Delta(\hat{P}_n) \text{ that satisfy (G.24) - (G.26)} \right\}. \quad (\text{G.32})$$

Next, we focus on the large-sample properties of the multipliers in  $\Lambda_n(\dot{P}_n)$ , under  $H_0$ . Let  $w \in \mathbb{Z}_+ \cup \{+\infty\}$ , and define the Banach spaces, as indexed by  $w$ ,

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

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

**Lemma G.3** (Asymptotic Bound for Lagrange Multipliers).

Let  $\Delta(\dot{P}_n)$  be given by (G.28) and  $\omega_n = |\Delta(\dot{P}_n)|$ . 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  $c_3 \in (0, +\infty)$ ,

1.  $\lim_{n \rightarrow +\infty} \inf_{P \in \mathcal{P}_0 \cap \mathcal{P}_1(c_1)} P \left[ \Delta(\dot{P}_n) \subseteq \Delta(P) \right] = 1$ .
2. Denote the vector of Lagrange multipliers on the constraints binding constraints by  $\dot{\mu}$  and the  $l_{\omega_n}^1$  norm of the vector  $\dot{\mu}$  by  $\|\dot{\mu}\|_{l_{\omega_n}^1}$ . Then  $\forall \epsilon > 0$

$$\lim_{n \rightarrow +\infty} \sup_{P \in \mathcal{P}_0^s(c_1, c_2, c_3)} P \left[ \sup_{\dot{\mu} \in \Lambda_n(\dot{P}_n)} \|\dot{\mu}\|_{l_{\omega_n}^1} > \epsilon \right] = 0, \quad \text{and} \quad (\text{G.34})$$

$$\sup_{\dot{\mu} \in \Lambda_n(\dot{P}_n)} \|\dot{\mu}\|_{l_{\omega_n}^1} = O_P(n^{-1/2}) \quad (\text{G.35})$$

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

*Proof.*

1. We show this result using proof by contrapositive, that is, we show that for large  $n$ ,

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

Given  $c_1 \in (0, 1)$ , for large enough  $n$  and uniformly in  $\mathcal{P}_0 \cap \mathcal{P}_1(c_1)$  Corollary F.1 and Part 2 of F.6 imply that  $\dot{p}$  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

$$\sum_{i=1}^n \dot{p}_i' g(X_i; t) \leq \frac{1}{n} \sum_{i=1}^n g(X_i; t) = \frac{1}{n} \sum_{i=1}^n g(X_i; t) - E_{P_0} [g(X; t)] + E_{P_0} [g(X; t)] \quad (\text{G.36})$$

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

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

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

$$\sum_{i=1}^n \dot{p}_i' g(X_i; t) < 0$$

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

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

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

Thus, our approach in the proof will be to construct an upper bound on  $\sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1}$  on the event  $\{\Delta(\dot{P}_n) \neq \emptyset\}$  that is  $o_p(1)$  uniformly in  $\mathcal{P}_0^s(c_1, c_2, c_3)$ .

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

$$\Delta(\dot{P}_n) = \{t_1, t_2, \dots, t_{\omega_n}\}. \quad (\text{G.38})$$

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

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

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

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

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

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

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

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

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

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

This implies that  $\sum_{j=1}^{\omega_n} \theta_j = 1$ . The system of equations defined by (G.40) for all  $t \in \Delta(\dot{P}_n)$  can be written in the following form

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



Define the quantity  $Y_i = (\hat{\boldsymbol{\mu}})^T \mathbf{g}_i$ . Using the manipulation  $\frac{1}{1+Y_i} = 1 - \frac{Y_i}{1+Y_i}$  and the fact that  $(\hat{\boldsymbol{\mu}})^T \mathbf{g}_i = \mathbf{g}_i^T \hat{\boldsymbol{\mu}}$  in equation (G.42) gives

$$\begin{aligned} \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i \left( 1 - \frac{\mathbf{g}_i^T \hat{\boldsymbol{\mu}}}{1 + Y_i} \right) \right) &= 0 \\ \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i \right) &= \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{g}_i \mathbf{g}_i^T \hat{\boldsymbol{\mu}}}{1 + Y_i} \right) \\ \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i \right) &= \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{g}_i \mathbf{g}_i^T \|\hat{\boldsymbol{\mu}}\| \boldsymbol{\theta}}{1 + Y_i} \right) \\ \therefore \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i \right) &= \|\hat{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{g}_i \mathbf{g}_i^T}{1 + Y_i} \right) \boldsymbol{\theta} \end{aligned} \quad (\text{G.43})$$

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

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

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

$$Y_{max} = \max_i \sum_{j=1}^{\omega_n} \hat{\mu}_j |g(\mathbf{X}_i; t_j)| \quad (\text{G.44})$$

$$\leq \sum_{j=1}^{\omega_n} \hat{\mu}_j \max_i |g(\mathbf{X}_i; t_j)| \quad (\text{G.45})$$

$$= \|\hat{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} s \max \{|\underline{t}|, |\bar{t}|\}^{s-1} \sum_{K=A,B} \max_i |X_i^K|^{s-1}, \quad (\text{G.46})$$

where we used the envelope function (G.1) to bound the moment functions, uniformly in  $t \in [\underline{t}, \bar{t}]$ .

Let  $X_{max} = s \max \{|\underline{t}|, |\bar{t}|\}^{s-1} \sum_{K=A,B} \max_i |X_i^K|^{s-1}$ , and consider

$$\begin{aligned} \|\hat{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \left( \boldsymbol{\theta}^T \hat{\Sigma}_{\Delta(\hat{P}_n)} \boldsymbol{\theta} \right) &= \|\hat{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \left( \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i \mathbf{g}_i^T \right) \boldsymbol{\theta} \right) \\ &\leq \|\hat{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \left( \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{g}_i \mathbf{g}_i^T}{1 + Y_i} \right) \boldsymbol{\theta} \right) (1 + Y_{max}) \\ &\leq \|\hat{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \left( \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{g}_i \mathbf{g}_i^T}{1 + Y_i} \right) \boldsymbol{\theta} \right) (1 + X_{max} \|\hat{\boldsymbol{\mu}}\|_{l_{\omega_n}^1}) \\ \therefore \|\hat{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \left( \boldsymbol{\theta}^T \hat{\Sigma}_{\Delta(\hat{P}_n)} \boldsymbol{\theta} \right) &\leq \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i \right) (1 + X_{max} \|\hat{\boldsymbol{\mu}}\|_{l_{\omega_n}^1}) \end{aligned} \quad (\text{G.47})$$

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

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

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

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

The last line follows from the fact that  $\sum_{j=1}^{\omega_n} \theta_j = 1$  by construction. The upper bound given by equation (G.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 (G.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  $\boldsymbol{\theta}^T (n^{-1} X_{max} \sum_{i=1}^n \mathbf{g}_i)$ . We will show that it is also  $o_P(1)$  uniformly in  $\mathcal{P}_0^s(c_1, c_2, c_3)$ . We have

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

Next, apply Lemma 11.2 of Owen (2001) and Theorem 2.8.2 of van der Vaart and Wellner (1996) to  $X_{max}$  and

$$\sup_{t \in [\underline{t}, \bar{t}]} \left| \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i; t) - E_P[g(\mathbf{X}; t)] \right|,$$

respectively, to deduce that  $X_{max} = o(n^{1/2})$  and

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

Therefore, the right side of (G.50) is  $o(n^{1/2})O_P(n^{-1/2}) = o_P(1)$ , uniformly in  $\mathcal{P}$ . In consequence,  $\boldsymbol{\theta}^T(n^{-1}X_{max} \sum_{i=1}^n \mathbf{g}_i) = o_P(1)$  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  $\Delta(\dot{P}_n) \subset \Delta(P)$ , with probability approaching unity uniformly over  $\mathcal{P}_0 \cap \mathcal{P}_1(c_1)$ . Whether or not  $\Delta(P) \neq \emptyset$  the following manipulation holds. Because  $\mathcal{P}_0^s(c_1, c_2, c_3) \subseteq \mathcal{P}_0 \cap \mathcal{P}_1(c_1)$ , Definition 4.1 implies that  $\boldsymbol{\theta}^T \Sigma_{\Delta(\dot{P}_n)} \boldsymbol{\theta} \geq c_3 > 0$  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 (G.49), we can rewrite (G.48) as

$$\|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \leq \frac{o_P(1)}{c_3 + o_P(1)} \quad \forall \dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n), \quad \text{uniformly in } \mathcal{P}_0^s(c_1, c_2, c_3). \quad (\text{G.51})$$

Consequently,

$$\sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \leq \frac{o_P(1)}{c_3 + o_P(1)}, \quad \text{uniformly in } \mathcal{P}_0^s(c_1, c_2, c_3). \quad (\text{G.52})$$

Therefore,  $\sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} = o_P(1)$  uniformly in  $\mathcal{P}_0^s(c_1, c_2, c_3)$ . Finally, to show that  $\sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} = O_P(n^{-1/2})$  uniformly in  $\mathcal{P}_0^s(c_1, c_2, c_3)$ , first note that the expression on the right side of (G.48) has this property. So that

$$\sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \leq \frac{O_P(n^{-1/2})}{c_3 + o_P(1)} \quad \text{uniformly in } \mathcal{P}_0^s(c_1, c_2, c_3), \quad (\text{G.53})$$

which implies

$$\sqrt{n} \sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \leq \frac{O_P(1)}{c_3 + o_P(1)} \quad \text{uniformly in } \mathcal{P}_0^s(c_1, c_2, c_3). \quad (\text{G.54})$$

Hence,  $\sqrt{n} \sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1}$ , 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

$$\sqrt{n} \sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} = O_P(1) \quad \text{uniformly in } \mathcal{P}_0^s(c_1, c_2, c_3).$$

□

### G.3 Relationship Between $E_{\hat{P}_n}[g(\mathbf{X}, \cdot)]$ and $E_{\dot{P}_n}[g(\mathbf{X}, \cdot)]$

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

**Proposition G.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  $c_3 \in (0, +\infty)$ ,

$$\sup_{t \in [\underline{t}, \bar{t}]} |E_{\hat{P}_n} [g(\mathbf{X}, t)] - E_{\dot{P}_n} [g(\mathbf{X}, t)]| = O_P(n^{-1/2}) \quad \text{uniformly over } \mathcal{P}_0^s(c_1, c_2, c_3). \quad (\text{G.55})$$

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

$$\begin{aligned} |E_{\hat{P}_n} [g(\mathbf{X}, t)] - E_{\dot{P}_n} [g(\mathbf{X}, t)]| &= E_{\hat{P}_n} [g(\mathbf{X}, t)] - E_{\dot{P}_n} [g(\mathbf{X}, t)] & (\text{G.56}) \\ &= \sum_{i=1}^n \frac{1}{n} g(X_i; t) - \sum_{i=1}^n \dot{p}_i g(X_i; t) \\ &= \sum_{i=1}^n \left( \frac{1}{n} - \dot{p}_i \right) g(X_i; t) \\ &= \sum_{i=1}^n \frac{1}{n} \left( 1 - \frac{1}{1 + \sum_{j=1}^N \dot{\mu}_j g(X_i; t_j)} \right) g(\mathbf{X}_i, t) \\ &= \sum_{i=1}^n \frac{1}{n} \cdot \frac{g(\mathbf{X}_i, t) \sum_{j=1}^N \dot{\mu}_j g(X_i; t_j)}{1 + \sum_{j=1}^N \dot{\mu}_j g(X_i; t_j)} \\ &= \sum_{i=1}^n \dot{p}_i g(\mathbf{X}_i, t) \sum_{j=1}^N \dot{\mu}_j g(X_i; t_j). & (\text{G.57}) \end{aligned}$$

Now using the envelope function (G.1), we can obtain the following upper bound on the term (G.57):

$$s^2 \max \{|\underline{t}|, |\bar{t}|\}^{2(s-1)} \sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \sum_{i=1}^n \dot{p}_i \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2, \quad (\text{G.58})$$

where  $\Lambda_n(\dot{P}_n)$  is the set of Lagrange multipliers on  $\Delta(\dot{P}_n)$  defined in (G.32). Part 2 of Lemma G.3 establishes

$$\sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} = O_P(n^{-1/2}) \quad \text{uniformly over } \mathcal{P}_0^s(c_1, c_2, c_3). \quad (\text{G.59})$$

Thus, to deduce the desired result, we need to show that

$$\sum_{i=1}^n \dot{p}_i \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2 = O_P(1) \quad \text{uniformly over } \mathcal{P}_0^s(c_1, c_2, c_3). \quad (\text{G.60})$$

For each  $i = 1, \dots, n$ , we will apply the expansion  $\frac{1}{1+Y_i} = 1 - \frac{Y_i}{1+Y_i}$  to

$$\dot{p}_i = \frac{1}{n} \left( 1 + \sum_{j=1}^N \dot{\mu}_j g(X_i; t_j) \right)^{-1},$$

where  $Y_i = \sum_{j=1}^N \dot{\mu}_j g(X_i; t_j)$ , to deduce that the left side of (G.60) equals

$$\frac{1}{n} \sum_{i=1}^n \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2 - \frac{1}{n} \sum_{i=1}^n \left( \frac{\sum_{j=1}^N \dot{\mu}_j g(X_i; t_j)}{1 + \sum_{j=1}^N \dot{\mu}_j g(X_i; t_j)} \right) \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2. \quad (\text{G.61})$$

Next, apply Jensen's inequality to the second term in (G.61) to obtain the following upper bound on (G.61)

$$\frac{1}{n} \sum_{i=1}^n \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2 - \frac{\sum_j \dot{\mu}_j \frac{1}{n} \sum_{i=1}^n \left( g(\mathbf{X}_i, t_j) \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2 \right)}{1 + \sum_j \dot{\mu}_j \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i, t_j)}. \quad (\text{G.62})$$

By Condition (iv) of Assumption 2.1, the term  $\frac{1}{n} \sum_{i=1}^n \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2$  converges in probability to  $E_P \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2$ , 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 (G.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 (G.62) is bounded above by

$$\frac{\sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \max_{i,j} |g(\mathbf{X}_i, t_j)| \frac{1}{n} \sum_{i=1}^n \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2}{\left| 1 + \sum_j \dot{\mu}_j \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i, t_j) \right|}. \quad (\text{G.63})$$

We tackle the numerator and denominator of (G.63) separately. Using the envelope function (G.1), the numerator is bounded above by

$$s \max \{ |\underline{t}|, |\bar{t}| \}^{(s-1)} \sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \left( \sum_{K=A,B} \max_i |X_i^K|^{s-1} \right) \frac{1}{n} \sum_{i=1}^n \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2. \quad (\text{G.64})$$

By Condition (iv) of Assumption 2.1, an application of Lemma 11.2 of Owen (2001) and Part 2 of Lemma G.3 imply that

$$\sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \left( \sum_{K=A,B} \max_i |X_i^K|^{s-1} \right) = o_P(1) \text{ uniformly in } P \in \mathcal{P}_0^s(c_1, c_2, c_3).$$

Furthermore,  $\frac{1}{n} \sum_{i=1}^n \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2 = O_P(1)$  uniformly in  $P \in \mathcal{P}_0^s(c_1, c_2, c_3)$ , which implies that the term (G.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  $\sum_j \dot{\mu}_j \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i, t_j) = o_P(1)$  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

$$\left| \sum_j \hat{\mu}_j \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i, t_j) \right| \leq \sup_{\hat{\boldsymbol{\mu}} \in \Lambda_n(\hat{P}_n)} \|\hat{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \sup_{t \in \Delta(\hat{P}_n)} |E_{\hat{P}_n} [g(\mathbf{X}, t)]| \quad (\text{G.65})$$

$$= O_P(n^{-1/2}) \sup_{t \in \Delta(\hat{P}_n)} |E_{\hat{P}_n} [g(\mathbf{X}, t)]| \quad (\text{G.66})$$

$$\leq O_P(n^{-1/2}) \sup_{t \in \Delta(P)} |E_{\hat{P}_n} [g(\mathbf{X}, t)]| \quad (\text{G.67})$$

$$= O_P(n^{-1/2}) O_P(n^{-1/2}) = O_P(n^{-1}) = o_P(1) \quad (\text{G.68})$$

by Lemma G.3 and the Uniform Central Limit Theorem. This concludes the proof.  $\square$

## H Technical Lemmas for Theorems 5.1 and 5.2

### H.1 Theorem 5.1

This subsection presents two technical lemmas that are useful for proving Theorem 5.1. They are a consequence of the condition  $P \left[ \sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}_1; t) < 0 \right] > 0$  being true. The first lemma is similar to Lemma F.1, but we now do not constrain  $P$  to satisfy  $H_0$ .

**Lemma H.1.** Suppose  $P_0 \in \mathcal{P}$  and let  $I_n^-$  be given by (F.1). Then

$$\lim_{n \rightarrow \infty} P \left[ I_n^- \neq \emptyset \right] = 1.$$

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

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

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

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

$$= \left( 1 - P_0 \left[ \sup_{t \in [\underline{t}, \bar{t}]} g(\mathbf{X}_1; t) < 0 \right] \right)^n \rightarrow 0 \quad (\text{H.2})$$

$n \rightarrow +\infty$  by Condition (i) of Assumption 2.1.  $\square$

The second lemma concerns the existence and uniqueness of the constrained empirical likelihood probability vector  $\hat{\mathbf{p}}$ . Recall that

$$\mathcal{H}_n = \left\{ p_i, i = 1, \dots, n; \sum_{i=1}^n p_i = 1, p_i \geq 0, \forall i = 1, \dots, n \right\}$$

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 H.2.** Suppose  $P_0 \in \mathcal{P}$ . Then

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

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

$$\exists i \in \{1, 2, \dots, n\} \quad g(\mathbf{X}_i, t) < 0 \quad \forall t \in \mathcal{T}_n \quad (\text{H.3})$$

occurs with probability approaching 1, since  $\mathcal{T}_n \subset [\underline{t}, \bar{t}]$  for each  $n$ . The rest of the proof proceeds using steps similar to those in the proof of Lemma F.5; therefore, we omit them for brevity.  $\square$

## H.2 Theorems 5.2

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

$$\{\{\mathbf{X}_{i,j}, i = 1, \dots, n\}, n = 1, 2, \dots\}, \quad (\text{H.4})$$

where for each  $n$   $\{\mathbf{X}_{i,j}, i = 1, \dots, n\}$  is bivariate random sample form  $P_n$  that satisfies Assumption 5.1.

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 H.3.** Suppose that Assumption 5.1 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

$$\forall \epsilon > 0, \lim_{n \rightarrow +\infty} P_n \left[ \frac{\max_{1 \leq i \leq n} \sum_{K=A,B} |X_{i,n}|^{s-1}}{\sqrt{n}} \leq \epsilon \right] = 1, \quad (\text{H.5})$$

which implies the desired result.

Under Assumption 5.1 and  $\mathcal{P}$ , we have that

$$\sup_n E_{P_n} \left[ \sum_{K=A,B} |X_{i,n}|^{s-1} \right]^2 \leq \sup_{P \in \mathcal{P}} E_P \left[ \sum_{K=A,B} |X_i|^{s-1} \right]^2 < +\infty, \quad (\text{H.6})$$

holds. Then, for every  $\epsilon > 0$  Markov's inequality implies that

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^n P_j \left[ \left( \sum_{K=A,B} |X_{i,j}|^{s-1} \right)^2 > n\epsilon \right] < +\infty. \quad (\text{H.7})$$

As  $\sum_{j=1}^n P_j [(\sum_{K=A,B} |X_{i,j}|^{s-1})^2 > n\epsilon]$  is a convergent series of non-negative terms, it follows that

$$\lim_{n \rightarrow +\infty} P_n \left[ \left( \sum_{K=A,B} |X_{i,n}|^{s-1} \right)^2 > n\epsilon \right] = 0, \quad (\text{H.8})$$

holds. In consequence, the limit of the complementary probabilities satisfies

$$\lim_{n \rightarrow +\infty} P_n \left[ \frac{\sum_{K=A,B} |X_{i,n}|^{s-1}}{\sqrt{n}} \leq \epsilon \right] = 1, \quad (\text{H.9})$$

which implies that

$$\lim_{n \rightarrow +\infty} P_n \left[ \frac{\max_{1 \leq i \leq n} \sum_{K=A,B} |X_{i,n}|^{s-1}}{\sqrt{n}} \leq \sqrt{\epsilon} \right] = 1, \quad (\text{H.10})$$

holds. The limit (H.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  $\epsilon$ .  $\square$

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

**Lemma H.4.** Suppose that the conditions of Theorem 5.2 hold. Then

1.  $\lim_{n \rightarrow +\infty} P_n [I_n^- \neq \emptyset] = 1$ , where  $I_n^-$  is defined in (F.1).
2.  $\lim_{n \rightarrow +\infty} P_n [\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^c \neq \emptyset] = 1$ .
3.  $\lim_{n \rightarrow +\infty} P_n [\mathcal{S}_n] = 1$ , where  $\mathcal{S}_n$  is the event defined in (G.27).
4.  $\lim_{n \rightarrow +\infty} P_n [E_{\hat{P}_n} [g(\mathbf{X}; t)] \leq E_{\hat{P}_n} [g(\mathbf{X}; t)] \quad \forall t \in [\underline{t}, \bar{t}]] = 1$ .
5.  $\sqrt{n} \sup_{\dot{\mu} \in \Lambda_n(\hat{P}_n)} \|\dot{\mu}\|_{l_{\omega_n}^1} = O_{P_n}(1)$ .

*Proof.* Under the conditions of Theorem 5.2, the steps for proving parts 1 to 4 of this lemma are identical to their counterparts in Appendix G, 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

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



The proof will follow steps similar to those of part (i) of Lemma G.3. Proceeding by contraposition, we need to show that

$$t \notin C \implies t \notin \Delta(\dot{P}_n) \quad (\text{H.11})$$

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

$$E_{\dot{P}_n} [g(\mathbf{X}; t)] \leq E_{\hat{P}_n} [g(\mathbf{X}; t)] = E_{\hat{P}_n} [g(\mathbf{X}; t)] - E_{P_n} [g(\mathbf{X}; t)] + E_{P_n} [g(\mathbf{X}; t)]. \quad (\text{H.12})$$

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

$$E_{\hat{P}_n} [g(\mathbf{X}; t)] - E_{P_n} [g(\mathbf{X}; t)] = o_{P_n}(1). \quad (\text{H.13})$$

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 from above by the square-integrable envelope function (G.1) under the local alternatives. Thus for sufficiently large  $n$ , the inequality (H.12) simplifies to

$$E_{\dot{P}_n} [g(\mathbf{X}; t)] \leq H(t) < 0 \quad \text{as } n \rightarrow +\infty. \quad (\text{H.14})$$

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

Using the notation of Lemma G.3, and following identical steps to those up to the inequality (G.48), we have that

$$\|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \left[ \boldsymbol{\theta}^T \hat{\Sigma}_{\omega_n} \boldsymbol{\theta} - \boldsymbol{\theta}^T \left( \frac{X_{max}}{n} \sum_{i=1}^n \mathbf{g}_i \right) \right] \leq \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i \right) \quad \forall \dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n), \quad (\text{H.15})$$

where

$$X_{max} = s \max \{ |\underline{t}|, |\bar{t}| \}^{s-1} \sum_{K=A,B} \max_{1 \leq i \leq n} |X_{i,n}^K|^{s-1}, \quad (\text{H.16})$$

$$\mathbf{g}_i = [g(X_i; t_1), g(X_i; t_2), \dots, g(X_i; t_{\omega_n})]^T, \quad (\text{H.17})$$

$$\Delta(\dot{P}_n) = \{t_1, t_2, \dots, t_{\omega_n}\} \quad (\text{H.18})$$

and  $\boldsymbol{\theta} \in \mathbb{R}_+^{\omega_n}$  with  $\|\boldsymbol{\theta}\|_{l_{\omega_n}^1} = 1$ . Noting that

$$\boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i \right) = \sum_{j=1}^{\omega_n} \theta_j \left( \frac{1}{n} \sum_{i=1}^n g(X_i; t_j) - E_{P_n} [g(X; t_j)] \right) + \sum_{j=1}^{\omega_n} \theta_j (\delta(t_j) / \sqrt{n}) \quad (\text{H.19})$$

$$\leq \sup_{t \in [\underline{t}, \bar{t}]} |E_{\hat{P}_n} [g(\mathbf{X}; t)] - E_{P_n} [g(\mathbf{X}; t)]| + \sup_{t \in [\underline{t}, \bar{t}]} \delta(t) / \sqrt{n} = o_{P_n}(1) \quad (\text{H.20})$$

by the Uniform WLLN for triangular arrays of random variables that are row-wise IID and that

$$\begin{aligned}
\boldsymbol{\theta}^T \left( \frac{X_{max}}{n} \sum_{i=1}^n \mathbf{g}_i \right) &= \sum_{j=1}^{\omega_n} \theta_j \left( \frac{X_{max}}{n} \sum_{i=1}^n g(X_i; t_j) - E_{P_n} [g(X; t_j)] \right) \\
&\quad + \sum_{j=1}^{\omega_n} \theta_j (\delta(t_j) / \sqrt{n}) \\
&\leq X_{max} \sup_{t \in [\underline{t}, \bar{t}]} |E_{\hat{P}_n} [g(\mathbf{X}; t)] - E_{P_n} [g(\mathbf{X}; t)]| + \frac{X_{max}}{\sqrt{n}} \sup_{t \in [\underline{t}, \bar{t}]} \delta(t) \\
&= \frac{X_{max}}{\sqrt{n}} \sup_{t \in [\underline{t}, \bar{t}]} |E_{\hat{P}_n} [g(\mathbf{X}; t)] - E_{P_n} [g(\mathbf{X}; t)]| \\
&\quad + \frac{X_{max}}{\sqrt{n}} \sup_{t \in [\underline{t}, \bar{t}]} \delta(t) \\
&= o_{P_n}(1) O_{P_n}(1) + o_{P_n}(1) = o_{P_n}(1)
\end{aligned}$$

by Lemma H.3 and Theorem 2.8.9 in van der Vaart and Wellner (1996) (i.e., uniform CLT), we have that

$$\sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\hat{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \leq \frac{\boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i \right)}{\left[ \boldsymbol{\theta}^T \hat{\Sigma}_{\omega_n} \boldsymbol{\theta} - \boldsymbol{\theta}^T \left( \frac{X_{max}}{n} \sum_{i=1}^n \mathbf{g}_i \right) \right]} \quad (\text{H.21})$$

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

Hence,

$$\sqrt{n} \sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\hat{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \leq \frac{\sqrt{n} \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i \right)}{\left[ \boldsymbol{\theta}^T \hat{\Sigma}_{\omega_n} \boldsymbol{\theta} - \boldsymbol{\theta}^T \left( \frac{X_{max}}{n} \sum_{i=1}^n \mathbf{g}_i \right) \right]}. \quad (\text{H.22})$$

To conclude the proof, all we need to do is to show that the numerator on the right side of the inequality (H.22) is  $O_{P_n}(1)$ . Noting the inequality (H.20) above, we have that

$$\sqrt{n} \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i \right) \leq \sqrt{n} \sup_{t \in [\underline{t}, \bar{t}]} |E_{\hat{P}_n} [g(\mathbf{X}; t)] - E_{P_n} [g(\mathbf{X}; t)]| + \sup_{t \in [\underline{t}, \bar{t}]} \delta(t), \quad (\text{H.23})$$

where  $\sqrt{n} \sup_{t \in [\underline{t}, \bar{t}]} |E_{\hat{P}_n} [g(\mathbf{X}; t)] - E_{P_n} [g(\mathbf{X}; t)]| = O_{P_n}(1)$  by Theorem 2.8.9 in van der Vaart and Wellner (1996), and  $\sup_{t \in [\underline{t}, \bar{t}]} \delta(t)$  is finite by Part (iii) of Assumption 5.1, which implies the desired result. Therefore,

$$\sqrt{n} \sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\hat{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} = O_{P_n}(1),$$

which concludes the proof.  $\square$

The next result is the counterpart of Proposition G.1 under the sequence of local alternatives demarcated by Assumption 5.1.

**Proposition H.1.** Suppose that Assumption 5.1 holds. Then For each  $s \in \mathbb{Z}_+$ ,

$$\sup_{t \in [\underline{t}, \bar{t}]} |E_{\hat{P}_n} [g(\mathbf{X}, t)] - E_{\dot{P}_n} [g(\mathbf{X}, t)]| = O_{P_n}(n^{-1/2}). \quad (\text{H.24})$$

*Proof.* The proof follows steps identical to those in the proof of Proposition G.1 except that the limits are taken under the sequence of local alternatives. Firstly, we can follow the same steps to deduce that  $\sup_{t \in [\underline{t}, \bar{t}]} |E_{\hat{P}_n} [g(\mathbf{X}, t)] - E_{\dot{P}_n} [g(\mathbf{X}, t)]|$  is bounded above by

$$s^2 \max \{|\underline{t}|, |\bar{t}|\}^{2(s-1)} \sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} \sum_{i=1}^n \dot{p}_i \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2, \quad (\text{H.25})$$

Then Part 5 of Lemma H.4 implies that  $\sup_{\dot{\boldsymbol{\mu}} \in \Lambda_n(\dot{P}_n)} \|\dot{\boldsymbol{\mu}}\|_{l_{\omega_n}^1} = O_{P_n}(n^{-1/2})$ , holds, which implies we need to show that

$$\sum_{i=1}^n \dot{p}_i \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2 = O_{P_n}(1), \quad (\text{H.26})$$

holds, in order to conclude the proof.

To show that (H.26) holds, we can implement the same decomposition for this term and apply Jensen's inequality as in the proof of Proposition G.1 to show that it is bounded above by the expression (G.62), which we repeat here for convenience:

$$\frac{1}{n} \sum_{i=1}^n \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2 - \frac{\sum_j \dot{\mu}_j \frac{1}{n} \sum_{i=1}^n \left( g(\mathbf{X}_i, t_j) \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2 \right)}{1 + \sum_j \dot{\mu}_j \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i, t_j)}. \quad (\text{H.27})$$

Condition (iv) of Assumption 2.1 implies that

$$\frac{1}{n} \sum_{i=1}^n \left( \sum_{K=A,B} |X_i^K|^{s-1} \right)^2 = O_{P_n}(1).$$

Then we can tackle the denominator and numerator of the second term in (H.27) separately. For the numerator, we follow the same steps as in Proposition G.1 but use Lemma H.3 and Part 5 of Lemma H.4 instead of Lemma 11.2 of Owen (2001) and Part 2 of Lemma G.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 G.1 except that we replace the contact set  $\Delta(P)$  with the set  $C$  and use Part 5 of Lemma H.4. and the Theorem 2.8.9 of van der Vaart and Wellner (1996) instead of Part 2 of Lemma G.3 and the Uniform Central Limit Theorem, respectively, to deduce that

$$\sum_j \dot{\mu}_j \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i, t_j) = o_{P_n}(1),$$

under the sequence of local alternatives. This concludes the proof.  $\square$