Bootstrap Tests of Stochastic Dominance with Asymptotic Similarity on the Boundary[1]

Oliver Linton[2], Kyungchul Song[3] and Yoon-Jae Whang[4]

London School of Economics, University of Pennsylvania and Seoul National University

September 30, 2008

Abstract

We propose a new method of testing stochastic dominance which improves on existing tests based on bootstrap or subsampling. The method allows the prospects to be indexed by infinite as well as finite dimensional unknown parameters, so that the variables may be residuals from nonparametric and semiparametric models. The proposed bootstrap tests have asymptotic sizes that are exactly equal to the nominal level uniformly over the boundary points of the null hypothesis and are therefore valid over the whole null hypothesis. As our simulation results show, these characteristics of our tests lead to an improved power property in general. The improvement stems from the design of the bootstrap test whose limiting behavior mimicks the discontinuity of the original test's limiting distribution. It is also shown that under certain cases, the approximation error of the rejection probability can be controlled nearly up to the rate of $N^{-1/2}$, despite the discontinuous limit behavior of the test.

Key words and Phrases: Set estimation; Size of test; Similarity; Bootstrap; Subsampling.

JEL Classifications: C12, C14, C52.

[1] We thank Yanqin Fan for reading the earlier version of this draft carefully and offering numerous useful comments. All errors are ours.

[2] Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: o.linton@lse.ac.uk. Thanks to the ESRC and Leverhulme foundations for financial support.

[3] Department of Economics, University of Pennsylvania, 528 McNeil Building, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6297. Email: kysong@sas.upenn.edu

[4] Department of Economics, Seoul National University, Seoul 151-742, Korea. E-mail address: whang@snu.ac.kr. Thanks to the Korea Research Foundation for financial support.

1 Introduction

There has been a growth of interest in testing stochastic dominance relations among variables such as investment strategies and income distributions. The ordering by stochastic dominance covers a large class of utility functions and hence is more useful in general than the notions of partial orderings that are specific to a certain class of utility functions. However, the main difficulty in testing stochastic dominance lies in the complexity that arises in computing approximate critical values of the test. A stochastic dominance relation is a weak inequality relation between distribution functions or integral transforms of distribution functions. Unlike nonparametric or semiparametric tests that are based on the equality of functions, the convergence of test statistics of Kolmogorov-Smirnov type or Cramér-von Mises type is not uniform over the probabilities under the null hypothesis. Discontinuity of convergence arises precisely between the "interior points" of the null hypothesis and the "boundary points" of the null hypothesis, where, in the case of Kolmogorov-Smirnov type tests, boundary points indicate those probabilities under which all pairs of the distribution functions meet at least one point in the interior of the support and interior points indicate probabilities under which at least one pair of the distribution functions do not meet at any point in the interior of the support. In general, the boundary points do not coincide with the least favorable subset of null hypothesis, i.e., the set of probabilities under which all the pairs of competing distribution functions are equal. As noted by Linton, Maasoumi, and Whang (2005), henceforth LMW, the usual approach of naive or recentered bootstrap does not provide tests with asymptotically valid sizes uniformly over the probabilities under the null hypothesis; the asymptotic size of the tests are exact only on the least favorable points and invalid for other points under the null hypothesis. LMW proposed a subsampling method to obtain tests with exact asymptotic sizes over the boundary points and asymptotically valid for each probability under the null hypothesis. However, LMW do not discuss uniformity of their results.

This paper proposes a bootstrap-based test that has asymptotically valid size uniformly under the null hypothesis and asymptotically exact size on the boundary points. The method is based on bootstrap test statistics that are constructed to mimic the phenomenon of discontinuous convergence by using the estimated boundary points. This requires the estimation of a certain set, we call the "contact set"; set estimation is a topic of considerable recent interest in microeconometrics, see for example Moon and Schorfheide (2006), Chernozhukov, Hong, and Tamer (2007), and Fan and Park (2007). This estimation of a contact set can be viewed as a continuum version of the general moment selection procedure in Andrews and Soares (2007). See Hansen (2005) for a similar suggestion in testing predictive abilities among a finite number of forecasting models. For a similar, related bootstrap procedure, see Bugni (2008).

Our method of bootstrap has the following remarkable properties. First, it is asymptotically similar on the boundary. Second, more importantly, the bootstrap method provides asymptotically exact sizes. Furthermore, under certain cases, we can make the rejection probability under the null hypothesis converge to the true rejection probability only slightly less fast than the asymptotic approximation. This implies that the convergence of rejection probability can be made faster than that of subsampling.

The proposed test of stochastic dominance admits variables that contain unknown finite-dimensional and infinite-dimensional parameters as long as those parameters have consistent estimators. Hence the proposed test can be used to test stochastic dominance relations among variables in the form of regression errors from the semiparametric regression models such as single index restrictions or partially parametric regressions. For example, one can detect the influence of a certain factor upon the stochastic dominance relation by comparing test results with and without the factor of interest while controlling for other factors. In some situations, it is desirable to control for publicly observed variables. It is a potential concern for the researcher in such cases that the result of testing may rely sensitively on the regression specification chosen to control certain factors stochastically. Semiparametric specification of the regression can be useful in this situation.

We perform Monte Carlo simulations that compare three methods: recentered bootstrap, subsampling method, and the bootstrap on the boundary domain proposed in this paper. The results verify the superior performance of our method.

Testing stochastic dominance has drawn attention in the literature over the last decade. McFadden (1986), Klecan, McFadden and McFadden (1991), Kaur, Prakasa Rao, and Singh (1994), Anderson (1996) and Davidson and Duclos (1997, 2000) are among the early works that considered testing stochastic dominance. Barrett and Donald (2003) proposed a consistent bootstrap test that has an asymptotically exact size on the least favorable points for the special case where the prospects are mutually independent. LMW suggested a subsampling method that has asymptotically valid size on the boundary points and applies to variables that contain unknown finite-dimensional parameters and allows the prospects to be mutually dependent. Recent work in this area also includes, e.g., Horváth, Kokoszka, and Zikitis (2006), Davidson and Duclos (2006) and Bennett (2007).

In Section 2, we define the null hypothesis of stochastic dominance and introduce notations. In Section 3, we suggest test statistics and develop asymptotic theory both under the null hypothesis and local alternatives. Section 4 is devoted to the bootstrap procedure, explaining the method of obtaining bootstrap test statistics and establishing their asymptotic properties. In Section 5, we describe Monte Carlo simulation studies and discuss results from them. Section 6 concludes. All the technical proofs are relegated to the appendix.

2 Stochastic Dominance

2.1 The Null Hypothesis

Let $\{X_k\}_{k=1}^K$ be a set of continuous outcome variables which may, for example, represent different points in time or different regions. Let $F_k(x)$ be the distribution function of the k-th random variable X_k. Let $D_k^{(1)}(x) = F_k(x)$ and

$$D_k^{(s)}(x) = \int_{-\infty}^{x} D_k^{(s-1)}(t)dt.$$

Then we say that X_1 *stochastically dominates* X_2 *at order* s if $D_1^{(s)}(x) \leq D_2^{(s)}(x)$ for all x. This is equivalent to an ordering of expected utility over a certain class of utility functions $\mathcal{U}_s$, see LMW.

Let $D_{kl}^{(s)}(x) = D_k^{(s)}(x) - D_l^{(s)}(x)$ and define a one-sided Cramer-vov Mises type functional

$$c_s = \min_{k \neq l} \int_{\mathcal{X}} \max\left\{D_{kl}^{(s)}(x), 0\right\}^2 w(x)dx,$$

where w is a nonnegative integrable weighting function and $\mathcal{X}$ denotes a set contained in the union of the supports of X_k. The weighting function can be chosen appropriately depending on the interest of the researcher. For instance, suppose that X_k indicates household income for individuals in the k-th cohort, and the object of main interest is to compare the stochastic dominance between X_1 and X_2 for the invidiuals with income below the median μ of the whole population. In this situation, one may consider using $w(x) = 1\{x \leq \mu\}$. Using this weighting fucntion instead of $w(x) = 1$ generally increases the power of the test against alternatives such that there is no stochastic dominance when X is restricted to below μ. Throughout this paper, we do not assume that the set $\mathcal{X}$ is bounded.

The null hypothesis that this paper focuses on takes the following form:

$$H_0 : c_s = 0 \text{ vs. } H_1 : c_s > 0. \tag{1}$$

The null hypothesis represents the presence of a stochastic dominance relationship between a pair of variables in $\{X_k\}_{k=1}^K$. The alternative hypothesis corresponds to no such incidence.

An alternative way of formulating the problem is through a one-sided Kolmogorov-

Smirnov type functional. More specifically define

$$d_s = \min_{k \neq l} \sup_{x \in \mathcal{X}} D_{kl}^{(s)}(x) w(x).$$

In this case, the null hypothesis can be written as

$$H_0 : d_s \leq 0 \text{ vs. } H_1 : d_s > 0. \tag{2}$$

Obviously the null hypothesis $d_s \leq 0$ implies the null hypothesis $c_s = 0$. On the contrary, when $c_s = 0$, $D_{kl}^{(s)}(x) \leq 0$, $w(x)dx$-a.e., for some pair (k, l). Therefore, if the $D_k^{(s)}$'s are continuous functions, both the null hypotheses in (2) and (1) are equivalent, but in general, the null hypothesis in (2) is stronger than that in (1). However, for brevity, we do not pursue a formal development along the one-sided Kolmogorov-Smirnov functional or other functionals in this paper.

Of crucial importance in the sequel are the "contact" sets, a subset of $\mathcal{X}$ such that $D_k^{(s)}(x)$ and $D_l^{(s)}(x)$ are equal, i.e.,

$$B_{kl} \equiv \{x \in \mathcal{X} : D_{kl}^{(s)}(x) = 0\}. \tag{3}$$

These sets can be empty, can contain a countable number of isolated points, or can be a countable union of disjoint intervals.

2.2 Test Statistics and Asymptotic Theory

In many practical situations, the variable of interest is specified as a residual from a certain model. This arises in particular when data is limited and one may want to use a model to adjust for systematic differences. Common practice is to cut the data into subsets, say of families with two children, and then make comparisons across homogenous populations. When data are limited, this can be difficult and a modelling approach can overcome this difficulty. We suppose that the variables X_k depend on an unknown finite dimensional parameter $\theta_0 \in \mathbb{R}^{d_\Theta}$ and on an infinite dimensional parameter $\tau_0 \in \mathcal{T}$ (here, $\mathcal{T}$ is a totally bounded class of functions with respect to a certain metric), so that we write $X_k = X_k(\theta_0, \tau_0)$. For example, the variable X_k may be the residual from the partially parametric regression $X_k(\theta_0, \tau_0) = Y_k - Z_{1k}^\top \theta_0 - \tau_0(Z_{2k})$ or the single index framework $X_k(\theta_0, \tau_0) = Y_k - \tau_0(Z_{1k}^\top \theta_0)$. Generically we let $X_k(\theta, \tau)$ be specified as

$$X_k(\theta, \tau) = \varphi_k(W; \theta, \tau), \tag{4}$$

where W is a random vector in $\mathbf{R}^{d_W}$ and $\varphi_k(\cdot;\theta,\tau)$ is a real-valued function known up to the parameter $(\theta,\tau)\in\Theta\times\mathcal{T}$. In the example of $X_k(\theta_0,\tau_0)=Y_k-\tau_0(Z_{1k}^\top\theta_0)$, we take $W=(Y,Z)$ and $\varphi_k(w;\theta,\tau)=y_k-\tau(z_k^\top\theta),\ w=(y,z)$. This set up is more general than in LMW who allowed only linear regression.

We next specify the precise properties that we require of the data generating process. Let $B_{\Theta\times\mathcal{T}}(\delta)=\{(\theta,\tau)\in\Theta\times\mathcal{T}:||\theta-\theta_0||+||\tau-\tau_0||_\infty<\delta\}$. The norms $||\cdot||_{P,2}$ and $||\cdot||_\infty$ denote the usual $L_2(P)$-norm and the sup norm. We introduce a bounded weight function $q(x)$ (see Assumption 3(iii)) and define

$$h_x(\varphi)=\frac{(x-\varphi)^{s-1}1\{\varphi\le x\}q(x)}{(s-1)!}. \tag{5}$$

Let $N_{[]}(\varepsilon,\mathcal{T},||\cdot||_\infty)$ denote the bracketing number of $\mathcal{T}$ with respect to $||\cdot||_\infty$, i.e. the smallest number of ε- brackets that are needed to cover the space $\mathcal{T}$ (e.g. van der Vaart and Wellner (1996)). The conditions in Assumptions 1 and 2 below are concerned with the data generating process of W and the map φ_k. Let $\mathcal{P}$ be the collection of all the potential distributions of W that satisfy Assumptions 1-3 below.

Assumption 1 : (i) $\{W_i\}_{i=1}^N$ is a random sample.
(ii) $\log N_{[]}(\varepsilon,\mathcal{T},||\cdot||_\infty)\le C\varepsilon^{-d}$ for some $d\in(0,1]$.
(iii) For some $\delta>0$ and $s>0$, $\sup_{P\in\mathcal{P}}\mathbf{E}_P[\sup_{(\theta,\tau)\in B_{\Theta\times\mathcal{T}}(\delta)}|X_{ki}(\theta,\tau)|^{2((s-1)\vee 1)+\delta}]<\infty$.
(iv) For some $\delta>0$, there exists a functional $\Gamma_{k,P}(x)[\theta-\theta_0,\tau-\tau_0]$ of $(\theta-\theta_0,\tau-\tau_0)$, $(\theta,\tau)\in B_{\Theta\times\mathcal{T}}(\delta)$, such that

$$\begin{aligned}&|\mathbf{E}_P\left[h_x(X_{ki}(\theta,\tau))\right]-\mathbf{E}_P\left[h_x(X_{ki})\right]-\Gamma_{k,P}(x)[\theta-\theta_0,\tau-\tau_0]|\\ \le\ & C_1||\theta-\theta_0||^2+C_2||\tau-\tau_0||_\infty^2,\end{aligned}$$

with constants C_1 and C_2 that do not depend on P.

The bracketing entropy condition for $\mathcal{T}$ in (ii) is satisfied by many classes of functions. For example, when $\mathcal{T}$ is a Hölder class of smoothness α with the common domain of $\tau(\cdot)\in\mathcal{T}$ that is convex, and bounded in the $d_{\mathcal{T}}$-dim Euclidean space with $d_{\mathcal{T}}/\alpha\in(0,1]$ (e.g. Corollary 2.7.2 in van der Vaart and Wellner (1996)), the bracketing entropy condition holds. In this case, we can take $d=d_{\mathcal{T}}/\alpha$. The condition is also satisfied when $\mathcal{T}$ is contained in a class of uniformly bounded functions of bounded variation. Condition (iii) is a moment condition with local uniform boundedness. The moment condition is widely used in the literature of semiparametric inferences. In the example of single-index restrictions where $Y_k=\tau_0(Z_{1k}^\top\theta_0)+\varepsilon_k$, we can write $\varphi(w;\theta,\tau)=\tau_0(z_{1k}^\top\theta_0)-\tau(z_{1k}^\top\theta)+\varepsilon_k$. If τ is uniformly bounded in the neighborhood of τ_0 in $L_2(P)$, the moment condition is immediately satisfied when $\mathbf{E}[|\varepsilon_i|^{2((s-1)\vee 1)+\delta}] < \infty$. We may check this condition for other semiparametric specifications in a similar manner.

Condition (iv) requires the pathwise differentiability of functional $\int h_x(X_k(\theta,\tau))dP$ in $(\theta,\tau) \in B_{\Theta\times\mathcal{T}}(\delta)$. Suppose that $X_{ki}(\theta,\tau) = \varphi_k(W_i;\theta,\tau)$. When $s \geq 2$, $h_x(\varphi)$ is continuously differentiable in φ with the derivative bounded by $C|x-\varphi|^{s-2}q(x)$. Hence Condition (iv) follows if the moment condition in (iii) is satisfied, and

$$\begin{aligned} & \left|\varphi_k(w;\theta,\tau) - \varphi_k(w;\theta_0,\tau_0) - \Gamma^{\varphi}_{k,P}(w)[\theta-\theta_0,\tau-\tau_0]\right| \\ \leq \ & C_1||\theta-\theta_0||^2 + C_2||\tau-\tau_0||^2_\infty \end{aligned}$$

where $\Gamma^{\varphi}_{k,P}(w)[\theta-\theta_0,\tau-\tau_0]$ is a measurable function linear in $\theta-\theta_0$ and $\tau-\tau_0$. Indeed, in this case, we can take

$$\Gamma_{k,P}(x)[\theta-\theta_0,\tau-\tau_0] = \mathbf{E}\left[Dh_x(\varphi(W;\theta_0,\tau_0))\Gamma^{\varphi}_{k,P}(W)[\theta-\theta_0,\tau-\tau_0]\right],$$

where Dh_x is the first-order derivative of h_x. When $s=1$, the lower level conditions for (iv) can be obtained using the specification of $\varphi(w;\theta,\tau)$. For example, suppose $\varphi(W_k;\tau,\theta) = Y_k - \tau(Z_{1k}^\top\theta)$. Define $B(\tau,\delta) = \{\tau \in \mathcal{T} : ||\tau-\tau_0||_\infty \leq \delta\}$ and $B(\theta_0,\delta) = \{\theta\in\Theta : ||\theta-\theta_0|| \leq \delta\}$. Assume that each τ is twice continuously differentiable and

$$\sup_{\tau_1\in B(\tau,\delta)}\mathbf{E}[\sup_{\theta\in B(\theta_0,\delta)}|D^j\tau_1(Z_{1k}^\top\theta)|^2||Z_{1k}||^2] < \infty, j=0,1,2,$$

where $D^j\tau$ denotes the j–th order derivative of τ. Furthermore, the conditional density $f(\cdot|Z_{1k})$ of Y_k given Z_{1k} is assumed to exist and second order continuously differentiable with bounded derivatives. Then, Condition (iv) is satisfied. To see this, first observe that

$$\begin{aligned} & \mathbf{E}[\tau(Z_{1k}^\top\theta) - \tau_0(Z_{1k}^\top\theta_0)]^2 \\ \leq \ & 2\mathbf{E}\left[\sup_{\theta\in B(\theta_0,\delta)}|D\tau(Z_{1k}^\top\theta)|^2||Z_{1k}||^2\right]||\theta-\theta_0||^2 + 2\mathbf{E}\left[\tau(Z_{1k}^\top\theta_0) - \tau_0(Z_{1k}^\top\theta_0)\right]^2 \\ & +o(||\theta-\theta_0||^2 + ||\tau-\tau_0||^2_\infty) \\ = \ & O(||\theta-\theta_0||^2 + ||\tau-\tau_0||^2_\infty). \end{aligned}$$

Now, by applying the above bound,

$$\begin{aligned}P\{Y_k - \tau(Z_{1k}^\top\theta) &\leq x|Z_{1k}\} - P\{Y_k - \tau_0(Z_{1k}^\top\theta_0) \leq x|Z_{1k}\} \\ &= f(x+\tau_0(Z_{1k}^\top\theta_0)|Z_{1k})\left[\tau(Z_{1k}^\top\theta) - \tau_0(Z_{1k}^\top\theta) + \tau_0(Z_{1k}^\top\theta) - \tau_0(Z_{1k}^\top\theta_0)\right] \\ &\quad +O(||\theta-\theta_0||^2 + ||\tau-\tau_0||_\infty^2) \\ &= f(x+\tau_0(Z_{1k}^\top\theta_0)|Z_{1k})\left[\tau(Z_{1k}^\top\theta) - \tau_0(Z_{1k}^\top\theta) + D\tau_0(Z_{1k}^\top\theta_0)Z_{1k}^\top(\theta-\theta_0)\right] \\ &\quad +O(||\theta-\theta_0||^2 + ||\tau-\tau_0||_\infty^2).\end{aligned}$$

Hence, we can take

$$\begin{aligned}&\Gamma_{k,P}(x)[\theta-\theta_0,\tau-\tau_0] \\ &= \mathbf{E}\left[f(x+\tau_0(Z_{1k}^\top\theta_0)|Z_{1k})\left[\tau(Z_{1k}^\top\theta) - \tau_0(Z_{1k}^\top\theta) + D\tau_0(Z_{1k}^\top\theta_0)Z_{1k}^\top(\theta-\theta_0)\right]\right].\end{aligned}$$

The computation of the pathewise derivative can be performed similarly in many semiparametric models.

Assumption 2 : (i) For each $k = 1,\cdots,K$, the distribution of $X_{ki}(\theta_0,\tau_0)$ is absolutely continuous with a bounded density.
(ii) Condition (A) below holds when $s=1$, and Condition (B), when $s>1$.
(A) There exist $\delta, C > 0$ and a subvector W_1 of W such that (a) the conditional density of W given W_1 is bounded uniformly over $(\theta,\tau) \in B_{\Theta\times\mathcal{T}}(\delta)$ and over $P \in \mathcal{P}$, (b) for each (θ,τ) and $(\theta',\tau') \in B_{\Theta\times\mathcal{T}}(\delta)$, $\varphi_k(W;\theta,\tau) - \varphi_k(W;\theta',\tau')$ is measurable with respect to the σ-field of W_1, and (c) for each $(\theta_1,\tau_1) \in B_{\Theta\times\mathcal{T}}(\delta)$ and for each $\varepsilon > 0$,

$$\sup_{P\in\mathcal{P}}\sup_{w_1}\mathbf{E}_P\left[\sup_{(\theta_2,\tau_2)\in B_{\Theta\times\mathcal{T}}(\varepsilon)}|\varphi_k(W;\theta_1,\tau_1) - \varphi_k(W;\theta_2,\tau_2)|^2\,|W_1 = w_1\right] \leq C\varepsilon^{2s_2} \tag{6}$$

for some $s_2 \in (d\lambda/2, 1]$ with $\lambda = 1\{s=1\} + 2\times\{s>1\}$ and d in Assumption 1(ii), where the supremum over w_1 runs in the support of W_1.
(B) There exist $\delta, C > 0$ such that Condition (c) above is satisfied with the conditional expectation replaced by the unconditional one.

Assumption 2(ii) contains two different conditions that are suited to each case of $s=1$ or $s>1$. This different treatment is due to the nature of the function $h_x(\varphi) = (x-\varphi)^{s-1}1\{\varphi \leq x\}q(x)/(s-1)!$ that is discontinuous in φ when $s=1$ and continuous in φ when $s>1$. Condition (A) can be viewed as a generalization of the set up of LMW. Condition (A)(a) is analogous to Assumption 1(iii) of LMW. Condition (A)(b) is satisfied by many semiparametric models. For example, in the case of a partially parametric specification: $X_k(\theta_0,\tau_0) = Y_k - Z_{1k}^\top\theta_0 - \tau_0(Z_{2k})$, we take $W = (Y, Z_1, Z_2)$ and $W_1 = (Z_1, Z_2)$. In the case of single index restrictions: $X_k(\theta_0,\tau_0) = Y_k - \tau_0(Z_k^\top\theta_0)$, $W = (Y, Z)$ and $W_1 = Z$. The condition in (6) requires the function $\varphi(W;\theta,\tau)$ to be (conditionally) locally uniformly $L_2(P)$-continuous in $(\theta,\tau) \in B_{\Theta\times\mathcal{T}}(\delta)$ uniformly over $P \in \mathcal{P}$. When $\varphi_k(W;\theta_1,\tau_1)$ is smooth in (θ_1,τ_1), the condition holds with $s_2 = 1$. However, when $\varphi_k(W;\theta,\tau)$ is discontinuous in (θ,τ), the condition may hold with s_2 smaller than 1. Sufficient conditions and discussions can be found in Chen, Linton, and van Keilegom (2003). We can weaken this condition to the unconditional version when we consider only the case $s > 1$.

We now turn to the test statistics. Let $\bar{F}_{kN}(x,\theta,\tau) = \frac{1}{N}\sum_{i=1}^N 1\{X_{ki}(\theta,\tau) \le x\}$ and

$$\bar{D}_{kl}^{(s)}(x,\theta,\tau) = \bar{D}_k^{(s)}(x,\theta,\tau) - \bar{D}_l^{(s)}(x,\theta,\tau) \text{ for } s \ge 1,$$

where $\bar{D}_k^{(1)}(x,\theta,\tau) = \bar{F}_{kN}(x,\theta,\tau)$ and $\bar{D}_k^{(s)}(x,\theta,\tau)$ is defined through the following recursive relation:

$$\bar{D}_{kl}^{(s)}(x,\theta,\tau) = \int_{-\infty}^{x} \bar{D}_{kl}^{(s-1)}(t,\theta,\tau)dt \text{ for } s \ge 2. \tag{7}$$

The numerical integration in (7) can be cumbersome in practice. Integrating by parts, we have an alternative form

$$\bar{D}_k^{(s)}(x,\theta,\tau) = \frac{1}{N(s-1)!}\sum_{i=1}^N (x - X_{ki}(\theta,\tau))^{s-1} 1\{X_{ki}(\theta,\tau) \le x\}.$$

Since $\bar{D}_k^{(s)}$ and $D_k^{(s)}$ are obtained by applying a linear operator to $\bar{F}_{kN}$ and F_k, the estimated function $\bar{D}_k^{(s)}$ is an unbiased estimator for $D_k^{(s)}$. From now on, we suppress the suprescripts (s) from the notations so that we write $\bar{D}_k$ and D_k for $\bar{D}_k^{(s)}$ and $D_k^{(s)}$.

The test statistics we consider are based on the weighted empirical analogues of c_s, namely,

$$T_N = \min_{k\neq l}\int_{\mathcal{X}} \max\left\{q(x)\sqrt{N}\bar{D}_{kl}(x,\hat{\theta},\hat{\tau}),0\right\}^2 w(x)dx.$$

Regarding the estimators $\hat{\theta}$ and $\hat{\tau}$ and the weight function q, we assume the following.

Assumption 3 : (i) $||\hat{\theta} - \theta_0|| + ||\hat{\tau} - \tau_0||_\infty = O_P(\delta_N)$ for some $\delta_N \to 0$, uniformly in $P \in \mathcal{P}$ and $\sup_{P\in\mathcal{P}} P\{\hat{\tau} \in \mathcal{T}\} \to 1$ as $N \to \infty$.
(ii) For each $\varepsilon > 0$,

$$\sup_{P\in\mathcal{P}} P\left\{\sup_{x\in\mathcal{X}}\left|\sqrt{N}\Gamma_{k,P}(x)[\hat{\theta}-\theta_0,\hat{\tau}-\tau_0] - \frac{1}{\sqrt{N}}\sum_{i=1}^N \psi_{x,k}(W_i;\theta_0,\tau_0)\right| > \delta_N M\right\} \to 0, \tag{8}$$

where $\psi_{x,k}(\cdot)$ satisfies that there exist $\eta, \delta, C > 0$ and $s_1 \in (d/2, 1]$ with d in Assumption 1(ii) such that for all $x \in \mathcal{X}$, $\mathbf{E}_P\left[\psi_{x,k}(W_i; \theta_0, \tau_0)\right] = 0$,

$$\sup_{P \in \mathcal{P}} \left\{\mathbf{E}\left[\sup_{x \in \mathcal{X}} |\psi_{x,k}(W; \theta, \tau)|^{2+\eta}\right]\right\}^{1/(2+\eta)} < \infty,$$

and for each $\varepsilon > 0$ and $x \in \mathcal{X}$,

$$\sup_{P \in \mathcal{P}} \mathbf{E}_P \left[\sup_{x_1 \in \mathcal{X}: d_k(x, x_1) \leq \varepsilon} |\psi_{x,k}(W_i; \theta_0, \tau_0) - \psi_{x_1,k}(W_i; \theta_0, \tau_0)|^2\right] \leq C\varepsilon^{2s_1}, \tag{9}$$

where $d_k(x, x_1) = |q(x)|D_k(x)| - q(x_1)|D_k(x_1)||$.
(iii) (a) $\sup_{x \in \mathcal{X}} \left(1 + |x|^{(s-1) \vee (1+\delta)}\right) q(x) < \infty$, for some $\delta > 0$ and for q, nonnegative, first order continuously differentiable function on $\mathcal{X}$ with a bounded derivative and the support of $q(x)$ contains $\mathcal{X}$.
(b) For any sequence $\varepsilon_N \to \infty$, $\int_{[B_{kl}]^{\varepsilon_N} \setminus B_{kl}} w(x)dx = O(\varepsilon_N)$ for all k, l, where $[B_{kl}]^{\varepsilon_N} = \{x \in \mathcal{X} : q(x)|D_{kl}(x)| \leq \varepsilon_N\}$.
(iv) For each $k \neq l$, there exist $\delta > 0$ and constants $C > 0$ and $p_{kl} \geq 1$ such that for each $x, x' \in \mathcal{X}$,

$$d_{kl}(x, x') \geq C\{\mathbf{E}[\sup_{(\theta,\tau) \in B_{\Theta \times \mathcal{T}}(\delta)} |V_{x,kl}(w; \theta, \tau) - V_{x',kl}(w; \theta, \tau)|^2] \wedge \delta\}^{p_{kl}/2},$$

where $d_{kl}(x, x') = |q(x)\,|D_{kl}(x)| - q(x')\,|D_{kl}(x')|\,|$, $V_{x,kl}(w; \theta, \tau) = (h^{\Delta}_{x,kl} + \psi^{\Delta}_{x,kl})(w; \theta, \tau)$,

$$\begin{aligned} h^{\Delta}_{x,kl}(w; \theta, \tau) &= h_x(\varphi_k(w; \theta, \tau)) - h_x(\varphi_l(w; \theta, \tau)), \text{ and} \\ \psi^{\Delta}_{x,kl}(w; \theta, \tau) &= \psi_{x,k}(w; \theta, \tau) - \psi_{x,l}(w; \theta, \tau). \end{aligned} \tag{10}$$

When $\hat{\theta}$ is a solution from the M-estimation problem, its rate of convergence can be obtained by following the procedure of Theorem 3.2.5 of van der Vaart and Wellner (1996). The uniformity in P in this case can be ensured by combining the uniform (in P) oscillation behavior of the population objective function and the associated empirical processes. For instance, one may employ a uniform oscillation exponential bound for empirical processes established by Giné (1997), Theorem 6.1. The rate of convergence for the nonparametric component $\hat{\tau}$ can also be established by controlling the bias and variance part uniformly in P. In order to control the variance part, one may employ the framework of Giné and Zinn (1991) and establish the central limit theorem that is uniform in P. The sufficient conditions for the last condition in (i) can be checked, for example, from the results of Andrews (1994).

The condition in (8) indicates that the functional $\Gamma_{k,P}$ at the estimators has an asymptotic linear representation. This condition can be established using the standard method of expanding the functional in terms of the estimators, $\hat{\theta}$ and $\hat{\tau}$, and using the asymptotic linear representation of these estimators. The asymptotic linear representation for these estimators is available in many semiparametric models. Since our procedure does not make use of its particular characteristic beyond the condition in (8), we keep this condition at a high level for the sake of brevity.

Condition (iii)(a) is fulfilled by an appropriate choice of the weight function. The use of the weight function is convenient as it enables us to allow $\mathcal{X}$ to be unbounded. Condition (iii)(a) is stronger than that of Horváth, Kokoszka, and Zikitis (2006) who under a set-up simpler than this paper, imposed that $\sup_{x\in\mathcal{X}}\left(1+(\max(x,0)^{(s-2)\vee 1}\right)q(x)<\infty$. Note that the condition in (iii)(a) implies that when $\mathcal{X}$ is a bounded set, we may simply take $q(x)=1$. When $s=1$ so that our focus is on the first order stochastic dominance relation, we may transform the variable $X_{ki}(\theta,\tau)$ into one that has a bounded support by taking a smooth strictly monotone transform. After this transformation, we can simply take $q(x)=1$. Condition (iii)(b) is a regularity condition for the weighting function and the contact set. When the Lebesgue measure of the set $[B_{kl}]^{\varepsilon_N}\backslash B_{kl}$ is $O(\varepsilon_N)$ and $w(x)$ is bounded, the condition obviously follows. This condition has an important implication when we consider a sequence of probabilities such that $D_{kl}(x)$ converges to zero over some points $x\in\mathcal{X}$, for some $k\neq l$. The condition requires that this convergence should be "regular" in the sense that the contact set B_{kl} behaves continuously as we move along such a sequence of probabilities. Condition (iv) is similar to Condition C.2 in Chernozhukov, Hong, and Tamer (2007) and is used to control the convergence rate of the estimated set with respect to an appropriate norm. This condition is needed only for the convergence rate of the rejection probability. The condition is stronger than is necessary for

The first result below is concerned with the convergence in distribution of T_N under the null hypothesis uniformly in a subset of $\mathcal{P}$. Let $\nu_{kl}(\cdot)$ be a mean zero Gaussian process on $\mathcal{X}$ with a covariance kernel given by

$$C_{kl}(x_1,x_2)=Cov_P(V_{x_1,kl}(W_i;\theta_0,\tau_0),V_{x_2,kl}(W_i;\theta_0,\tau_0)), \tag{11}$$

where Cov_P denotes the covariance under P and $V_{x,kl}(w;\theta_0,\tau_0)$ is defined in Assumption 3(iv). The asymptotic critical values are based on this Gaussian process. By convention, we assume that the supremum over an empty set is equal to $-\infty$. Let $\mathcal{P}_0$ be the set of probabilities under the null hypothesis.

Theorem 1 : *Suppose that* Assumptions 1-3 *hold. Then under the null hypothesis, as*

$N \to \infty$,

$$T_N \underset{D}{\to} \min_{k \neq l} \int_{B_{kl}} \max\{\nu_{kl}(x), 0\}^2 w(x)dx, \text{ uniformly in } P \in \mathcal{P}_0.$$

The asymptotic result in Theorem 1 is uniform over the probabilities under the null hypothesis. For fixed probabilities, the limiting behavior of the test statistic is discontinuous. Let $\mathcal{P}_{00}$ be the set of probabilities in $\mathcal{P}_0$ such that for each $k \neq l$, $\int_{B_{kl}} w(x)dx > 0$. The set $\mathcal{P}_{00}$ constitutes the "boundary points" of the null hypothesis. On the other hand, the "interior points", $\mathcal{P}_0 \backslash \mathcal{P}_{00}$, are probabilities such that for some $k \neq l$, $\int_{B_{kl}} w(x)dx = 0$. Under a fixed probability $P \in \mathcal{P}_{00}$, the test statistic has a nondegenerate limiting distribution. But under a fixed probability $P \in \mathcal{P}_0 \backslash \mathcal{P}_{00}$, the limiting distribution is degenerate at zero. This phenomenon of discontinuity arises often in moment inequality models. (e.g. Moon and Schorfheide (2006), Chernozhukov, Hong, and Tamer (2007), and Andrews and Guggenberger (2006)). We introduce the following definition of a test having an asymptotically exact size.

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

$$\text{limsup}_{N \to \infty} \text{sup}_{P \in \mathcal{P}_0} \mathbf{E}_P \varphi_\alpha \leq \alpha.$$

$$\text{limsup}_{N \to \infty} \text{sup}_{P \in \mathcal{P}_0'} \mathbf{E}_P \varphi_\alpha = \alpha. \tag{12}$$

(ii) When a test φ_α satisfies (12), we say that the test is *asymptotically similar on* $\mathcal{P}_0'$.

The Gaussian process ν_{kl} in Theorem 1 depends on the unknown elements of the null hypothesis, precluding the use of first order asymptotic critical values in practice. Barrett and Donald (2003) suggested a bootstrap procedure and LMW, a subsampling approach. Both studies have not paid attention to the issue of uniformity in the convergence of tests. In order to deal with uniformity, we separate $\mathcal{P}_0$ into the boundary points and the interior points and deal with asymptotic theory of the test statistic separately. In fact, the proof of Theorem 1 establishes the weak convergence of $q(\cdot)\sqrt{N}\{\bar{D}_{kl}(\cdot, \hat{\theta}, \hat{\tau}) - D_{kl}(\cdot, \theta_0, \tau_0)\}$ to $\nu_{kl}(\cdot)$ uniformly over $P \in \mathcal{P}$. This result is obtained by showing that the class of functions indexing the empirical process is a uniform Donsker class (Giné and Zinn (1991)). The remaining step is to deal with the limiting behavior of the test statistic uniformly over the interior points.

One might consider a typical bootstrap procedure for this situation. The difficulty for a bootstrap test of stochastic dominance lies mainly in the fact that it is hard to impose the null hypothesis upon the test. There have been approaches that consider only *least favorable* subset of the models of the null hypothesis as in the following.

$$F_1(x) = \cdots = F_K(x) \text{ for all } x \in \mathcal{X}. \tag{13}$$

This leads to the problem of asymptotic nonsimilarity in the sense that when the true data generating process lies away from the least favorable subset of the models of the null hypothesis, and yet still lie at the boundary points, the bootstrap sizes become misleading even in large samples. LMW calls this phenomenon *asymptotic nonsimilarity on the boundary.* Bootstrap procedures that employ the usual recentering implicitly impose restrictions that do not hold outside the least favorable set and hence asymptotically biased against such probabilities outside the set.

To illustrate heuristically why a test that uses critical values from the least favorable case of a composite null hypothesis can be asymptotically biased, let us consider a simple example in the finite dimensional case. (For a related result in the context of comparing predictabilities of forecasting models, see Hansen (2005)). Suppose that the observations $\{X_i = (X_{1i}, X_{2i}) : i = 1, \ldots, N\}$ are mutually independently and identically distributed with unknown mean $\mu = (\mu_1, \mu_2)$ and known variance $\Sigma = diag(1,1)$. Let the hypotheses of interest be given by:

$$H_0 : \mu_1 \leq 0 \text{ and } \mu_2 \leq 0 \ \text{ vs. } H_1 : \mu_1 > 0 \text{ or } \mu_2 > 0. \tag{14}$$

The "boundary" of the null hypothesis is given by $\mathcal{B}_{BD} = \{(\mu_1, \mu_2) : \mu_1 \leq 0 \text{ and } \mu_2 \leq 0\} \cap \{(\mu_1, \mu_2) : \mu_1 = 0 \text{ or } \mu_2 = 0\}$, while the "least favorable case (LFC)" is given by $\mathcal{B}_{LF} = \{(\mu_1, \mu_2) : \mu_1 = 0 \text{ and } \mu_2 = 0\} \subset \mathcal{B}_{BD}$. To test (14), one may consider the following the t-statistic:

$$T_N = \max\{N^{1/2}\overline{X}_1, N^{1/2}\overline{X}_2\},$$

where $\overline{X}_k = \sum_{i=1}^N X_{ki}/N$. Then, the asymptotic null distribution of T_N is non-degenerate provided the true μ lies on the boundary $\mathcal{B}_{BD}$, but the distribution depends on the location of μ. That is, we have

$$T_N \xrightarrow{d} \begin{cases} \max\{Z_1, Z_2\} & \text{if} \quad \mu = (0,0) \\ Z_1 & \text{if} \quad \mu_1 = 0, \mu_2 < 0 \\ Z_2 & \text{if} \quad \mu_1 < 0, \mu_2 = 0 \end{cases},$$

where Z_1 and Z_2 are mutually independent standard normal random variables. On the other hand, T_N diverges to $-\infty$ in the "interior" of the null hypothesis. Suppose that z_α^* satisfies $P(\max\{Z_1, Z_2\} > z_\alpha^*) = \alpha$ for $\alpha \in (0,1)$. Then, the test based on the least favorable case is asymptotically non-similar on the boundary because, for example, $\lim_{N\to\infty} P(T_N > z_\alpha^* | \mu = (0,0)) \neq \lim_{N\to\infty} P(T_N > z_\alpha^* | \mu = (0,-1))$. Now consider the following sequence of local alternatives: for $\delta > 0$,

$$H_N : \mu_1 = \frac{\delta}{\sqrt{N}} \text{ and } \mu_2 < 0.$$

Then, under H_N, it is easy to see that $T_N \xrightarrow{d} N(\delta, 1)$. However, the test based on the LFC critical value may be biased against this local alternatives, because $\lim_{N\to\infty} P(T_N > z_\alpha^*) = P(N(\delta,1) > z_\alpha^*) < \alpha$ for some values of δ. To see this, in Figure 1, we draw a c.d.f.'s of $\max\{Z_1, Z_2\}$ and $N(\delta, 1)$ for $\delta = 0.0$, 0.2, and 1.5. Clearly, the distribution of $\max\{Z_1, Z_2\}$ first-order stochastic dominates that of $N(0.2, 1)$, i.e., T_N is asymptotically biased against H_N for $\delta = 0.2$.

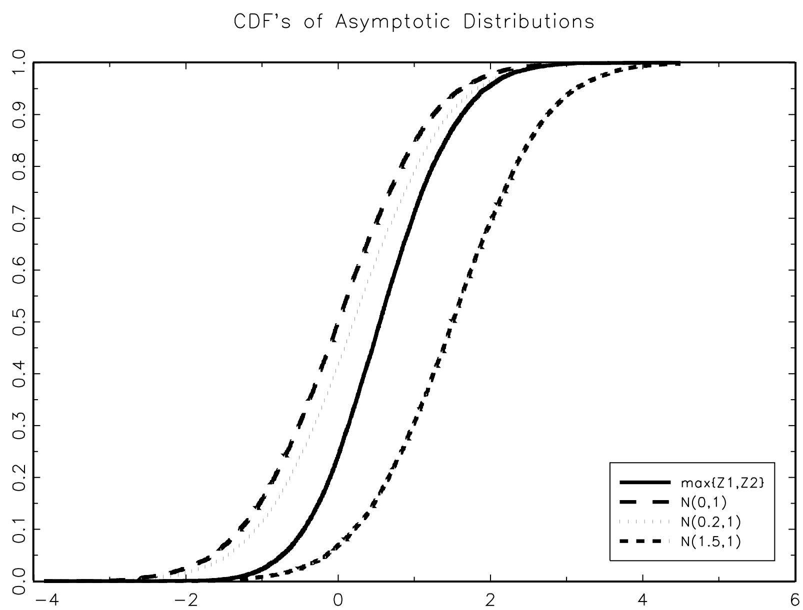

Figure 1. Shows that T_N is asymptocally biased against H_N for $\delta = 0.2$.

3 Bootstrap Procedure

In this section we discuss our method for obtaining asymptotically valid critical values. We propose a bootstrap procedure as follows. We draw $\{W_{i,b}^*\}_{i=1}^N$, $b = 1, \cdots, B$, with replacement from the empirical distribution of $\{W_i\}_{i=1}^N$. Then, we construct estimators $\hat{\theta}_b^*$ and $\hat{\tau}_b^*$, for each $b = 1, \ldots, B$, using the bootstrap sample $\{W_{i,b}^*\}_{i=1}^N$. Given the bootstrap estimators

$\hat{\theta}_b^*$ and $\hat{\tau}_b^*$, we define $\tilde{X}^*_{ki,b} = \varphi_k(W^*_{i,b}; \hat{\theta}^*, \hat{\tau}^*)$. The bootstrap estimators $\hat{\theta}_b^*$ and $\hat{\tau}_b^*$ should be constructed so that their bootstrap distribution mimicks the distribution of $\hat{\theta}$ and $\hat{\tau}$. For example, when θ_0 and τ_0 are identified through a moment condition:

$$\mathbf{E}\left[Z_i \varphi_k(W_i; \theta_0, \tau_0)\right] = 0, \;\; k = 1, 2, \cdots, K,$$

for some random vector Z_i, one needs to consider the *recentered* moment conditions for the bootstrap estimators. (See Hall and Horowitz (1996)). We do not detail the method of constructing $\hat{\theta}_b^*$ and $\hat{\tau}_b^*$ as it depends on the way the parameters θ_0 and τ_0 are identified. All we require for these estimators is subsumed in Assumption 4 below.

Now, we introduce the following bootstrap empirical process

$$\bar{D}^*_{kl,b}(x) = \frac{1}{N}\sum_{i=1}^{N}\left\{h_x(\tilde{X}^*_{ki,b}) - h_x(\tilde{X}^*_{li,b}) - \frac{1}{N}\sum_{i=1}^{N}\{h_x(\hat{X}_{ki}) - h_x(\hat{X}_{li})\}\right\}, \; b = 1, 2, \ldots, B,$$

where $\hat{X}_{ki} = X_{ki}(\hat{\theta}, \hat{\tau})$. The quantity $\bar{D}^*_{kl,b}(x)$ denotes the bootstrap counterpart of $\bar{D}_{kl}(x)$. Given a sequence $c_N \to 0$ and $c_N\sqrt{N} \to \infty$, we construct an estimated contact set:

$$\hat{B}_{kl} = \left\{x \in \mathcal{X} : q(x)|\bar{D}_{kl}(x)| < c_N\right\}. \tag{15}$$

As for the weight function $q(x)$, we may consider the following type of function. For $z_1 < z_2$ and for constants $a, \delta > 0$, we set

$$q(x) = \begin{cases} 1 \text{ if } x \in [z_1, z_2] \\ a/(a + |x - z_2|^{(s-1)\vee(1+\delta)}) \text{ if } x > z_2 \\ a/(a + |x - z_1|^{(s-1)\vee(1+\delta)}) \text{ if } x < z_1. \end{cases}$$

This is a modification of the weighting function considered by Horváth, Kokoszka, and Zikitis (2006).

Let us introduce a bootstrap test statistic. The examination of Theorem 1 reveals that the limiting distribution of the test statistic becomes degenerate when for some $k \neq l$, $\int_{B_{kl}} w(x)dx$ converges to zero too fast. Since the estimated set $\hat{B}_{kl}$ has a sampling variation, there is a limitation in bootstrap's ability to mimick this sequence of probabilities leading to degeneracy. Hence instead of considering all the probabilities in the boundary points, we consider a sequence of expanding subsets of boundary points that expand toward the entire boundary set as $N \to \infty$. More specifically, let $z_N \to 0$ such that $c_N/z_N \to 0$. Then we define $\mathcal{P}_{0,N} = \{P \in \mathcal{P}_0 : \int_{B_{kl}} w(x)dx > z_N \text{ for some } k \neq l\}$. Instead of using the demarcation of $\mathcal{P}_0$ into $\mathcal{P}_{00}$ and $\mathcal{P}_0 \backslash \mathcal{P}_{00}$, we use that into $\mathcal{P}_{0,N}$ and $\mathcal{P}_0 \backslash \mathcal{P}_{0,N}$. The sample analogue of this dichotomy is $\int_{\hat{B}_{kl}} w(x)dx > z_N$ for all $k \neq l$ and $\int_{\hat{B}_{kl}} w(x)dx \leq z_N$ for some $k \neq l$. Using this scheme, we propose the following bootstrap test statistic:

$$T^*_{N,b} = \begin{cases} \min_{k \neq l} \int_{\hat{B}_{kl}} \max\left\{q(x)\sqrt{N}\bar{D}^*_{kl,b}(x), 0\right\}^2 w(x)dx, & \text{if } \int_{\hat{B}_{kl}} w(x)dx > z_N \text{ for all } k \neq l \\ \\ \pi_N \quad \text{if } \int_{\hat{B}_{kl}} w(x)dx \leq z_N \text{ for some } k \neq l, \end{cases} \tag{16}$$

where π_N is a positive sequence such that $\pi_N \to \infty$. The bootstrap critical values are obtained by $c^*_{\alpha,N,B} = \inf\{t : B^{-1}\Sigma^B_{b=1} 1\{T^*_{N,b} \leq t\} \geq 1 - \alpha\}$, yielding an α-level bootstrap test: $\varphi_\alpha \equiv 1\{T_N > c^*_{\alpha,N,B}\}$.

The main step we need to justify the bootstrap procedure uniformly in $P \in \mathcal{P}$ under this general semiparametric environment is to establish the bootstrap uniform central limit theorem for the empirical processes in the test statistic, where uniformity holds over probability measures in $\mathcal{P}$. The uniform central limit theorem for empirical processes has been developed by Giné and Zinn (1991) and Sheehy and Wellner (1992) through the characterization of uniform Donsker classes. The bootstrap central limit theorem for empirical processes was established by Giné and Zinn (1990) who considered a nonparametric bootstrap procedure. See also a nice lecture note on the bootstrap by Giné (1997). We use these results for the proof of Theorem 2 below. Let $\mathcal{G}_N$ be the σ-field generated by $\{W_i\}^N_{i=1}$.

Assumption 4: (i) For $\psi_{x,k}(\cdot)$ in Assumption 3(ii), there exists $r_N \to 0$ such that $r_N/\delta_N = O(1)$ and for any $\varepsilon > 0$,

$$P\left\{\sup_{x\in\mathcal{X}} \left|\sqrt{N}\hat{\Gamma}_{k,P}(x) - \frac{1}{\sqrt{N}}\sum_{i=1}^N \left\{\psi_{x,k}(W^*_{i,b}; \hat{\theta}, \hat{\tau}) - \frac{1}{N}\sum_{i=1}^N \psi_{x,k}(W_i; \hat{\theta}, \hat{\tau})\right\}\right| > r_N \varepsilon | \mathcal{G}_N\right\} \to_P 0,$$

uniformly in $P \in \mathcal{P}$ for any $\varepsilon > 0$, where $\hat{\Gamma}_{k,P}(x) = \frac{1}{N}\sum_{i=1}^N\{h_x(\tilde{X}^*_{ki,b}) - h_x(\hat{X}^*_{ki,b})\}$ and $\hat{X}^*_{ki,b} = \varphi_k(W^*_{i,b}; \hat{\theta}, \hat{\tau})$.

(ii) There exist $\delta > 0$ and $s_3 \in (\frac{d}{2} \vee s_2, 1]$ such that

$$\sup_{P\in\mathcal{P}}\left\{\mathbf{E}\left[\sup_{(x,\theta,\tau)\in\mathcal{X}\times B_{\Theta\times\mathcal{T}}(\delta)}|\psi_{x,k}(W;\theta,\tau)|^{2+\eta}\right]\right\}^{1/(2+\eta)} < \infty,$$

and for each $x \in \mathcal{X}$ and $(\theta_1, \tau_1) \in B_{\Theta\times\mathcal{T}}(\delta)$, for all $\varepsilon \in (0, \delta]$,

$$\sup_{P\in\mathcal{P}} \mathbf{E}_P\left[\sup_{(x,\theta,\tau)\in\mathcal{X}\times B_{\Theta\times\mathcal{T}}(\varepsilon):d_k(x,x_1)\leq\varepsilon} |\psi_{x,k}(W_i;\theta,\tau) - \psi_{x_1,k}(W_i;\theta_1,\tau_1)|^2\right] \leq C\varepsilon^{2s_3}. \tag{17}$$

Condition (i) assumes the bootstrap analogue of the asymptotic linearity of $\sqrt{N}\Gamma_{k,P}(x)[\hat{\theta}-\theta_0, \hat{\tau}-\tau_0]$ with the same influence function $\psi_{x,k}$. (See e.g. Koul and Lahiri (1994)). Condition (ii) is typically proved when one establishes the validity of bootstrap confidence sets for $\hat{\theta}^*$ and $\hat{\tau}^*$. In this regard, the bootstrap validity of M-estimators is established by Arcones and Giné (1992). Infinite dimensional Z-estimators are dealt with by Wellner and Zhan (1996). See also Abrevaya and Huang (2005) for a bootstrap inconsistency result for a case where the estimators are not asymptotically linear. The condition for $\psi_{x,k}$ is stronger than (9) as it requires uniformity over $(\theta, \tau) \in B_{\Theta\times\mathcal{T}}(\delta)$. In the following, we establish the asymptotic validity for the proposed bootstrap procedure.

Theorem 2 : *Suppose that the conditions of* Theorem 1 *and* Assumption 4 *hold and that* $c_N = o(1)$, $c_N\sqrt{N} \to \infty$, *and* $c_N\sqrt{N}(-\log c_N)^{-1/4} = O(1)$ as $N \to \infty$. *Suppose further that for all* $P \in \mathcal{P}_{0,N}$, ν_{kl} *has a positive definite covariance function. Then*

$$\begin{aligned} P\left\{T_N > c^*_{\alpha,N,\infty}\right\} &\leq \alpha + O_P(\delta_N^{\kappa}) + O_P(c_N\sqrt{-\log c_N}) \text{ uniformly over } P \in \mathcal{P}_0 \text{ and} \\ P\left\{T_N > c^*_{\alpha,N,\infty}\right\} &= \alpha + O_P(\delta_N^{\kappa}) + O_P(c_N\sqrt{-\log c_N}) \text{ uniformly over } P \in \mathcal{P}_{0,N}, \end{aligned}$$

where $\kappa = \frac{\lambda}{2}(s_2 - \frac{d\lambda}{2})$ *and* s_2 *and* λ *are constants that appear in* Assumption 2(ii)(A).

The conditions on the tuning parameter c_N are satistfied if, e.g., $c_N \propto (\log N)^{1/4}/N^{1/2}$ or $(\log\log N)^{1/4}/N^{1/2}$. The first result of Theorem 2 says that the bootstrap tests have asymptotically correct sizes. The second result tells us that the bootstrap tests are asymptotically similar on the expanding subsets of the boundary. The second result combined with the first result establishes that the bootstrap tests have exact asymptotic size equal to α. The result is uniform over $P \in \mathcal{P}_0$. The limit behavior of the bootstrap test statistic mimics the discontinuity in Theorem 1.

The approximation error rate in Theorem 2 contains two components. The first component involves the rate of convergence of the estimators $\hat{\theta}$ and $\hat{\tau}$. The second component stems from the estimation error of the estimated contact set. For example, suppose that there is no nonparametric component τ and $\hat{\theta} = \theta_0 + O(N^{-1/2})$, so that $\delta_N = N^{-1/2}$. If $s_2 = 1$ and $s > 1$ (and hence $\lambda = 2$), we may have $\kappa = 1$. Then, the term $O(\delta_N^{\kappa})$ is dominated by $O_P(c_N\sqrt{-\log c_N})$, achieving the approximation error rate close to $N^{-1/2}$. When there is nonparametric component τ, this error rate would usually depend on the rate of convergence of nonparametric estimators $\hat{\tau}$, which in turn depends on the smoothness of underlying functions and dimension of regressors. Hence even in this complex situation, we expect that the proposed bootstrap test improves on the subsampling method whose rate is known to be $N^{-1/3}$ at best for tests of finite dimensional models. (See. e.g. Horowitz (2001) and Politis and Romano (1999)). However, the error rate in Theorem 2 is not an exact rate as it relies on the Berry-Esseen bounds of normalized sums of random elements. For a more sophisticated result, one may need a formal Edgeworth expansion of the test statistic's distribution.

Our requirement for the sequence π_N to increase to infinity is used to obtain results that are uniform in P. To see this, consider the following:

$$\sqrt{N}q(x)\bar{D}_{kl}(x,\hat{\theta},\hat{\tau}) = \sqrt{N}q(x)\left\{\bar{D}_{kl}(x,\hat{\theta},\hat{\tau}) - D_{kl}(x)\right\} + \sqrt{N}q(x)D_{kl}(x). \tag{18}$$

The first component on the right-hand side is uniformly asymptotically tight and its tail probability goes to zero only when we increase the cut-off value to infinity. The behavior of the last term depends on the sequence of probabilities $P_N \in \mathcal{P}_0 \backslash \mathcal{P}_{00}$. The worst case arises when P_N converges to a boundary point so fast that for some $k \neq l$, the second term in (18) also converges to zero fast. Considering uniformity in P, we should take this case into account. Therefore, π_N should increase to infinity in order to control the tail probability of the combined quantities in (18) appropriately. On the other hand, for only pointwise results for each $P \in \mathcal{P}_0 \backslash \mathcal{P}_{00}$, it suffices to set π_N to be any nonnegative number.

The result of Theorem 2 allows for a wide range of choice for the sequence π_N in (16) in the bootstrap test statistic. In practice, we may consider two extreme cases $\pi_N = \log(N)$ and $\pi_N = \infty$, and see the robustness of the results. The latter case with $\pi_N = \infty$ is tantamount to the rule of setting the test $\varphi = 0$ (i.e. do not reject the null) when the contact set $\int_{\hat{B}_{kl}} w(x)dx = 0$.

4 Asymptotic Power Properties

In this section, we investigate asymptotic power properties of the bootstrap test. First, we consider consistency of the test.

Theorem 3 : *Suppose that the conditions of* Theorem 2 *hold and that we are under a fixed alternative* $P \in \mathcal{P} \backslash \mathcal{P}_0$ *such that* $\min_{k \neq l} \int_{B_{kl}} \max\{q(x)D_{kl}(x), 0\}^2 w(x)dx > 0$. *Then,*

$$\lim_{N\to\infty} P\left\{T_N > c^*_{\alpha,N,\infty}\right\} \to 1.$$

Therefore, the bootstrap test is consistent against all types of alternatives. This property is shared by other tests of LMW and Barrett and Donald (2003) for example.

Let us turn to asymptotic local power properties. We consider a sequence of probabilities $P_N \in \mathcal{P} \backslash \mathcal{P}_0$ and denote $D_{k,N}(x)$ to be $D_k(x)$ under P_N. That is, $D_{k,N}(x) = \mathbf{E}_{P_N} h_x(\varphi_k(W;\theta_0,\tau_0))$ using the notation of h_x and the specification of X_k in a previous section, where $\mathbf{E}_{P_N}$ denotes the expectation under P_N. We confine our attention to $\{P_N\}$ such that for each $k \in \{1, \ldots, K\}$, there exists functions $H_k(\cdot)$ and $\delta_k(\cdot)$ such that

$$D_{k,N}(x) = H_k(x) + \delta_k(x)/\sqrt{N}. \tag{19}$$

We assume the following for the functions $H_k(\cdot)$ and $\delta_k(\cdot)$.

Assumption 5: (i) $C_{kl} = \{x \in \mathcal{X} : H_k(x) - H_l(x) = 0\}$ is nonempty for all $k \neq l$.
(ii) $\min_{k \neq l} \sup_{x \in \mathcal{X}} (H_k(x) - H_l(x)) \leq 0$.
(iii) $\int_{C_{kl}} \max\{\delta_k(x) - \delta_l(x), 0\}^2 w(x) dx > 0$, for all $k \neq l$.

Assumption 5 enables the local sequences in (19) to constitute non-void local alternatives. Note that when $H_k(x) = \mathbf{E}_P h_x(\varphi_k(W; \theta_0, \tau_0))$ for some $P \in \mathcal{P}$, these conditions for H_k imply that $P \in \mathcal{P}_0$, that is, the probability associated with H_k belongs to the null hypothesis, in particular, the boundary points. When the distributions are continuous, the contact sets are nonempty under the alternatives. If the contact sets satisfy the condition $\int_{B_{kl}} w(x) dx > 0$, any local alternatives that converge to the null hypothesis will have a limit in the set of boundary points. The conditions for $\delta_k(x)$ indicate that for each N, the probability P_N belongs to the alternative hypothesis $\mathcal{P} \backslash \mathcal{P}_0$. Therefore, the sequence P_N represents local alternatives that converge to the null hypothesis (in particular, to the boundary points $\mathcal{P}_{00}$) maintaining the convergence of $D_{k,N}(x)$ to $H_k(x)$ at the rate of $\sqrt{N}$ in the direction of $\delta_k(x)$.

It is interesting to compare the asymptotic power of the bootstrap procedure that is based on the least favorable set of the null hypothesis. Using the same bootstrap sample $\{X^*_{ki,b} : k = 1, \cdot\cdot\cdot, K\}_{i=1}^n$, $b = 1, \ldots, B$, this bootstrap procedure alternatively consider the following bootstrap test statistics

$$T^{*LF}_{N,b} = \min_{k \neq l} \int_{\mathcal{X}} \max\left\{q(x)\sqrt{N}\bar{D}^*_{kl,b}(x), 0\right\}^2 w(x) dx. \tag{20}$$

Let the bootstrap critical values with $B \to \infty$ be denoted by $c^{*LF}_{\alpha,N,\infty}$. The results of this paper easily imply the following fact that this bootstrap procedure is certainly inferior to the procedure that this paper proposes. For brevity, we state the result for $T^{*LF}_{N,b}$.

Theorem 4 : *Suppose that the conditions of* Theorem 2 *and* Assumption 5 *hold. Under the local alternatives $P_N \in \mathcal{P} \backslash \mathcal{P}_0$ satisfying the condition in* (19),

$$\lim_{N \to \infty} P_N \left\{T_N > c^*_{\alpha,N,\infty}\right\} \geq \lim_{N \to \infty} P_N \left\{T_N > c^{*LF}_{\alpha,N,\infty}\right\}. \tag{21}$$

Furthermore, assume that for all $k \neq l$,

$$\int_{\mathcal{X}} \max\{\nu_{kl}(x), 0\}^2 w(x)dx > \int_{C_{kl}} \max\{\nu_{kl}(x), 0\}^2 w(x)dx \text{ for all } k \neq l.$$

Then the inequality in (21) *is strict.*

The result of Theorem 4 is remarkable that the bootstrap test of this paper weakly dominates the bootstrap in (20) regardless of the Pitman local alternative directions. Furthermore, when the union of the closures of the contact sets is a proper subset of the interior of $\mathcal{X}$, our test strictly dominates the bootstrap in (20) uniformly over the Pitman local alternatives. However, this latter condition for contact sets can be strong when the number K of prospects are many. In this case, a full characterization of the case for strict dominance appears to be complicated. The result of Theorem 4 is based on the nonsimilarity of the bootstrap tests in (20) on the boundary. In fact, Theorem 4 implies that the test based on the bootstrap procedure using (20) is inadmissible. This result is related to Hansen (2003)'s finding in an environment of finite-dimensional composite hypothesis testing that a test that is not asymptotically similar on the boundary is asymptotically inadmissible.

5 Monte Carlo Experiments

5.1 Simulation Designs

In this section, we examine the finite sample performance of our tests using Monte Carlo simulations. In particular, we compare our procedure with the subsampling method and recentered bootstrap method which were suggested by LMW.

For a fair comparison of the simulation results, the simulation designs we consider are exactly same as those considered by LMW: the Burr distributions also examined by Tse and Zhang (2004), the lognormal distributions also considered by Barrett and Donald (2003) and the exchangeable normal processes of Klecan et. al. (1991).

We first describe the details of the simulation designs. We consider the five different designs from the Burr Type XII distribution, $B(\alpha, \beta)$ which is a two parameter family with cdf defined by:

$$F(x) = 1 - (1 + x^{\alpha})^{-\beta}, \quad x \geq 0.$$

The parameters and the population values of d_1^*, d_2^* of the Burr designs are given below.

Design	X_1	X_2	d_1^*	d_2^*
1a	$B(4.7, 0.55)$	$B(4.7, 0.55)$	$0.000(FSD)$	$0.0000(SSD)$
1b	$B(2.0, 0.65)$	$B(2.0, 0.65)$	$0.0000(FSD)$	$0.0000(SSD)$
1c	$B(4.7, 0.55)$	$B(2.0, 0.65)$	0.1395	0.0784
1d	$B(4.6, 0.55)$	$B(2.0, 0.65)$	0.1368	0.0773
1e	$B(4.5, 0.55)$	$B(2.0, 0.65)$	0.1340	0.0761

On the other hand, we consider four different designs from the lognormal distribution $LN(\mu_k, \sigma_k^2)$. That is, we consider the distributions of $X_k = \exp(\mu_k + \sigma_k Z_k)$, where Z_k are i.i.d. $N(0, 1)$. The parameters and the population values of d_1^*, d_2^* of the lognormal designs are given below.

Design	X_1	X_2	d_1^*	d_2^*
2a	$LN(0.85, 0.6^2)$	$LN(0.85, 0.6^2)$	$0.0000(FSD)$	$0.0000(SSD)$
2b	$LN(0.85, 0.6^2)$	$LN(0.7, 0.5^2)$	$0.0000(FSD)$	$0.0000(SSD)$
2c	$LN(0.85, 0.6^2)$	$LN(1.2, 0.2^2)$	0.0834	$0.0000(SSD)$
2d	$LN(0.85, 0.6^2)$	$LN(0.2, 0.1^2)$	0.0609	0.0122

To define the multivariate normal designs, we let

$$X_{ki} = (1-\lambda)\left[\alpha_k + \beta_k\left(\sqrt{\rho}Z_{0i} + \sqrt{1-\rho}Z_{ki}\right)\right] + \lambda X_{k,i-1},$$

where (Z_{0i}, Z_{1i}, Z_{2i}) are i.i.d. standard normal random variables, mutually independent. The parameters $\lambda = \rho = 0.1$ determine the mutual correlation of X_{1i} and X_{2i} and their autocorrelation. The parameters α_k, β_k are actually the mean and standard deviation of the marginal distributions of X_{1i} and X_{2i}. The marginals and the true values of the multivariate normal designs are:

Design	X_1	X_2	d_1^*	d_2^*
3a	$N(0, 1)$	$N(-1, 16)$	0.1981	$0.0000(SSD)$
3b	$N(0, 16)$	$N(1, 16)$	$0.0000(FSD)$	$0.0000(SSD)$
3c	$N(0, 1)$	$N(1, 16)$	0.1981	0.5967

In the simulations, we consider the Kolmogorov-Smirnov type test and the Cramer-von Mises type test. In computing the suprema in the Kolmogorov-Smirnov type test, we took a maximum over an equally spaced grid of size n on the range of the pooled empirical distribution. To compute the integral for the Cramer-von Mises type test, we took the sum over the same grid points. We chose a total of 5 different subsamples for each sample size $N \in \{50, 500, 1000\}$. We took an equally spaced grid of subsample sizes: for $N = 50$, the subsample sizes are $\{20, 25, \ldots, 45\}$; for $N = 500$ the subsample sizes are $\{50, 100, \ldots, 250\}$; for $N = 1000$ the subsample sizes are $\{100, 200, \ldots, 500\}$.[5] This grid of subsamples are then used to compute the mean of the critical values from the grid, which had the best overall performance among the automatic methods considered by LMW. We used a total of 200 bootstrap repetitions in each case. In computing the suprema and integral in each subsample and bootstrap test statistic, we took the same grid of points as was used in the original test statistic.

To estimate the contact set

$$\hat{B}_{kl} = \left\{ x \in \mathcal{X} : |\bar{D}_{kl}(x)| < c_N \right\},$$

we took the tuning parameter to be $c_N = cN^{-1/3}$, where $c \in \{0.25, 0.50, 0.75\}$ for $s = 1$ and $c \in \{2.0, 3.0, 4.0\}$ for $s = 2$. To compute the test statistics and the contact set, we took the weight function $q(x) = 1$ for all $x \in \mathcal{X}$, i.e., no weighting, and set $\pi_N = N^{1/4}$. There was almost no incidence of empty contact sets and hence the choice of π_N did not affect the simulation results. In each experiment, the number of replications was $1,000$.

5.2 Simulation Results

Tables 1F - 3S present the rejection probabilities for the tests with nominal size 5%. The simulation standard error is approximately 0.007.

The overall impression is that all the methods considered work reasonably well in samples above 500. The full- sample methods work better than the subsample method under the null hypothesis in small samples . Our procedure is always more powerful than the recentered bootstrap method. In cases (1c,1d,1e,2d below) where the subsample method is better than the recentered bootstrap, our procedure is strictly more powerful than the subsample method. The performance of our procedure depend on the choice of the tuning parameter c_N. If it takes larger values, then the rejection probabilities get closer to those of the recentered bootstrap, but the rejection probabilities increase as c_N decreases. This is consistent with our theory because the rejection probability is a monotonically non-increasing function of c_N. In a reasonable range of the tuning parameter, our procedure strictly dominates the the subsample and recentered bootstrap methods in terms of power without sacrificing size in

[5] LMW considered a finer grid, i.e., 20, of different subsample sizes but their simulation results are very close to ours.

all of the designs we considered.

The first two designs in Tables 1F and 1S are to evaluate the size performance, especially in the least favorable case of the equality of two distributions, while the other three are to evaluate the power performance. The subsample method tends to over-reject under the null when $N = 50$ but the size distortions are negligible when $N \geq 500$. Our procedure has a good size performance over all tuning parameters we considered in Table 1F, but the case 1b in Table 1S shows that it tends to over-reject when the tuning parameter is chosen to be too small. The last three designs (case 1c, 1d, and 1e in Tables 1F and 1S) are quite conclusive. For moderate and large samples ($N \geq 500$), the subsample method is more powerful than the recentered bootstrap method. However, in these cases, our procedure strictly dominates the subsample method uniformly over all values of c_N we considered. This is so even at small samples ($N = 50$). This remarkable results forcefully demonstrate the power of our procedure.

Tables 2F and 2S give the rejection probabilities under the lognormal designs. The results under the least favorable case 2a are similar to those of 1a and 1b. The design 2b (in Table 2F) is quite instructive because it corresponds to the boundary case under which the two distributions "kiss" at points in the interior of the support. The recentered bootstrap tends to under-reject the null hypothesis and perform very differently from the least favorable case, while our procedure tends to have correct size for a suitable value of c_N. The design 2c in Table 2F shows that, in small sample $N = 50$, the full sample method is more powerful than the subsample method. In design 2d (Tables 2F and 2S), the subsample method is more powerful than the recentered bootstrap method for all sample sizes, but again our procedure dominates the subsample method.

Tables 3F and 3S consider the multivariate normal designs. The designs 3a (Table 2F) and 3b (Tables 3F and 3S) are to evaluate the size characteristics of the tests. The latter designs correspond to the "interior" of the null hypothesis and the rejection probabilities tend to zero as the sample size increases. This is consistent with our theory. The other designs are to evaluate the power performance. They show that the subsample method is less powerful than the recentered bootstrap and the latter is again dominated by our procedure.

Among the Kolmogorov-Smirnov and Cramer-von Mises type tests, the power performance depends on the alternatives. In designs 1c,1d,1e, and 2d, the Kolmogorov-Smirnov type tests are more powerful, while in designs 2c, 3a, and 3c, the Cramer-von Mises type tests are more powerful.

6 Conclusion

This paper proposes a new method for testing stochastic dominance that improves on the existing methods. Specifically, our tests have asymptotic sizes that are exactly correct uniformly over the entire null hypothesis. In addition, we have extended the domain of applicability of our tests to a more general class of situations where the outcome variable is the residual from some semiparametric model. Our simulation study demonstrates that our method works better than existing methods for quite modest sample sizes.

Our setting throughout has been i.i.d. data, but many time series applications call for the treatment of dependent data. In that case, one may have to use a block bootstrap algorithm in place of our wild bootstrap method. We expect, however, that similar results will obtain in that case.

7 Appendix: Mathematical Proofs

Throughout the proof, C denotes a constant that can assume different values in different places.

7.1 The Heuristics of the Proof of Theorem 2

We present the heuristics of the proof of Theorem 2 assuming simply that θ_0 and τ_0 are known. The objects of interests are then

$$\begin{aligned} a_{kl}^*(A) &= \int_A \max\{q(x)\sqrt{N}\bar{D}_{kl,b}^*(x), 0\}^2 w(x)dx \text{ and} \\ \hat{a}_{kl}(A) &= \int_A \max\{q(x)\sqrt{N}\bar{D}_{kl}(x), 0\}^2 w(x)dx, \end{aligned}$$

where $\hat{\theta}$ and $\hat{\tau}$ are replaced by θ_0 and τ_0 in the definitions of $\bar{D}_{kl,b}^*$ and $\bar{D}_{kl}$. We define $[B_{kl}]^{\varepsilon} = \{x : q(x)|D_{kl}(x)| \leq \varepsilon\}$. As preliminary results, we show the following:

Claim A:$P\{B_{kl} \subset \hat{B}_{kl} \subset [B_{kl}]^{2c_N}\} \to 1$, uniformly in $P \in \mathcal{P}$.
Claim B: $|a_{kl}^*([B_{kl}]^{2c_N}) - a_{kl}^*(B_{kl})| = O_{P^*}(c_N\sqrt{-\log c_N})$ in P, uniformly in $P \in \mathcal{P}$.
Claim C: $T_N = \int_{B_{kl}} \max\{v_{kl}(x), 0\}^2 w(x)dx + O_P(\delta_N^{\kappa} + c_N\sqrt{-\log c_N})$ in P, uniformly in $P \in \mathcal{P}_0$.

Claim A follows from the weak convergence of the empirical process $\sqrt{N}q(\cdot)\{\bar{D}_{kl}(\cdot) - D_{kl}(\cdot)\}$. Claim B is obtained by translating the difference into the oscillation of the bootstrap empirical processes over the set $[B_{kl}]^{2c_N} \backslash B_{kl}$. The perturbation over $[B_{kl}]^{2c_N} \backslash B_{kl}$ is at best of order $O(c_N)$. The maximal inequality of the process bounds this oscillation of the bootstrap empirical process by the bracketing integral entropy of the process, resulting in the order $c_N\sqrt{-\log c_N}$. Claim C is rather involved requiring a delicate choice of c_N as in Theorem 2. The result is essentially obtained in the proof of Theorem 1. Given Claims A, B and C, the steps of the proof proceeds as follows.

Suppose that $P \in \mathcal{P}_{0,N}$. By Claim A, we can show that with probability approaching one,

$$\min_{k \neq l} a_{kl}^*(B_{kl}) \leq T_{N,b}^* \leq \min_{k \neq l} a_{kl}^*([B_{kl}]^{2c_N}),$$

which implies by Claim B that $T^*_{N,b} = \min_{k\neq l} a^*_{kl}(B_{kl}) + O_{P^*}(c_N\sqrt{-\log c_N})$. Hence the treatment of the estimated contact sets is completed. The remaining step is to apply the bootstrap CLT to $q(x)\sqrt{N}\bar{D}^*_{kl,b}(x)$ and the Berry Esseen bound to obain the normal approximation error.

Now, suppose that $P \in \mathcal{P}_0\backslash\mathcal{P}_{0,N}$. In this case, we can show that with probability approaching one,

$$\pi_N \leq T^*_{N,b},$$

so that $c^*_{\alpha,N,\infty} \geq \pi_N$ with probability approaching one. Writing $T_N = \min_{k\neq l}\hat{a}_{kl}(\mathcal{X})$, we note that

$$\begin{aligned} P\left\{\min_{k\neq l}\hat{a}_{kl}(\mathcal{X}) \geq c^*_{\alpha,N,\infty}\right\} &\leq P\left\{\max_{k\neq l} 4\int_{\mathcal{X}} \max\{q(x)\sqrt{N}(\bar{D}_{kl}(x) - D_{kl}(x)), 0\}^2 w(x)dx \geq \pi_N/2\right\} \\ &\quad + P\left\{\min_{k\neq l} 4\int_{\mathcal{X}} \max\{q(x)\sqrt{N}D_{kl}(x), 0\}^2 w(x)dx \geq \pi_N/2\right\}, \end{aligned}$$

with probability approaching one. By the asymptotic tightness of the empirical process, the first probability converges to zero because $\pi_N \to \infty$. Since we are under the null hypothesis, $\int_{\mathcal{X}} \max\{q(x)\sqrt{N}D_{kl}(x), 0\}^2 w(x)dx = 0$, making the second probability zero. Hence the rejection probability converges to zero in this case.

7.2 The Proofs of the Main Results

Lemma A1 : *For a bounded map $V(x) : \mathcal{X} \to [-M, M]$, $M \in (0, \infty)$, let us introduce a pseudo metric $d_V(x, x') = |V(x) - V(x')|$. Then, for each $\varepsilon \in (0, 1]$,*

$$\log N(\varepsilon, \mathcal{X}, d_V) \leq C\{1 - \log(\varepsilon)\}.$$

Proof of Lemma A1 : Fix $\varepsilon > 0$ and partition $[-M, M] = \cup_{j=1}^N \mathcal{I}_j$, where $\mathcal{I}_j$ is an interval of length ε and $N = 2M/\varepsilon$. Then, choose $\{v_j\}_{j=1}^N$ to be the centers of $\mathcal{I}_j$, $j = 1, \cdots, N$, so that the sets $V_j = \{x \in \mathcal{X} : |V(x) - v_j| \leq \varepsilon/2\}$, $j = 1, \cdots, N$, cover $\mathcal{X}$. These sets V_j have radius bounded by $\varepsilon/2$ with respect to d_V. Redefining constants, we obtain the inequality. ■

Lemma A2 : *Let $\mathcal{F} = \{h_x(\varphi_k(\cdot; \theta, \tau)) : (x, \theta, \tau) \in \mathcal{X} \times B_{\Theta\times\mathcal{T}}(\delta)\}$. Then, for each $\varepsilon \in (0, 1]$,*

$$\sup_{P\in\mathcal{P}} \log N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||_{P,2}) \leq C\varepsilon^{-d\lambda/s_2}.$$

Proof of Lemma A2 : Let $\mathcal{H} = \{h_x : x \in \mathcal{X}\}$ and $\gamma_x(y) = 1\{y \leq x\}$. Simply write $\varphi = \varphi_k$. Define $\Phi = \{\varphi(\cdot; \theta, \tau) : (\theta, \tau) \in B_{\Theta\times\mathcal{T}}(\delta)\}$. Then by the local uniform L_2-continuity condition in Assumption 2(ii)(A)(c), we have

$$\log N_{[]}(\varepsilon, \Phi, ||\cdot||_{P,2}) \leq \log N(C\varepsilon^{1/s_2}, B_{\Theta\times\mathcal{T}}(\delta), ||\cdot|| + ||\cdot||_\infty) \leq C\varepsilon^{-d/s_2}. \quad (22)$$

Choose ε-brackets $(\varphi_j, \Delta_{1,j})_{j=1}^{N_1}$ of Φ such that $\int \Delta^2_{1,j}dP \leq \varepsilon^2$. For each j, let $Q_{j,P}$ be the distribution of $\varphi_j(W)$, where W is distributed as P. By Assumption 2(ii)(A)(b), we can take $\Delta_{1,j}$ such that $\Delta_{1,j}(W)$ is measurable with respect to the σ-field of W_1.

Suppose $s = 1$. We let $V(x) = xq(x)$ and $h(x) = (\log x)^{s_2/d}$. Fix $\varepsilon > 0$. Since $|V(x)x^\delta| < \infty$ as $|x| \to \infty$ by Assumption 3(iii), for some $C > 0$, $V(x) < C/h(x)$ for all $x > 1$. Put $x = h^{-1}(2C/\varepsilon)$ where

$h^{-1}(y) =$inf$\{x : h(x) \geq y : x \in \mathcal{X}\}$, so that we obtain

$$V(h^{-1}(2C/\varepsilon)) < \varepsilon/2.$$

Then, we can choose $L(\varepsilon) \leq Ch^{-1}(2C/\varepsilon)$ such that $L(\varepsilon) \geq 1$ and $\sup_{|x|\geq L(\varepsilon)}V(x) < \varepsilon/2$. Partition $[-L(\varepsilon), L(\varepsilon)]$ into intervals $\{\mathcal{I}_m\}_{m=1}^N$of length ε with the number of intervals N not greater than $2L(\varepsilon)/\varepsilon$. Let $\mathcal{X}_0 = \mathcal{X}\backslash[-L(\varepsilon), L(\varepsilon)]$ and $\mathcal{X}_m = \mathcal{I}_m \cap \mathcal{X}$. Then $(\cup_{m=1}^N \mathcal{X}_m) \cup \mathcal{X}_0$ constitutes the partition of $\mathcal{X}$ with the number of partitions bounded by $2L(\varepsilon)/\varepsilon + 1$ or by $Ch^{-1}(2C/\varepsilon)/\varepsilon + 1$.

Now, choose any $(\varphi, x) \in \Phi \times \mathcal{X}$ and let (φ_j, x_m) be such that $|\varphi - \varphi_j| < \Delta_{1,j}$ and x_m is the center of $\mathcal{I}_m$ if $x \in \mathcal{X}_m$ for some $m \geq 1$ and x_m is any arbitrary member of $\mathcal{X}_m$ if $x \in \mathcal{X}_m$ with $m = 0$ Define $\Phi_j = \{\varphi \in \Phi : |\varphi - \varphi_j| \leq \Delta_{1,j}\}$. Then observe that

$$\begin{aligned} &\sup_{\varphi\in\Phi j}\sup_{x\in\mathcal{X}_m} \left|\gamma_x(\varphi(w))q(x) - \gamma_{x_m}(\varphi_j(w))q(x_m)\right| \\ &\leq \sup_{\varphi\in\Phi j}\sup_{x\in\mathcal{X}_m} \left|\gamma_x(\varphi(w))q(x) - \gamma_{x_m}(\varphi_j(w))q(x)\right| \\ &+\sup_{\varphi\in\Phi j}\sup_{x\in\mathcal{X}_m} \left|\gamma_{x_m}(\varphi_j(w))q(x) - \gamma_{x_m}(\varphi_j(w))q(x_m)\right| = \Delta^*_{m,j}(w), \text{ say.} \end{aligned} \tag{23}$$

As for the first term on the right-hand side of the inequality,

$$\begin{aligned} & \mathbf{E}[\sup_{\varphi\in\Phi j}\sup_{x\in\mathcal{X}_m}|\gamma_x(\varphi(W)) - \gamma_{x_m}(\varphi_j(W))|^2 q(x)^2|W_1] \\ \leq\ & \mathbf{E}[\sup_{x\in\mathcal{X}_m} 1\{x - \Delta_{1,j}(W_1) - |x - x_m| \leq \varphi_j(W) \leq x + \Delta_{1,j}(W) + |x - x_m|\}q(x)^2|W_1] \\ \leq\ & \sup_{x\in\mathcal{X}_0} q(x)^2 1\{m = 0\} + C\{\varepsilon + \Delta_{1,j}(W)\} 1\{m \geq 1\} \\ \leq\ & C\varepsilon^2 1\{m = 0\} + C\{\varepsilon + \Delta_{1,j}(W)\} 1\{m \geq 1\} \leq C\varepsilon + C\Delta_{1,j}(W). \end{aligned}$$

The second inequality is obtained by splitting the supremum into the case $m = 0$ and the case $m \geq 1$. Here we also used the fact that $\Delta_{1,j}(W)$ is measurable with respect to the σ-field of W_1. The third inequality follows by Assumption 2(ii)(A)(a) and by the fact that $\sup_{x\in\mathcal{X}_0} q(x)^2 = \sup_{|x|>L(\varepsilon)} q(x)^2 \leq \sup_{|x|>L(\varepsilon)} |x|^2 q(x)^2 = \sup_{|x|>L(\varepsilon)} V^2(x) \leq \varepsilon^2/4$. Note also that

$$\mathbf{E}\left[\sup_{\varphi\in\Phi j}\sup_{x\in\mathcal{X}_m} \left|\gamma_{x_m}(\varphi_j(W))q(x) - \gamma_{x_m}(\varphi_j(W))q(x_m)\right|^2 |W_1\right] \leq C|q(x) - q(x_m)|^2 \leq C\varepsilon^2,$$

because q is Lipschitz continuous. We conclude that $||\Delta^*_{k,j}||_{P,2} \leq C\varepsilon^{1/2}$. Hence by taking appropriate constants C,

$$\begin{aligned} \sup_{P\in\mathcal{P}} \log N_{[]}(C\varepsilon^{1/2}, \mathcal{F}, ||\cdot||_{P,2}) &\leq \log N_{[]}(C\varepsilon, \Phi, ||\cdot||_{P,2}) + \log\left(Ch^{-1}(2C/\varepsilon)/\varepsilon + 1\right) \\ &\leq C\{\varepsilon^{-d/s_2} - \log(\varepsilon)\}. \end{aligned}$$

Suppose $s > 1$. Then $(x - \varphi)^{s-1}\gamma_x(\varphi)q(x)$ is Liptschitz continuous in φ with the coefficient bounded by $C|x - \varphi|^{s-2}q(x) < C$. Therefore,

$$\log N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||_{P,2}) \leq \log N_{[]}(C\varepsilon, \Phi, ||\cdot||_{P,2}).$$

Observe that the constants above do not depend on the choice of the measure P. Combined with (22), we obtain the wanted result. ■

Proof of Theorem 1 : For $h^{\Delta}_{x,kl}$ defined in (10), we write

$$\begin{aligned}
&\sqrt{N}\bar{D}_{kl}(x,\hat{\theta},\hat{\tau}) - \sqrt{N}\mathbf{E}h^{\Delta}_{x,kl}(W_i;\theta_0,\tau_0) \\
= \;& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{h^{\Delta}_{x,kl}(W_i;\theta_0,\tau_0) - \mathbf{E}h^{\Delta}_{x,kl}(W_i;\theta_0,\tau_0)\right\} + \sqrt{N}(\Gamma_{k,P}-\Gamma_{l,P})(x)[\hat{\theta}-\theta_0,\hat{\tau}-\tau_0] + \zeta_{1N} + \zeta_{2N},
\end{aligned}$$

where

$$\begin{aligned}
\zeta_{1N} &= \sqrt{N}\left\{\bar{D}_{kl}(x,\hat{\theta},\hat{\tau}) - \bar{D}_{kl}(x,\theta_0,\tau_0) - (D_{kl}(x,\hat{\theta},\hat{\tau}) - D_{kl}(x,\theta_0,\tau_0))\right\} \text{ and} \\
\zeta_{2N} &= \sqrt{N}(D_{kl}(x,\hat{\theta},\hat{\tau}) - D_{kl}(x,\theta_0,\tau_0)) - \sqrt{N}(\Gamma_{k,P}-\Gamma_{l,P})(x)[\hat{\theta}-\theta_0,\hat{\tau}-\tau_0].
\end{aligned}$$

By Assumption 3(ii), $\zeta_{2N} = O_P(\delta_N)$ uniformly in $P \in \mathcal{P}$. We now show that $\zeta_{1N} = o_P(1)$. Let $\mathcal{H}_{kl} = \{h^{\Delta}_{x,kl}(\cdot;\theta,\tau) : (x,\theta,\tau) \in \mathcal{X} \times B_{\Theta\times\mathcal{T}}(\delta)\}$ where $h^{\Delta}_{x,kl}(\cdot;\theta,\tau) = h_x(\varphi_k(\cdot;\theta,\tau)) - h_x(\varphi_l(\cdot;\theta,\tau))$. The bracketing entropy of this class at $\varepsilon \in (0,1]$ is bounded by $C\varepsilon^{-d\lambda/s_2}$ by Lemma A2. It is easy to show that the $L_2(P)$-norm of its envelope is $O(\delta_N^{\lambda s_2/2})$. Hence by using the maximal inequality and using the fact that $d\lambda/s_2 < 2$, we obtain $\zeta_{1N} = O_P(\delta_N^{\lambda(s_2-d\lambda/2)/2})$. Hence by Assumption 3(ii) and noting that $\mathbf{E}h^{\Delta}_{x,kl}(W_i;\theta_0,\tau_0) = q(x)D_{kl}(x)$,

$$\sqrt{N}q(x)\{\bar{D}_{kl}(x,\hat{\theta},\hat{\tau}) - D_{kl}(x)\} = \eta_N^{kl}(x) + O_P(\delta_N^{\kappa}), \text{ uniformly in } x \in B_{\Theta\times\mathcal{T}}(\delta_N), \tag{24}$$

where $\kappa = \lambda(s_2 - d\lambda/2)/2$,

$$\eta_N^{kl}(x) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{(h^{\Delta}_{x,kl}+\psi^{\Delta}_{x,kl})(W_i;\theta_0,\tau_0) - \mathbf{E}\left[(h^{\Delta}_{x,kl}+\psi^{\Delta}_{x,kl})(W_i;\theta_0,\tau_0)\right]\right\}. \tag{25}$$

We turn to $\eta_N^{kl}(x)$. Define $\mathcal{H}^0_{kl} = \{h^{\Delta}_{x,kl}(\cdot) : x \in \mathcal{X}\}$ and $\Psi_{kl} = \{\psi^{\Delta}_{x,kl} : x \in \mathcal{X}\}$. Consider the class of functions $\mathcal{F}_{kl} = \{h + \psi : (h,\psi) \in \mathcal{H}_{kl} \times \Psi_{kl}\}$. Using Assumption 3(ii), Lemmas A1 and A2 above, we can show that $\sup_{P\in\mathcal{P}} \log N_{[]}(\varepsilon, \mathcal{F}_{kl}, ||\cdot||_{P,2}) < C\log\varepsilon$. Also note that $\mathcal{H}_{kl}$ and Ψ_{kl} are uniformly bounded. By Proposition 3.1 and Theorem 2.3 of Giné and Zinn (1991) (See also Theorem 2.8.4 of van der Vaart and Wellner (1996)), $\mathcal{F}_{kl}$ is a uniform P-Donsker class. The computation of the covariance kernel of the limiting Gaussian process is straightforward. Therefore,

$$\eta_N^{kl}(\cdot) \Longrightarrow \nu_{kl}(\cdot), \tag{26}$$

uniformly in $P \in \mathcal{P}$.

Now we show that for any sequence $\varepsilon_N \to 0$ such that $\sqrt{N}\varepsilon_N \to \infty$ and $\lim_{N\to\infty}\sqrt{N}\varepsilon_N/(-\log\varepsilon_N)^{1/4} < \infty$,

$$T_N = \min_{k\neq l}\int_{B_{kl}} \max\left\{\eta_N^{kl}(x), 0\right\}^2 w(x)dx + O_P(\delta_N^{\kappa} + \varepsilon_N\sqrt{-\log\varepsilon_N}), \tag{27}$$

uniformly in $P \in \mathcal{P}_0$. The rate above is used in the proof of Theorem 2. To show this, first write (using

(24)),

$$\int_{\mathcal{X}} \max\left\{\sqrt{N}q(x)\bar{D}_{kl}(x,\hat{\theta},\hat{\tau}),0\right\}^2 w(x)dx \tag{28}$$

$$= \int_{[B_{kl}]^{\varepsilon_N}} \max\left\{\eta_N^{kl}(x)+\sqrt{N}q(x)D_{kl}(x),0\right\}^2 w(x)dx$$

$$+\int_{\mathcal{X}\backslash[B_{kl}]^{\varepsilon_N}} \max\{\eta_N^{kl}(x)+\sqrt{N}q(x)D_{kl}(x),0\}^2 w(x)dx + O_P(\delta_N^{\kappa}),$$

where $[B_{kl}]^{\varepsilon_N} = \{x : q(x)|D_{kl}(x)| \leq \varepsilon_N\}$. If the second integral diverges to infinity, this pair k and l does not affect the test statistic because we are under the null hypothesis, so we assume that k and l are such that the second integral is stochastically bounded. Since for any $x \in \mathcal{X}\backslash[B_{kl}]^{\varepsilon_N}$, $\sqrt{N}q(x)|D_{kl}(x)| \geq \sqrt{N}\varepsilon_N \to \infty$, this is possible only when for all $x \in \mathcal{X}\backslash[B_{kl}]^{\varepsilon_N}$ except for $w(x)dx$-null sets, $\sqrt{N}q(x)D_{kl}(x) \leq -\sqrt{N}\varepsilon_N \to -\infty$. Observe that for any sequence $q_N \to 0$,

$$P\left\{\int_{\mathcal{X}\backslash[B_{kl}]^{\varepsilon_N}} \max\{\eta_N^{kl}(x)+\sqrt{N}q(x)D_{kl}(x),0\}^2 w(x)dx > q_N\right\}$$

$$\leq P\left\{\int_{\mathcal{X}\backslash[B_{kl}]^{\varepsilon_N}} \max\{\eta_N^{kl}(x)-\sqrt{N}\varepsilon_N,0\}^2 w(x)dx > q_N\right\}.$$

The integral in the probability is not zero only when $\eta_N^{kl}(x) > \sqrt{N}\varepsilon_N$, $w(x)dx$-a.e. Hence the above probability is bounded by

$$P\left\{\sup_{x\in\mathcal{X}}|\eta_N^{kl}(x)| > \sqrt{N}\varepsilon_N\right\}.$$

This probability converges to zero because the sequence $\sup_{x\in\mathcal{X}}|\eta_N^{kl}(x)|$ is asymptotically tight by Theorem 1. Therefore, the last integral (28) vanishes at an arbitrary rate. Hence it suffices to consider only the second to the last integral in (28) which we write as

$$\int_{B_{kl}} \max\left\{\eta_N^{kl}(x),0\right\}^2 w(x)dx + \int_{[B_{kl}]^{\varepsilon_N}\backslash B_{kl}} \max\left\{\eta_N^{kl}(x)+\sqrt{N}q(x)D_{kl}(x),0\right\}^2 w(x)dx. \tag{29}$$

The last integral is bounded by

$$\int_{[B_{kl}]^{\varepsilon_N}\backslash B_{kl}} \max\left\{\eta_N^{kl}(x)+2\sqrt{N}\varepsilon_N,0\right\}^2 w(x)dx \tag{30}$$

$$\leq \int_{[B_{kl}]^{\varepsilon_N}\backslash B_{kl}} \max\left\{\eta_N^{kl}(x),0\right\}^2 w(x) + \int_{[B_{kl}]^{\varepsilon_N}\backslash B_{kl}} (2\sqrt{N}\varepsilon_N)^2 w(x)dx.$$

The $w(x)dx$ measure of the set $[B_{kl}]^{\varepsilon_N}\backslash B_{kl}$ is of order $O(\varepsilon_N)$ by Assumption 3(iii)(b) while the integrand is of order $O(N\varepsilon_N^2)$. Therefore, the last integral is $O(N\varepsilon_N^3) = O(\varepsilon_N\sqrt{-\log\varepsilon_N})$ by the design of ε_N. The first integral in the last line of (30) is bounded by

$$\int_{\mathcal{X}} \left|\max\left\{\eta_N^{kl}(x),0\right\}^2 1\{x\in[B_{kl}]^{\varepsilon_N}\} - \max\left\{\eta_N^{kl}(x),0\right\}^2 1\{x\in B_{kl}\}\right| w(x)dx$$

$$\leq C\sup_{x,x':d_{kl}(x,x')<C\varepsilon_N} \left|\max\left\{\eta_N^{kl}(x),0\right\}^2 - \max\left\{\eta_N^{kl}(x'),0\right\}^2\right|,$$

where d_{kl} is defined in Assumption 3(iv). Using $|\max(a,0)^2 - \max(b,0)^2| \leq |a^2-b^2| = (a+b)(a-b)$ and the Cauchy-Schwartz inequality, the expectation of the above is bounded by

$$4\left\{\mathbf{E}\sup_{x\in\mathcal{X}}\left|\eta_N^{kl}(x)\right|^2\right\}^{1/2}\times\left\{\mathbf{E}\sup_{x,x':d_{kl}(x,x')<C\varepsilon_N}\left|\eta_N^{kl}(x)-\eta_N^{kl}(x')\right|^2\right\}^{1/2}.$$

The first supremum is $O_P(1)$ by the weak convergence in (26). By using Assumption 3(iv), the second expectation is $O(\varepsilon_N)$. Therefore, we conclude that the second to the last integral in (28) is $O(\varepsilon_N\sqrt{-\log\varepsilon_N})$. This gives us the wanted result of (27). ■

The following lemma is a key step for the bootstrap consistency of the test (Theorem 2). In the following, we use the usual bootstrap stochastic convergence notations o_{P^*} and O_{P^*} with respect to the conditional distribution given $\mathcal{G}_N$. Let $l_\infty(\mathcal{X}\times B_{\Theta\times\mathcal{T}}(\delta))$ be the space of real bounded functions on $\mathcal{X}\times B_{\Theta\times\mathcal{T}}(\delta)$. Let $V_{x,kl}(w;\theta,\tau)$ be as defined in Assumption 3(iv) and denote

$$\nu_{kl}^*(x;\theta,\tau)=\frac{1}{\sqrt{N}}\sum_{i=1}^N\left\{V_{x,kl}(W_{i,b}^*;\theta,\tau)-\frac{1}{N}\sum_{i=1}^N V_{x,kl}(W_i;\theta,\tau)\right\}. \quad (31)$$

Lemma A3 : (i) *Under the assumptions of* Theorem 2,

$$\sqrt{N}\bar{D}_{kl,b}^*(x)=\nu_{kl}^*(x,\hat{\theta},\hat{\tau})+O_{P^*}(\delta_N),$$

in P uniformly in $P\in\mathcal{P}$.
(ii) *For any $\delta>0$, $\nu_{kl}^*(\cdot;\hat{\theta},\hat{\tau})\to\nu_{kl}$ weakly in $l_\infty(\mathcal{X}\times B_{\Theta\times\mathcal{T}}(\delta))$ conditional on $\mathcal{G}_N$ in P uniformly in $P\in\mathcal{P}$, where ν_{kl} is a Gaussian process on $\mathcal{X}\times B_{\Theta\times\mathcal{T}}(\delta)$ whose covariance kernel is given by $C_{kl}(\cdot,\cdot)$.*

Proof of Lemma A3 : (i) Recall the definitions of $h_{x,kl}^\Delta$ and $\psi_{x,kl}^\Delta$ in (10). We write

$$\begin{aligned}\sqrt{N}\bar{D}_{kl,b}^*(x) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N h_{x,kl}^\Delta(W_i^*;\hat{\theta}^*,\hat{\tau}^*)-\frac{1}{\sqrt{N}}\sum_{i=1}^N h_{x,kl}^\Delta(W_i^*;\hat{\theta},\hat{\tau})\\ &\quad+\frac{1}{\sqrt{N}}\sum_{i=1}^N h_{x,kl}^\Delta(W_i^*;\hat{\theta},\hat{\tau})-\frac{1}{\sqrt{N}}\sum_{i=1}^N h_{x,kl}^\Delta(W_i;\hat{\theta},\hat{\tau})\\ &= \hat{\Gamma}_{k,P}(x)-\hat{\Gamma}_{l,P}(x)+\frac{1}{\sqrt{N}}\sum_{i=1}^N h_{x,kl}^\Delta(W_i^*;\hat{\theta},\hat{\tau})-\frac{1}{\sqrt{N}}\sum_{i=1}^N h_{x,kl}^\Delta(W_i;\hat{\theta},\hat{\tau}),\end{aligned}$$

where $\hat{\Gamma}_{k,P}(x)$ is as defined in Assumption 4. By Assumption 4, the leading term $\hat{\Gamma}_{k,P}(x)-\hat{\Gamma}_{l,P}(x)$ is written as

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N\left\{\psi_{x,kl}^\Delta(W_i^*;\hat{\theta},\hat{\tau})-\frac{1}{N}\sum_{i=1}^N\psi_{x,kl}^\Delta(W_i;\hat{\theta},\hat{\tau})\right\}+O_{P_*}(\delta_N). \quad (32)$$

Hence we obtain the wanted result.

(iii) Let $\mathcal{H}_{kl}(\delta)=\{V_{x,kl}(\cdot;\theta,\tau):(x,\theta,\tau)\in\mathcal{X}\times B_{\Theta\times\mathcal{T}}(\delta)\}$. Then, it suffices to show Pollard's entropy condition:

$$\int_0^\infty\sup_{P\in\mathcal{P}}\sqrt{\log N_{[]}(\varepsilon,\mathcal{H}_{kl}(\delta),||\cdot||_{P,2})}d\varepsilon<\infty. \quad (33)$$

Then by Proposition 3.1 of Giné and Zinn (1991), $\mathcal{H}_{kl}$ is uniformly pregaussian and hence by Corollary 2.7 of Giné and Zinn (1991), the wanted uniform weak convergence follows.

We split the proof of (33) into that for $\mathcal{H}_{1,kl}(\delta) = \{h^{\Delta}_{x,kl}(\cdot;\theta,\tau) : (x,\theta,\tau) \in \mathcal{X} \times B_{\Theta\times\mathcal{T}}(\delta)\}$ and that for $\mathcal{H}_{2,kl}(\delta) = \{\psi^{\Delta}_{x,kl}(\cdot;\theta,\tau) : (x,\theta,\tau) \in \mathcal{X} \times B_{\Theta\times\mathcal{T}}(\delta)\}$. As we saw in the proof of Theorem 1 the bracketing entropy of $\mathcal{H}_{1,kl}(\delta)$ is bounded by $C\varepsilon^{-d\lambda/s_2}$ by Lemma A2. Introduce the pseudo metric $d_V(x,x') = |D_{kl}(x)q(x) - D_{kl}(x')q(x')|$. By the local uniform L_2-continuity condition for the elements in $\mathcal{H}_{2,kl}(\delta)$ (Assumption 4), we deduce that

$$\begin{aligned} \sup_{P\in\mathcal{P}} \log N_{[]}(\varepsilon, \mathcal{H}_{2,kl}(\delta), ||\cdot||_{P,2}) &\leq C\sup_{P\in\mathcal{P}} \log N(\varepsilon, \mathcal{X}, d_V) \\ &\quad + C\sup_{P\in\mathcal{P}} \log N_{[]}(\varepsilon^{1/s_3}, B_{\Theta\times\mathcal{T}}(\delta), ||\cdot|| + ||\cdot||_\infty) \\ &\leq -C\log(\varepsilon) + C\varepsilon^{-d/s_3}, \end{aligned} \tag{34}$$

by Lemma A1. By applying Theorem 6 of Andrews (1994) and using the condition that $d/s_3 < 2$ and $d\lambda/s_2 < 2$, we obtain (33). ■

Proof of Theorem 2 : (i) Define $\nu^*_{kl,b}(x;\theta,\tau)$ as in (31). We also define $1_K = 1\{\int_{B_{kl}} w(x)dx > z_N$ for all $k \neq l\}$ and $1_K^{2c_N} = 1\{\int_{[B_{kl}]^{2c_N}} w(x)dx > z_N$ for all $k \neq l\}$ and $\hat{1}_K = 1\{\int_{\hat{B}_{kl}} w(x)dx > z_N$ for all $k \neq l\}$. For any set $A \in \mathcal{X}$, let us denote:

$$\begin{aligned} \tilde{a}^*_{kl}(A) &= \int_A \max\{q(x)\sqrt{N}\bar{D}^*_{kl,b}(x), 0\}^2 w(x)dx, \\ \hat{a}^*_{kl}(A) &= \int_A \max\{\nu^*_{kl}(x;\hat{\theta},\hat{\tau}), 0\}^2 w(x)dx, \\ a^*_{kl}(A) &= \int_A \max\{\nu^*_{kl}(x;\theta_0,\tau_0), 0\}^2 w(x)dx, \text{ and} \\ a_{kl}(A) &= \int_A \max\{\nu_{kl}(x), 0\}^2 w(x)dx, \end{aligned}$$

where ν^*_{kl} is defined in (31) and ν_{kl} is the Gaussian process in Theorem 1. We will show the following at the end of the proof:

Claim 1: $P\{B_{kl} \subset \hat{B}_{kl} \subset [B_{kl}]^{2c_N}\} \to 1$, uniformly in $P \in \mathcal{P}$.
Claim 2: For any $A \subset \mathcal{X}$, $|\tilde{a}^*_{kl}(A) - a^*_{kl}(A)| = O_{P^*}(\delta_N^{\kappa})$ in P, uniformly in $P \in \mathcal{P}$.
Claim 3: $|a^*_{kl}([B_{kl}]^{2c_N}) - a^*_{kl}(B_{kl})| = O_{P^*}(c_N\sqrt{-\log c_N})$ in P, uniformly in $P \in \mathcal{P}$.

Write $T^*_{N,b} = \min_{k\neq l} \tilde{a}^*_{kl}(\hat{B}_{kl})\hat{1}_K + \pi_N(1-\hat{1}_K)$. By Claim 1, $1_K \leq \hat{1}_K \leq 1_K^{2c_N}$ with probability approaching one. It follows that with probability (uniformly over $P \in \mathcal{P}$) approaching one, the conditional probability given $\mathcal{G}_N$ of

$$\min_{k\neq l} \tilde{a}^*_{kl}(B_{kl})1_K + \pi_N(1 - 1_K^{2c_N}) \leq T^*_{N,b} \leq \min_{k\neq l} \tilde{a}^*_{kl}([B_{kl}]^{2c_N})1_K^{2c_N} + \pi_N(1 - 1_K) \tag{35}$$

is equal to one. Suppose that $P \in \mathcal{P}_{0,N}$. In this case, $1_K = 1_K^{2c_N} = 1$ by the definition of $\mathcal{P}_{0,N}$. From above inequalities, we deduce that with probability approaching one,

$$\min_{k\neq l} \tilde{a}^*_{kl}(B_{kl}) \leq T^*_{N,b} \leq \min_{k\neq l} \tilde{a}^*_{kl}([B_{kl}]^{2c_N}).$$

Therefore,

$$\begin{aligned}\left|T_{N,b}^{*}-\min_{k\neq l}\tilde{a}_{kl}^{*}(B_{kl})\right| &\leq \left|\min_{k\neq l}\tilde{a}_{kl}^{*}([B_{kl}]^{2c_N})-\min_{k\neq l}\tilde{a}_{kl}^{*}(B_{kl})\right|\leq\sum_{k\neq l}\left|\tilde{a}_{kl}^{*}([B_{kl}]^{2c_N})-\tilde{a}_{kl}^{*}(B_{kl})\right| \\ &\leq \sum_{k\neq l}\left|a_{kl}^{*}([B_{kl}]^{2c_N})-a_{kl}^{*}(B_{kl})\right|+O_{P^*}(\delta_N^{\kappa})=O_{P^*}(c_N\sqrt{-\log c_N})+O_{P^*}(\delta_N^{\kappa})\end{aligned}$$

uniformly in P, by Claims 2 and 3. This implies that

$$\begin{aligned}T_{N,b}^{*} &= \min_{k\neq l}\tilde{a}_{kl}^{*}(B_{kl})+O_{P^*}(c_N\sqrt{-\log c_N})+O_{P^*}(\delta_N^{\kappa}) \\ &= \min_{k\neq l}a_{kl}^{*}(B_{kl})+O_{P^*}(c_N\sqrt{-\log c_N})+O_{P^*}(\delta_N^{\kappa}),\end{aligned}$$

again by Claim 2.

Now, we focus on $a_{kl}^{*}(B_{kl})$. Let $l_{\infty}^{0}(\mathcal{X})$ be the space of bounded real functions $f:\mathcal{X}\to\mathbf{R}$ such that $\{x\in\mathcal{X}:f(x)=0\}$ has Lebesgue measure zero. Since the Gaussian process ν_{kl} has continuous sample paths and a positive definite covariance function, almost all the realizations of ν_{kl} belong to $l_{\infty}^{0}(\mathcal{X})$. It is easy to see that the functional F on $l_{\infty}^{0}(\mathcal{X})$ defined by $Fg=\int_{B_{kl}}\max\{g(x),0\}^2w(x)dx$ is infinite-order Frechet differentiable. Furthermore, h_x and $\psi_{x,k}$ are uniformly bounded. Lastly, the bootstrap empirical process $\nu_{kl}^{*}(\cdot;\theta_0,\tau_0)$ weakly converges to $\nu_{kl}(\cdot)$ conditional on $\mathcal{G}_N$ in P by Lemma A3(ii). Hence, we can apply the Berry-Esseen type bound for the tail probability to obtain that (e.g. Theorem 1.1. of Bentkus and Götze (1993))

$$P\left\{a_{kl}^{*}(B_{kl})>c|\mathcal{G}_N\right\}=P\left\{a_{kl}(B_{kl})>c\right\}+O_P(N^{-1/2}),$$

uniformly on c. Therefore, $c_{\alpha,N,\infty}^{*}=c_{\alpha}+O_P(\delta_N^{\kappa}+c_N\sqrt{-\log c_N})$, uniformly in $P\in\mathcal{P}_{0,N}$, where c_{α} is such that

$$c_{\alpha}=\inf\left\{c\in\mathbf{R}_{+}:P\left\{\min_{k\neq l}a_{kl}(B_{kl})\leq c\right\}\geq 1-\alpha\right\}.$$

By the rate result in (27) (using c_N in place of ε_N there) and applying the Berry-Esseen type bound to the weak convergence in (26), we obtain that

$$c_{\alpha,N}=c_{\alpha}+O_P(\delta_N^{\kappa}+c_N\sqrt{-\log c_N}),$$

where $c_{\alpha,N}$ =inf$\{c\in\mathbf{R}_{+}:P\left\{T_N\leq c\right\}\geq 1-\alpha\}$. Therefore, we obtain the wanted result for $P\in\mathcal{P}_{0,N}$.

Now, suppose that $P\in\mathcal{P}_0\backslash\mathcal{P}_{0,N}$. In this case, for some k,l, $\int_{B_{kl}}w(x)dx\leq z_N$ and $1_K=0$. Recall that by Claim 1, with probability approaching one,

$$P\{\pi_N(1-\hat{1}_K)\leq T_{N,b}^{*}|\mathcal{G}_N\}=1.$$

Observe that

$$\begin{aligned}P\left\{\hat{1}_K=0\right\} &= P\left\{\int_{\hat{B}_{kl}}w(x)dx\leq z_N\right\} \\ &= P\left\{\int_{\hat{B}_{kl}\backslash B_{kl}}w(x)dx\leq z_N\right\}\geq 1\left\{\int_{[B_{kl}]^{c_N}\backslash B_{kl}}w(x)dx\leq z_N\right\}.\end{aligned}$$

The last indicator function is equal to one after some large N by Assumption 3(iii)(b) because $c_N/z_N\to 0$.

Therefore, $P\{\pi_N \leq T^*_{N,b}|\mathcal{G}_N\} = 1$ with probability approaching one, which implies that $c^*_{\alpha,N,\infty} \geq \pi_N$ with probability approaching one. Now, observing that $\max(a+b,0) \leq \max(a,0) + \max(b,0)$,

$$\begin{aligned}
& P\left\{\min_{k\neq l}\int_{\mathcal{X}} \max\{\sqrt{N}q(x)\bar{D}_{kl}(x),0\}^2 w(x)dx > c^*_{\alpha,N,\infty}\right\} \\
\leq\ & P\left\{\max_{k\neq l} 4\int_{\mathcal{X}} \max\{\sqrt{N}q(x)(\bar{D}_{kl}(x)-D_{kl}(x)),0\}^2 w(x)dx > \pi_N/2\right\} \\
& + P\left\{\min_{k\neq l} 4\int_{\mathcal{X}} \max\{\sqrt{N}q(x)D_{kl}(x),0\}^2 w(x)dx \geq \pi_N/2\right\},
\end{aligned}$$

with probability approaching one. By the asymptotic tightness of the empirical process, the first probability converges to zero because $\pi_N \to \infty$. The second probability is zero because under the null hypothesis, the integral is zero for some $k \neq l$.

(*Proof of Claim 1*): Since the empirical process $\sqrt{N}q(x)\{D_k(x) - \bar{D}_k(x)\}$ is asymptotically tight for each $k = 1, \ldots, K$, $P\{\sup_x(q(x)|D_l(x) - \bar{D}_l(x)| + q(x)|D_k(x) - \bar{D}_k(x)|) > c_N\} \to 0$ by the choice of $c_N \to 0$ and $c_N\sqrt{N} \to \infty$. For any $x \in B_{kl}$ so that $D_{kl}(x) = 0$, by the triangular inequality,

$$q(x)\left|\bar{D}_k(x)\right| \leq q(x)\left|D_l(x) - \bar{D}_l(x)\right| + q(x)\left|D_k(x) - \bar{D}_k(x)\right| \leq c_N,$$

with probability approaching one. Thus we deduce that $P\{B_{kl} \subset \hat{B}_{kl}\} \to 1$. Now, for any $x \in \hat{B}_{kl}$, by the triangular inequality,

$$q(x)\left|D_{kl}(x)\right| \leq c_N + q(x)\left|D_l(x) - \bar{D}_l(x)\right| + q(x)\left|D_k(x) - \bar{D}_k(x)\right| \leq 2c_N,$$

with probability approaching one. Therefore, $P\{\hat{B}_{kl} \subset [B_{kl}]^{2c_N}\} \to 1$. The convergence uniform over $P \in \mathcal{P}$ follows from the fact that $\sqrt{N}q(x)\{D_k(x) - \bar{D}_k(x)\}$ is asymptotically tight uniformly over $P \in \mathcal{P}$.

(*Proof of Claim 2*): By Lemma A3(i), we have $|\tilde{a}^*_{kl}(A) - \hat{a}^*_{kl}(A)| = O_{P^*}(\delta_N)$. Now, consider

$$|\hat{a}^*_{kl}(A) - a^*_{kl}(A)| \leq \left|\int_A |\nu^*_{kl}(x,\hat{\theta},\hat{\tau})|^2 w(x)dx - \int_A |\nu^*_{kl}(x,\theta_0,\tau_0)|^2 w(x)dx\right|.$$

By using $a^2 - b^2 = (a+b)(a-b)$ and the Cauchy-Schwartz inequality, the above term is bounded by

$$\begin{aligned}
& \mathbf{E}\left[|\hat{a}^*_{kl}(A) - a^*_{kl}(A)|\,|\mathcal{G}_N\right] \\
\leq\ & \left\{\mathbf{E}\left[\int_A |\nu^*_{kl}(x,\hat{\theta},\hat{\tau}) - \nu^*_{kl}(x,\theta_0,\tau_0)|^2 w(x)dx|\mathcal{G}_N\right]\right\}^{1/2} \\
& \times\left\{\mathbf{E}\left[\int_A |\nu^*_{kl}(x,\hat{\theta},\hat{\tau}) + \nu^*_{kl}(x,\theta_0,\tau_0)|^2 w(x)dx|\mathcal{G}_N\right]\right\}^{1/2} \\
=\ & \left\{\mathbf{E}\left[\int_A |\nu^*_{kl}(x,\hat{\theta},\hat{\tau}) - \nu^*_{kl}(x,\theta_0,\tau_0)|^2 w(x)dx|\mathcal{G}_N\right]\right\}^{1/2} \times O_P(1),
\end{aligned}$$

because both the processes $\nu^*_{kl}(\cdot,\hat{\theta},\hat{\tau})$ and $\nu^*_{kl}(\cdot,\theta_0,\tau_0)$ are asymptotically tight by the weak convergence result of Lemma A3(ii). Using the Fubini's theorem, the last expectation is written as

$$\int_A \mathbf{E}[|\nu^*_{kl}(x,\hat{\theta},\hat{\tau}) - \nu^*_{kl}(x,\theta_0,\tau_0)|^2|\mathcal{G}_N]w(x)dx.$$

Now, by using Theorem 2.14.5 of van der Vaart and Wellner (1996), we deduce that for any $p \geq 2$,

$$\begin{aligned}\left\{\mathbf{E}[|\nu_{kl}^*(x,\hat{\theta},\hat{\tau})-\nu_{kl}^*(x,\theta_0,\tau_0)|^2|\mathcal{G}_N]\right\}^{1/2} &\leq \left\{\mathbf{E}[|\nu_{kl}^*(x,\hat{\theta},\hat{\tau})-\nu_{kl}^*(x,\theta_0,\tau_0)|^p|\mathcal{G}_N]\right\}^{1/p} \\ &\leq C\mathbf{E}[|\nu_{kl}^*(x,\hat{\theta},\hat{\tau})-\nu_{kl}^*(x,\theta_0,\tau_0)||\mathcal{G}_N]+CN^{-1/2+1/p}.\end{aligned}$$

By taking large p, we can make $CN^{-1/2+1/p} = O(\delta_N^\kappa)$. Using Le Cam's poissonization lemma and following the proof of Theorem 2.2 of Giné (1997), we obtain that for $e = \exp(1)$,

$$\begin{aligned}&\mathbf{E}\left(\mathbf{E}[\sup\nolimits_{(\theta,\tau)\in B_{\Theta\times\mathcal{T}}(\delta_N)}|\nu_{kl}^*(x,\theta,\tau)-\nu_{kl}^*(x,\theta_0,\tau_0)||\mathcal{G}_N]\right) \qquad (36)\\ \leq\ &\frac{e}{e-1}\mathbf{E}\left[\sup\nolimits_{(\theta,\tau)\in B_{\Theta\times\mathcal{T}}(\delta_N)}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N(N_i-1)\{V_{x,kl}(W_i,\theta,\tau)-V_{x,kl}(W_i,\theta_0,\tau_0)\}\right|\right]\\ &+\frac{e}{e-1}\mathbf{E}\left[\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N(N_i-1)\right|\times\sup\nolimits_{(\theta,\tau)\in B_{\Theta\times\mathcal{T}}(\delta_N)}\left|\frac{1}{N}\sum_{i=1}^N\{V_{x,kl}(W_i,\theta,\tau)-V_{x,kl}(W_i,\theta_0,\tau_0)\}\right|\right],\end{aligned}$$

where $V_{x,kl}(W_i,\theta,\tau)$ is defined in Assumption 3(iv) and $\{N_i\}$ is i.i.d. Poisson random variables of parameter 1 and independent of $\{W_i\}_{i=1}^N$.

Redefine $\mathcal{H}_{kl}(\delta)$ as in the proof of Lemma A3(ii) with $V_{x,kl}(\cdot,\theta,\tau)$ replaced by $V_{x,kl}(\cdot,\theta,\tau)-V_{x,kl}(\cdot,\theta_0,\tau_0)$. Then by Assumption 2(ii)(A) and Assumption 4, we can take an envelope of $\mathcal{H}_{kl}(\delta_N)$, whose $L_2(P)$-norm is bounded by $C(\delta_N^{\lambda s_2/2}+\delta_N^{s_3})$. Using the maximal inequality similarly as in (34), we can bound the second expectation in (36) by

$$\begin{aligned}&C\int_0^{C\delta_N^{(\lambda s_2/2)\wedge s_3}}\sqrt{\log N(\varepsilon,\mathcal{H}_k(\delta_N),||\cdot||_{P,2})}d\varepsilon\\ \leq\ &C\int_0^{C\delta_N^{(\lambda s_2/2)\wedge s_3}}\varepsilon^{-d\lambda/(2s_2)}d\varepsilon+C\int_0^{C\delta_N^{(\lambda s_2/2)\wedge s_3}}\sqrt{\log N_{[]}(\varepsilon^{1/s_3},B_{\Theta\times\mathcal{T}}(\delta_N),||\cdot||+||\cdot||_\infty)}d\varepsilon\\ =\ &O(\delta_N^{((\lambda s_2/2)\wedge s_3)(1-d\lambda/(2s_2))}+\delta_N^{(\lambda s_2/2)\wedge s_3)(1-d/(2s_3))}).\end{aligned}$$

Since $s_2 \leq s_3$, the last rate becomes $\delta_N^{(s_2/2)(1-d/(2s_2))} = \delta_N^{(s_2/2-d/4)}$ when $s = 1$ and $\delta_N^{s_2-d}$ when $s > 1$. Therefore, it is $O(\delta_N^\kappa)$, where $\kappa = \frac{\lambda}{2}(s_2 - \frac{d\lambda}{2})$. Similarly, we can show that the third expectation in (36) is $O(\delta_N^\kappa)$.

(*Proof of Claim 3*): We bound

$$\left|a_{kl}^*([B_{kl}]^{2c_N})-a_{kl}^*(B_{kl})\right| \leq \left|\int_{[B_{kl}]^{2c_N}\setminus B_{kl}}\max\{\nu_{kl}^*(x;\theta_0,\tau_0),0\}^2w(x)dx\right|.$$

Using the weak convergence result in Lemma A3 and the Berry-Esseen approximation bound in Banach spaces, we observe that for any number $M > 0$,

$$\begin{aligned}
& P\left\{\left|\int_{[B_{kl}]^{2c_N}\setminus B_{kl}} \max\{\nu_{kl}^*(x;\theta_0,\tau_0),0\}^2 w(x)dx\right| > Mc_N\sqrt{-\log c_N}|\mathcal{G}_N\right\} \qquad (37)\\
= \ & P\left\{\left|\int_{[B_{kl}]^{2c_N}\setminus B_{kl}} \max\{\nu_{kl}(x),0\}^2 w(x)dx\right| > Mc_N\sqrt{-\log c_N}\right\} + O_P(N^{-1/2})\\
= \ & P\left\{\left|\int_{[B_{kl}]^{2c_N}\setminus B_{kl}} \max\{\bar{\nu}_{kl}(x;\theta_0,\tau_0),0\}^2 w(x)dx\right| > Mc_N\sqrt{-\log c_N}\right\} + O_P(N^{-1/2}),
\end{aligned}$$

where $\bar{\nu}_{kl}(x;\theta_0,\tau_0) = q(x)\sqrt{N}\bar{D}_{kl}(x;\theta_0,\tau_0)$ and $\bar{D}_{kl}(x;\theta_0,\tau_0)$ is $\bar{D}_{kl}$ with $\hat{\theta}$ and $\hat{\tau}$ replaced by θ_0 and τ_0. The last probability above is bounded by

$$\begin{aligned}
& \mathbf{E}\left[\sup_{x,x;:d_V(x,x')\leq 2c_N}\left|\bar{\nu}_{kl}^2(x;\theta_0,\tau_0)-\bar{\nu}_{kl}^2(x';\theta_0,\tau_0)\right|\right]/(Mc_N\sqrt{-\log c_N})\\
\leq \ & J_N\times\left\{\mathbf{E}[\sup_{x,x;:d_V(x,x')\leq 2c_N}\left|\bar{\nu}_{kl}(x;\theta_0,\tau_0)-\bar{\nu}_{kl}(x';\theta_0,\tau_0)\right|^2]\right\}^{1/2}/(Mc_N\sqrt{-\log c_N}),
\end{aligned}$$

where $J_N = 2\{\mathbf{E}[\sup_{x\in\mathcal{X}}|\bar{\nu}_{kl}(x;\theta_0,\tau_0)|^2]\}^{1/2}$. The functions indexing the process $\bar{\nu}_{kl}(x;\theta_0,\tau_0)$ are uniformly bounded. By Theorem 2.14.5 of van der Vaart and Wellner (1996),

$$\lim_{N\to\infty} J_N \leq \lim_{N\to\infty}\mathbf{E}[\sup_{x\in\mathcal{X}}|\bar{\nu}_{kl}(x;\theta_0,\tau_0)|] + C < C$$

by the maximal inequality applied to the expectation above. Now, using Theorem 2.14.5 of van der Vaart and Wellner (1996) again, we deduce that for any $p \geq 2$,

$$\begin{aligned}
& \left\{\mathbf{E}[\sup_{x,x;:d_V(x,x')\leq 2c_N}\left|\bar{\nu}_{kl}(x;\theta_0,\tau_0)-\bar{\nu}_{kl}(x';\theta_0,\tau_0)\right|^2]\right\}^{1/2}\\
\leq \ & \left\{\mathbf{E}[\sup_{x,x;:d_V(x,x')\leq 2c_N}\left|\bar{\nu}_{kl}(x;\theta_0,\tau_0)-\bar{\nu}_{kl}(x';\theta_0,\tau_0)\right|^p]\right\}^{1/p}\\
\leq \ & C\mathbf{E}[\sup_{x,x;:d_V(x,x')\leq 2c_N}\left|\bar{\nu}_{kl}(x;\theta_0,\tau_0)-\bar{\nu}_{kl}(x';\theta_0,\tau_0)\right|] + CN^{-1/2+1/p}.
\end{aligned}$$

By taking p sufficiently large, we can make $CN^{-1/2+1/p} = O(c_N)$. We consider the leading term in the last line which we can bound by

$$\mathbf{E}[\sup_{x,x;:d_{kl}(x,x')\leq Cc_N^{1/p_{kl}}}\left|\bar{\nu}_{kl}(x;\theta_0,\tau_0)-\bar{\nu}_{kl}(x';\theta_0,\tau_0)\right|]$$

using Assumption 3(iv). By Lemma A2 and the maximal inequality again, we bound the last expectation by

$$\int_0^{Cc_N^{1/p_{kl}}}\sqrt{\log N(\varepsilon,\mathcal{X},Cd_{kl})}d\varepsilon \leq \int_0^{Cc_N}\sqrt{\log N(\varepsilon,\mathcal{X},Cd_{\tilde{V}})}d\varepsilon \leq \int_0^{Cc_N}\sqrt{-\log\varepsilon}d\varepsilon = O(c_N\sqrt{-\log c_N}).$$

By taking large M, we can bound the last probability in (37) arbitrarily small. Hence, we obtain the wanted result. ■

Proof of Theorem 3: From the proof of Theorem 2, $c_{\alpha,N,\infty}^* = c_\alpha + o_P(1)$. However, the test statistic diverges to infinity under the alternative hypothesis as we can see in the proof of Theorem 1. Hence the rejection probability converges to one. ■

Proof of Theorem 4 : Since the test statistics are the same, it suffices to compare the bootstrap critical values as $B \to \infty$. By the construction of the local alternatives, we are under the probability on the boundary. Since this bootstrap test statistic $T_{N,b}^{*LF}$ is recentered, it converges in distribution (conditional on $\mathcal{G}_N$) to the distribution of $\min_{k \neq l} \int_{\mathcal{X}} \max\{\nu_{kl}(x), 0\}^2 w(x)dx$, while the distribution of the bootstrap test statistic $T_{N,b}^*$ converges to that of $\min_{k \neq l} \int_{C_{kl}} \max\{\nu_{kl}(x), 0\}^2 w(x)dx$. Note that

$$\int_{\mathcal{X}} \max\{\nu_{kl}(x), 0\}^2 w(x)dx \geq \int_{C_{kl}} \max\{\nu_{kl}(x), 0\}^2 w(x)dx$$

because $C_{kl} \subset \mathcal{X}$. Hence $c_{\alpha,N,\infty}^{*LF} \geq c_{\alpha,N,\infty}^* + o_P(1)$. This implies that under the local alternatives,

$$\begin{aligned} \lim_{N\to\infty} P_N \left\{ T_N > c_{\alpha,N,\infty}^* \right\} &= P\left\{ \min_{k\neq l} \int_{C_{kl}} \max\left\{ \nu_{kl}(x) + \delta_{kl}(x), 0 \right\}^2 w(x)dx > c_{\alpha,N,\infty}^* \right\} \quad (38) \\ &\geq P\left\{ \min_{k\neq l} \int_{C_{kl}} \max\left\{ \nu_{kl}(x) + \delta_{kl}(x), 0 \right\}^2 w(x)dx > c_{\alpha,N,\infty}^{*LF} \right\}. \end{aligned}$$

Now, by the assumption that

$$\int_{\mathcal{X}} \max\left\{ \nu_{kl}(x), 0 \right\}^2 w(x)dx > \int_{C_{kl}} \max\left\{ \nu_{kl}(x), 0 \right\}^2 w(x)dx \text{ for all } k \neq l,$$

$c_{\alpha,N,\infty}^{*LF} > c_{\alpha,N,\infty}^* + o_P(1)$. Hence the inequality in (38) is strict. ■

References

[1] Abadie, A. (2002), "Bootstrap tests of distributional treatment effects in instrumental variables," *The Journal of the American Statistical Association*, 97, 284-292.

[2] Anderson, G. J. (1996), "Nonparametric tests of stochastic dominance in income distributions," *Econometrica*, 1183-1193.

[3] Andrews, D. W. K. (1994), "Empirical process methods in econometrics," in *The Handbook of Econometrics*, Vol. IV ed. by R. F. Engle and D. L. McFadden, Amsterdam: North-Holland.

[4] Andrews, D. W. K. (1997), "A conditional Kolmogorov test," *Econometrica,* 65, 1097-1128.

[5] Andrews, D. W. K. and P. Guggenberger (2007), "The limit of exact size and a problem with subsampling and with the m out of n bootstrap," Working paper.

[6] Andrews, D. W. K. and G. Soares (2007), "Inference for parameters defined by moment inequalities using generalized moment selection," Working paper.

[7] Arcones, M. A. and E. Giné (1992), "On the bootstrap of M-estimators and other statistical functionals," in *Exploring the Limits of Bootstrap*, ed. by R. LePage and L. Billard, Wiley, 13-47.

[8] Arcones, M. A. and E. Giné (1993), "Limit theorems for U-processes," *Annals of Statistics*, 21, 1494-1542.

[9] Barrett, G. and S. Donald (2003), "Consistent tests for stochastic dominance," *Econometrica*, 71,71-104.

[10] Bennett, C. J. (2007), "New consistent integral-type tests for stochastic dominance," Working paper.

[11] Bentkus, V. and F. Götze (1993), "On smoothness conditions and convergence rates in the CLT in Banach spaces," *Probability Theory and Related Fields*, 96, 137-151.

[12] Bickel, P. J. , A.J. Klaassen, Y. Ritov, and J. A. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*, Springer Verlag, New York.

[13] Birman, M. S. and M. Z. Solomjak (1967), "Piecewise-polynomial approximation of functions of the classes W_p^α," *Mathematics of the USSR-Sbornik* 2, 295-317.

[14] Bugni, F. (2008), "Bootstrap inference in partially identified models," Working paper.

[15] Chen, X., O. Linton, and I. van Keilegom (2003), "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica* 71, 1591-1608.

[16] Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models. *Econometrica,* 75, 1243-1284.

[17] Davidson, R. and J. Y. Duclos (2000), "Statistical inference for stochastic dominance and measurement for the poverty and inequality," *Econometrica*, 68, 1435-1464.

[18] Davidson, R. and J. Y. Duclos (2006), "Testing for restricted stochastic dominance," Working paper.

[19] Fan, Y. and S. Park (2007), "Confidence sets for some partially identified models," Working Paper.

[20] Giné, E. (1997), *Lecture Notes on Some Aspects of the Bootstrap, Ecole de Éte de Calcul de Probabilités de Saint-Flour. Lecture Notes in Mathematics*, 1665, Springer, Berlin.

[21] Giné, E. and J. Zinn (1990), "Bootstrapping general empirical measures," *Annals of Probability*, 18, 851-869.

[22] Giné, E. and J. Zinn (1991), "Gaussian characterization of uniform Donsker classes of functions," *Annals of Probability*, 19, 758-782.

[23] Hall, P. and J. L. Horowitz (1996), "Bootstrap critical values for tests based on generalized-method-of-moments estimators," *Econometrica*, 64, 891-916.

[24] Hansen, P. (2003), "Asymptotic tests of composite hypotheses," Working Paper.

[25] Hansen, P. (2005), "A test for superior predictive ability," *Journal of Buiness, Economics, and Statistics*, 23, 365-380.

[26] Härdle, W. and E. Mammen (1993), "Comparing nonparametric versus parametric regression fits," *Annals of Statistics*, 21, 1926-1947.

[27] Horowitz, J. L. (2001), "The Bootstrap", *Handbook of Econometrics*, ed. by J. J. Heckman and E.E. Leamer. North-Holland, Amsterdam.

[28] Horváth, L., P. Kokoszka, and R. Zikitis (2006), "Testing for stochastic dominance using the weighted McFadden-type statistic," *Journal of Econometrics*, 133, 191-205.

[29] Kaur, A., B.L.S. Prakasa Rao and H. Singh (1994), "Testing for second order stochastic dominance of two distributions," *Econometric Theory*, 10, 849-866.

[30] Kim, J. and D. Pollard (1990), "Cube-root asymptotics," *Annals of Statistics*, 18, 191-219.

[31] Klecan, L., R. McFadden, and D. McFadden (1991), "A robust test for stochastic dominance," Working paper.

[32] Koul, H. L. and S. N. Lahiri (1994), "On bootstrapping M-estimated residual processes in multiple linear regression models," *Journal of Multivariate Analysis*, 49, 255-265.

[33] Linton, O., E. Maasoumi, and Y-J. Whang (2005), "Consistent testing for stochastic dominance under general sampling schemes," *Review of Economic Studies,* 72, 735-765.

[34] McFadden, D. (1989), "Testing for stochastic dominance," in *Studies in the Economics of Uncertainty: In Honor of Josef Hadar*, ed. by T. B. Fomby and T. K. Seo. New York, Berlin, London and Tokyo: Springer.

[35] Moon, R. H. and F. Schorfheide (2006), "Boosting your instruments: estimation with overidentifying inequality moment conditions," Working Paper.

[36] Politis, D. N. and J. P. Romano (1999), *Subsampling*, Springer-Verlag, New York.

[37] Paulauskas, V. I. (1976), "On the rate of convergence in the central limit theorem in certain Banach spaces," *Theory of Probability and Applications*, 21, 754-769.

[38] Sheehy, A. and J. A. Wellner (1992), "Uniform Donsker classes of functions," *Annals of Probability*, 20, 1983-2030.

[39] Tse, Y. K. and X. B. Zhang (2004), "A Monte Carlo investigation of some tests for stochastic dominance," *Journal of Statistical Computation and Simulation*, 74, 361-378.

[40] van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag.

[41] Wellner, J. A. and Y. Zhan (1996), "Bootstrapping Z-estimators," Working Paper.

[42] Whang, Y. J. (2000), "Consistent bootstrap tests of parametric regression functions," *Journal of Econometrics*, 98, 27-46.

Design	N	KS					CM				
				LSW					LSW		
		SUB	BTS	I	II	III	SUB	BTS	I	II	III
	50	.137	.063	.070	.063	.063	.105	.064	.084	.064	.064
1a, $d_1^* = 0$	500	.046	.056	.057	.056	.056	.050	.057	.058	.057	.057
	1000	.052	.049	.049	.049	.049	.046	.060	.061	.060	.060
	50	.122	.055	.061	.055	.055	.082	.071	.080	.072	.071
1b, $d_1^* = 0$	500	.052	.051	.051	.051	.051	.058	.059	.061	.059	.059
	1000	.048	.059	.060	.051	.051	.056	.046	.048	.046	.046
	50	.399	.685	.762	.707	.689	.279	.411	.581	.457	.424
1c, $d_1^* > 0$	500	.981	.983	.983	.983	.983	.975	.981	.983	.982	.982
	1000	.994	.995	.995	.995	.995	.993	.994	.994	.994	.994
	50	.381	.684	.781	.706	.689	.262	.426	.578	.464	.437
1d, $d_1^* > 0$	500	.986	.984	.987	.985	.985	.985	.982	.983	.982	.982
	1000	.995	.994	.994	.994	.994	.995	.993	.993	.993	.993
	50	.397	.646	.738	.670	.653	.265	.371	.528	.408	.387
1e, $d_1^* > 0$	500	.986	.992	.992	.992	.992	.975	.991	.991	.991	.991
	1000	.987	.992	.992	.992	.992	.987	.989	.991	.991	.991

Table 1F. Rejection frequencies for the test of First Order Stochastic Dominance for Design 1. SUB refers to the subsampling method with critical values computed by the automatic method "Mean" described by LMW(2005) for the 5% null rejection probabilities. BT refers to the recentered bootstrap. LSW refers to the recentered bootstrap with set estimation, where the tuning parameter is given by $c_N = c \cdot N^{-1/3}$ with $c = 0.25, 0.50,$ and 0.75 for cases I, II, and III, respectively.

Design	N	KS					CM				
				LSW					LSW		
		SUB	BTS	I	II	III	SUB	BTS	I	II	III
	50	.110	.066	.068	.066	.066	.090	.062	.063	.062	.062
1a, $d_2^* = 0$	500	.055	.055	.055	.055	.055	.049	.057	.057	.057	.057
	1000	.052	.050	.050	.050	.050	.048	.050	.050	.050	.050
	50	.084	.061	.132	.075	.066	.064	.066	.154	.080	.071
1b, $d_2^* = 0$	500	.068	.060	.178	.091	.071	.063	.060	.210	.096	.070
	1000	.073	.050	.177	.084	.058	.060	.051	.219	.086	.063
	50	.261	.336	.846	.525	.368	.159	.200	.850	.380	.212
1c, $d_2^* > 0$	500	.942	.451	.995	.985	.983	.738	.003	.996	.986	.983
	1000	.993	.545	.997	.996	.995	.969	.003	.997	.996	.995
	50	.233	.329	.841	.544	.368	.149	.197	.851	.370	.209
1d, $d_2^* > 0$	500	.958	.423	.994	.987	.985	.710	.004	.994	.987	.984
	1000	.995	.524	.999	.996	.994	.955	.003	.999	.998	.994
	50	.233	.299	.798	.478	.326	.138	.179	.810	.346	.189
1e, $d_2^* > 0$	500	.933	.424	.995	.993	.992	.653	.009	.996	.993	.991
	1000	.984	.484	.997	.995	.992	.934	.001	.998	.997	.992

Table 1S. Rejection frequencies for the test of Second Order Stochastic Dominance for Design 1. SUB refers to the subsampling method with critical values computed by the automatic method "Mean" described by LMW(2005) for the 5% null rejection probabilities. BT refers to the recentered bootstrap. LSW refers to the recentered bootstrap with set estimation, where the tuning parameter is given by $c_N = c \cdot N^{-1/3}$ with $c = 2.0, 3.0$, and 4.0 for cases I, II, and III, respectively.

Design	N	KS SUB	KS BTS	KS LSW I	KS LSW II	KS LSW III	CM SUB	CM BTS	CM LSW I	CM LSW II	CM LSW III
	50	.130	.054	.058	.054	.054	.082	.058	.071	.059	.058
2a, $d_1^* = 0$	500	.049	.055	.055	.055	.055	.054	.057	.057	.057	.057
	1000	.070	.044	.044	.044	.044	.050	.061	.062	.061	.061
	50	.078	.072	.088	.075	.073	.031	.056	.068	.056	.056
2b, $d_1^* = 0$	500	.010	.026	.104	.046	.028	.002	.004	.028	.005	.004
	1000	.020	.018	.170	.076	.034	.007	.001	.031	.009	.002
	50	.300	.453	.787	.626	.574	.272	.718	.918	.872	.831
2c, $d_1^* > 0$	500	.979	1.00	1.00	1.00	1.00	.988	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	50	.257	.173	.555	.427	.377	.184	.016	.396	.223	.159
2d, $d_1^* > 0$	500	.969	.988	1.00	1.00	1.00	.954	.644	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	1.00	.995	1.00	1.00	1.00

Table 2F. Rejection frequencies for the test of First Order Stochastic Dominance for Design 2. SUB refers to the subsampling method with critical values computed by the automatic method "Mean" described by LMW(2005) for the 5% null rejection probabilities. BT refers to the recentered bootstrap. LSW refers to the recentered bootstrap with set estimation, where the tuning parameter is given by $c_N = c \cdot N^{-1/3}$ with $c = 0.25, 0.50,$ and 0.75 for cases I, II, and III, respectively.

Design	N	KS					CM				
				LSW					LSW		
		SUB	BTS	I	II	III	SUB	BTS	I	II	III
	50	.077	.056	.063	.056	.056	.068	.061	.074	.061	.061
2a, $d_2^* = 0$	500	.061	.046	.050	.046	.046	.054	.058	.066	058.	.058
	1000	.063	.065	.065	.065	.065	.057	.064	.064	.064	.064
	50	.063	.078	.119	.078	.078	.037	.077	.135	.077	.077
2b, $d_2^* = 0$	500	.002	.006	.154	.018	.006	.001	.002	.171	.006	.002
	1000	.004	.001	.215	.023	.001	.001	.000	.239	.014	.000
	50	.032	.006	.007	.006	.006	.001	.006	.006	.006	.006
2c, $d_2^* = 0$	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	50	.177	.027	.823	.499	.069	.145	.001	.851	.407	.006
2d, $d_2^* > 0$	500	.936	.332	1.00	1.00	1.00	.857	.000	1.00	1.00	1.00
	1000	1.00	.860	1.00	1.00	1.00	.991	.000	1.00	1.00	1.00

Table 2S. Rejection frequencies for the test of Second Order Stochastic Dominance for Design 2. SUB refers to the subsampling method with critical values computed by the automatic method "Mean" described by LMW(2005) for the 5% null rejection probabilities. BT refers to the recentered bootstrap. LSW refers to the recentered bootstrap with set estimation, where the tuning parameter is given by $c_N = c \cdot N^{-1/3}$ with $c = 2.0, 3.0$, and 4.0 for cases I, II, and III, respectively.

Design	N	KS					CM				
				LSW					LSW		
		SUB	BTS	I	II	III	SUB	BTS	I	II	III
	50	.637	.959	.993	.986	.977	.612	.972	.997	.992	.988
3a, $d_1^* > 0$	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	50	.060	.025	.032	.026	.025	.037	.026	.040	.028	.026
3b, $d_1^* = 0$	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	50	.611	.949	.993	.989	.985	.575	.970	.996	.992	.990
3c, $d_1^* > 0$	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3F. Rejection frequencies for the test of First Order Stochastic Dominance for Design 3. SUB refers to the subsampling method with critical values computed by the automatic method "Mean" described by LMW(2005) for the 5% null rejection probabilities. BT refers to the recentered bootstrap. LSW refers to the recentered bootstrap with set estimation, where the tuning parameter is given by $c_N = c \cdot N^{-1/3}$ with $c = 0.25, 0.50$, and 0.75 for cases I, II, and III, respectively.

Design	N	KS					CM				
				LSW					LSW		
		SUB	BTS	I	II	III	SUB	BTS	I	II	III
	50	.031	.021	.025	.022	.021	.000	.016	.022	.016	.016
3a, $d_2^* = 0$	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	50	.048	.044	.123	.053	.046	.039	.050	.149	.065	.051
3b, $d_2^* = 0$	500	.000	.000	.020	.005	.002	.000	.000	.030	.006	.002
	1000	.000	.000	.011	.001	.000	.000	.000	.018	.002	.000
	50	.546	.934	.942	.940	.938	.458	.920	.939	.926	.924
3c, $d_2^* > 0$	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3S. Table 1F. Rejection frequencies for the test of Second Order Stochastic Dominance for Design 3. SUB refers to the subsampling method with critical values computed by the automatic method "Mean" described by LMW(2005) for the 5% null rejection probabilities. BT refers to the recentered bootstrap. LSW refers to the recentered bootstrap with set estimation, where the tuning parameter is given by $c_N = c \cdot N^{-1/3}$ with $c = 2.0, 3.0,$ and 4.0 for cases I, II, and III, respectively.