HAL
open science

An improved bootstrap test of stochastic dominance

Oliver Linton, Kyungchul Song, Yoon-Jae Whang

To cite this version:

Oliver Linton, Kyungchul Song, Yoon-Jae Whang. An improved bootstrap test of stochastic dominance. Econometrics, 2009, 154 (2), pp.186. 10.1016/j.jeconom.2009.08.002 . hal-00617323

HAL Id: hal-00617323

https://hal.science/hal-00617323v1

Submitted on 27 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

An improved bootstrap test of stochastic dominance

Oliver Linton, Kyungchul Song, Yoon-Jae Whang

PII: S0304-4076(09)00188-2
DOI: 10.1016/j.jeconom.2009.08.002
Reference: ECONOM 3226

To appear in: *Journal of Econometrics*

Received date: 9 December 2008
Revised date: 9 December 2008
Accepted date: 4 August 2009

Please cite this article as: Linton, O., Song, K., Whang, Y.-J., An improved bootstrap test of stochastic dominance. *Journal of Econometrics* (2009), doi:10.1016/j.jeconom.2009.08.002

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

An Improved Bootstrap Test of Stochastic Dominance

Oliver Linton*
Department of Economics, London School of Economics
Houghton Street, London WC2A 2AE, United Kingdom.

Kyungchul Song†
Department of Economics, University of Pennsylvania,
528 McNeil Building, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6297.

Yoon-Jae Whang‡
Department of Economics, Seoul National University, Seoul 151-742, Korea.

July 17, 2009

Abstract

We propose a new method of testing stochastic dominance that improves on existing tests based on the standard bootstrap or subsampling. The method admits prospects involving infinite as well as finite dimensional unknown parameters, so that the variables are allowed to be residuals from nonparametric and semiparametric models. The proposed bootstrap tests have asymptotic sizes that are less than or equal to the nominal level uniformly over probabilities in the null hypothesis under regularity conditions. This paper also characterizes the set of probabilities that the asymptotic size is exactly equal to the nominal level uniformly. 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 mimics the discontinuity of the original test's limiting distribution.

*E-mail address: `o.linton@lse.ac.uk`.
†Email: `kysong@sas.upenn.edu`
‡E-mail address: `whang@snu.ac.kr`.

ACCEPTED MANUSCRIPT

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. 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. Barret and Donald (2003) proposed a consistent bootstrap test, for the special case of independent prospects, and showed that it has an asymptotically exact size on the least favorable points in the null hypothesis. Linton, Massoumi, and Whang (2005) (LMW hereafter) suggested a subsampling method that has asymptotically exact size on the boundary points in the null hypothesis and applies to variables that are mutually dependent and contain unknown finite-dimensional parameters. Recent works in this area also include, e.g., Horváth, Kokoszka, and Zitikis (2006) and Bennett (2007).

This paper has three main contributions. First, we propose a new bootstrap test and show that it is asymptotically similar over a much wider set of probabilities than the bootstrap test of Barret and Donald (2003). For example, in the case of first order stochastic dominance, the bootstrap test has an asymptotic size exactly equal to the nominal level as long as the two distribution functions coincide on a set of a positive measure. This is a significant improvement over the bootstrap test of Barret and Donald (2003) that has an asymptotic size equal to the nominal level only when the two distribution functions coincide almost everywhere. Hence as our simulation studies demonstrate, our bootstrap test has improved power properties.

The basic method of this paper is to introduce estimation of what we call the "contact set". In the case of first order stochastic dominance, the contact set is the subset of the domain on which the two distribution functions coincide. The use of the estimated contact set is similar to the general moment selection procedure in Chernozhukov, Hong ,and Tamer (2007) and Andrews and Soares (2007) except that the object here is infinite dimensional. See Hansen (2005) for a similar suggestion in testing predictive abilities among a finite number of forecasting models and Bugni (2008) for a related bootstrap procedure in partially identified moment inequality models.

Second, the bootstrap test is asymptotically valid uniformly over the probabilities in the null hypothesis under certain regularity conditions. Unlike nonparametric or semiparametric tests that are based on the equality of functions, the convergence of test statistics of

ACCEPTED MANUSCRIPT

Kolmogorov-Smirnov type or Cramér-von Mises type is not uniform over the probabilities under the null hypothesis. The asymptotic non-uniformity arises in many situations, such as moment inequalities (Moon and Schorfheide (2007), Chernozhukov, Hong, and Tamer (2007), Andrews and Guggenberger (2006), Fan and Park (2007)) and constructing confidence sets of the coefficient of the autoregressive processes (Mikusheva (2007)). The non-uniformity raises concerns about the validity of asymptotic approximations in the tests or estimators. Despite the fact that the test statistic has a discontinuous limit behavior, the bootstrap test that this paper proposes is uniformly asymptotically valid under the null hypothesis. The result is related to the suggestion by Andrews and Soares (2007) who proposed a method to construct uniformly valid confidence sets under a model defined by finite moment inequalities.

The third contribution of this paper's proposal is that it accommodates the situation where the object of interest depends on unknown parameters finite dimensional or infinite dimensional. As we explain in the main text, this accommodation is useful when one attempts to "explain" the observed stochastic dominance relationship through covariates. The paper's framework is quite flexible enough to encompass various semiparametric models.

We perform Monte Carlo simulations that compare three methods: the bootstrap based on the least favorable configuration, subsampling method, and the bootstrap proposed by this paper. The results verify the superior performance of our method.

It is also worth noting that in the appendix, we present a uniform continuous mapping theorem and a result on bootstrap uniform Donsker classes. These tools are very general, and can be useful in other contexts of methodological developments.

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 $\{X_0, X_1\}$ be a pair of continuous outcome variables that may, for example, represent different points in time or different regions. Let $F_k(x)$ be the distribution function of the

ACCEPTED MANUSCRIPT

k-th random variable X_k, $k = 0, 1$. Let $D_k^{(1)}(x) \equiv F_k(x)$ and for each subsequent s,

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

Then we say that X_0 *stochastically dominates* X_1 *at order* s if $D_0^{(s)}(x) \leq D_1^{(s)}(x)$ for all x. This is equivalent to an ordering of expected utility over a certain class of utility functions $\mathcal{U}_s$, see LMW.

The main focus of this paper is on whether a designated outcome variable X_0 stochastically dominates X_1. Let $D_{01}^{(s)}(x) \equiv D_0^{(s)}(x) - D_1^{(s)}(x)$. The null hypothesis of main interest takes the following form:

$$H_0 : D_{01}^{(s)}(x) \leq 0 \text{ for all } x \in \mathcal{X}, \text{ vs. } H_1 : D_{01}^{(s)}(x) > 0 \text{ for some } x \in \mathcal{X}, \tag{1}$$

where $\mathcal{X}$ denotes a (possibly unbounded) set contained in the union of the supports of X_1 and X_0. The null hypothesis represents the stochastic dominance of X_0 over X_1. The alternative hypothesis corresponds to no such incidence.

The null hypothesis involves the sign of a function. It is convenient to formulate it in terms of a scalar quantity that represents this. In particular, this paper focuses on a one-sided Cramér-von Mises type functional

$$c_s \equiv \int_{\mathcal{X}} \max\left\{D_{01}^{(s)}(x), 0\right\}^2 w(x)dx,$$

where w is a nonnegative integrable weighting function. The weighting function can be chosen appropriately depending on the interest of the researcher. For instance, suppose that X_0 and X_1 indicate household income for individuals in two different cohorts and the object of main interest is to see if X_0 stochastically dominates X_1 for individuals 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 function instead of $w(x) = 1$ generally increases the power of the test against alternatives such that the stochastic dominance fails for X restricted to below μ. Then the null hypothesis can be checked by looking at whether $c_s = 0$ or $c_s > 0$.

An alternative way of formulating the problem is through a one-sided Kolmogorov-Smirnov type functional:

$$d_s \equiv \sup_{x \in \mathcal{X}} D_{01}^{(s)}(x)w(x),$$

and check whether $d_s \leq 0$ or $d_s > 0$. Obviously when $d_s \leq 0$, it follows that $c_s = 0$. On the contrary, when $c_s = 0$, $D_{01}^{(s)}(x) \leq 0$, $w(x)dx$-a.e. Therefore, if $D_0^{(s)}$ and $D_1^{(s)}$ are continuous

ACCEPTED MANUSCRIPT

functions and $w(x)dx$ is absolutely continuous with respect to the Lebesgue measure, the statements $d_s \leq 0$ and $c_s = 0$ are equivalent. For brevity, we do not pursue a formal development of the one-sided Kolmogorov-Smirnov functional or other functionals in this paper.

This paper's framework allows for the case where X_k depends on unknown parameters. Generically we let $X_k(\theta, \tau)$ be specified as

$$X_k(\theta, \tau) = \varphi_k(W; \theta, \tau), \ k = 0, 1, \tag{2}$$

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 $(\theta, \tau) \in \Theta \times \mathcal{T}$. Here we assume that Θ is contained in a finite-dimensional Euclidean space and $\mathcal{T}$ is an infinite-dimensional space. Hence the specification $X_k(\theta, \tau)$ allows for many semiparametric models. Throughout this paper, we assume that θ and τ are identified and can be consistently estimated, so that we use $X_k(\hat{\theta}, \hat{\tau})$ eventually in the test statistic. 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 $X_k(\theta_0, \tau_0) = Y_k - \tau_0(Z_{1k}^\top \theta_0)$. 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 $\varphi_k(w; \theta, \tau) = y_k - \tau(z_k^\top \theta), \ w = (y, z)$.

2.2 Test Statistics and Asymptotic Theory

In this section, we specify the characteristics of the data generating process that define the scope of this paper. Let $B_{\Theta \times \mathcal{T}}(\delta) \equiv \{(\theta, \tau) \in \Theta \times \mathcal{T} : ||\theta - \theta_0|| + ||\tau - \tau_0||_\infty < \delta\}$. The notations $|| \cdot ||$, $|| \cdot ||_{P,2}$ and $|| \cdot ||_\infty$ denote the Euclidean norm, the $L_2(P)$-norm and the sup norm, respectively. We introduce a bounded weight function $q(x)$ (see Assumption 3(iii)) and define

$$h_x(\varphi) \equiv \frac{(x - \varphi)^{s-1} 1\{\varphi \leq x\} q(x)}{(s-1)!}. \tag{3}$$

The functions $h_x(\varphi)$ play a significant role here as they constitute the index space for the empirical processes later. (See the discussions below (5).) Let $N(\varepsilon, \mathcal{T}, || \cdot ||_\infty)$ denote the ε-covering number of $\mathcal{T}$ with respect to $|| \cdot ||_\infty$, i.e. the smallest number of ε-balls 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) $\log N(\varepsilon, \mathcal{T}, || \cdot ||_\infty) \leq C\varepsilon^{-d}$ for some $d \in (0, 1]$.

ACCEPTED MANUSCRIPT

(iii) For some $\delta > 0$, $\sup_{P\in\mathcal{P}}\mathbf{E}_P[\sup_{(\theta,\tau)\in B_{\Theta\times\mathcal{T}}(\delta)}|X_{ki}(\theta,\tau)|^{2((s-1)\vee 1)+\delta}] < \infty$, $k = 0, 1$.
(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

$$\begin{aligned} & |\mathbf{E}_P\left[h_x(X_{ki}(\theta,\tau))\right] - \mathbf{E}_P\left[h_x(X_{ki}(\theta_0,\tau_0))\right] - \Gamma_{k,P}(x)[\theta-\theta_0, \tau-\tau_0]| \\ \leq\ & C_1||\theta-\theta_0||^2 + C_2||\tau-\tau_0||_\infty^2,\ k = 0, 1, \end{aligned}$$

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

The 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 2.7.1 in van der Vaart and Wellner (1996)), the entropy condition holds. In this case, we can take $d = d_{\mathcal{T}}/\alpha$. Condition (iii) is a moment condition with local uniform boundedness. The moment condition is widely used in the semiparametric literature. In the example of single-index models where $Y_k = \tau_0(Z_{1k}^\top\theta_0) + \varepsilon_k$, we can write $\varphi_k(w;\theta,\tau) = \tau_0(z_{1k}^\top\theta_0) - \tau(z_{1k}^\top\theta) + \varepsilon_k$. If τ is uniformly bounded in the neighborhood of τ_0 in $||\cdot||_\infty$, the moment condition is immediately satisfied when $\sup_{P\in\mathcal{P}}\mathbf{E}_P[|\varepsilon_i|^{2((s-1)\vee 1)+\delta}] < \infty$. We may check this condition for other semiparametric specifications in a similar manner.

Condition (iv) requires pathwise differentiability of the 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 Lipschitz in φ with the coefficient bounded by $C|x-\varphi|^{s-2}q(x)$. Hence, Condition (iv) follows if the moment condition in (iii) is satisfied, and

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

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

$$\Gamma_{k,P}(x)[\theta-\theta_0, \tau-\tau_0] = \mathbf{E}\left[Dh_x(\varphi_k(W;\theta_0,\tau_0))\Gamma_{k,P}^{\varphi}(W)[\theta-\theta_0, \tau-\tau_0]\right] q(x)/(s-1)!$$

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 that $\varphi_k(W;\tau,\theta) = Y_k - \tau(Z_{1k}^\top\theta)$. Define $B(\tau,\delta) = \{\tau \in \mathcal{T} : ||\tau-\tau_0||_\infty \leq \delta\}$ and $B(\theta_0,\delta) = \{\theta \in \Theta : ||\theta-\theta_0|| \leq$

ACCEPTED MANUSCRIPT

$\delta\}$. Assume that each τ is twice continuously differentiable and that

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

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 to be second order continuously differentiable with bounded derivatives. Then, Condition (iv) is satisfied. To see this, first observe that

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

Now, by applying the above bound,

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

Hence, we can take

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

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

Assumption 2 : (i) $X_{0i}(\theta_0, \tau_0)$ and $X_{1i}(\theta_0, \tau_0)$ are continuous random variables.
(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

ACCEPTED MANUSCRIPT

of W_1, and (c) for each $(\theta_1, \tau_1) \in B_{\Theta\times\mathcal{T}}(\delta)$ and for each $\varepsilon > 0$,

$$\sup_{P\in\mathcal{P}} \sup_{w_1} \mathbf{E}_P \left[\sup_{(\theta_2,\tau_2)\in B_{\Theta\times\mathcal{T}}(\varepsilon)} \left|\varphi_k(W;\theta_1,\tau_1) - \varphi_k(W;\theta_2,\tau_2)\right|^2 |W_1 = w_1 \right] \leq C\varepsilon^{2s_2} \tag{4}$$

for some $s_2 \in (d\lambda/2, 1]$ with $\lambda = 2 \times 1\{s = 1\} + 1\{s > 1\}$ 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 (4) requires the function $\varphi_k(W; \theta, \tau)$ 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 $\bar{F}_{kN}(x, \theta, \tau) \equiv \frac{1}{N} \sum_{i=1}^N 1\{X_{ki}(\theta, \tau) \leq x\}$ and

$$\bar{D}_{01}^{(s)}(x, \theta, \tau) \equiv \bar{D}_0^{(s)}(x, \theta, \tau) - \bar{D}_1^{(s)}(x, \theta, \tau) \text{ for } s \geq 1,$$

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

$$\bar{D}_k^{(s)}(x, \theta, \tau) = \int_{-\infty}^x \bar{D}_k^{(s-1)}(t, \theta, \tau) dt \text{ for } s \geq 2. \tag{5}$$

The numerical integration in (5) can be cumbersome in practice. Integrating by parts, we have an alternative form

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

ACCEPTED MANUSCRIPT

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 superscripts (s) from the notations so that we write $\bar{D}_k$ and D_k for $\bar{D}_k^{(s)}$ and $D_k^{(s)}$. Similarly we write D_{01} and $\bar{D}_{01}$ instead of $D_{01}^{(s)}$ and $\bar{D}_{01}^{(s)}$.

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

$$T_N \equiv \int_{\mathcal{X}} \max\left\{q(x)\sqrt{N}\bar{D}_{01}(x), 0\right\}^2 w(x)dx,$$

where $\bar{D}_{01}(x) \equiv \bar{D}_{01}(x, \hat{\theta}, \hat{\tau})$. Regarding $\hat{\theta}$ and $\hat{\tau}$ and the weight function q, we assume the following.

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

$$\sup_{P\in\mathcal{P}} P\left\{\sup_{x\in\mathcal{X}}\left|\sqrt{N}\Gamma_{k,P}(x)[\hat{\theta} - \theta_0, \hat{\tau} - \tau_0] - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi_{x,k}(W_i; \theta_0, \tau_0)\right| > \varepsilon\right\} \to 0, \qquad (6)$$

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

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

(iii) $\sup_{x\in\mathcal{X}}\left(1 + |x|^{(s-1)\vee(1+\delta)}\right)q(x) < \infty$, 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 $\mathcal{X}$.
(iv) There exist a bounded function V on $\mathcal{X}$ and constants $C, \delta > 0$ and $s_1 \in (d/2, 1]$ with d in Assumption 1(ii) such that for each $(x_1, \theta_1, \tau_1) \in \mathcal{X} \times B_{\Theta\times\mathcal{T}}$ and for each $\varepsilon > 0$,

$$\mathbf{E}[\sup_{x\in\mathcal{X}:d_V(x,x_1)\le\varepsilon}\sup_{(\theta,\tau)\in B_{\Theta\times\mathcal{T}}(\delta):||\theta-\theta_1||+||\tau-\tau_1||_\infty\le\varepsilon}|\psi_x^\Delta(W; \theta, \tau) - \psi_{x_1}^\Delta(W; \theta_1, \tau_1)|^2] \le C\varepsilon^{2s_1},$$

where $d_V(x, x') \equiv |V(x) - V(x')|$ and $\psi_x^\Delta(w; \theta, \tau) \equiv \psi_{x,0}(w; \theta, \tau) - \psi_{x,1}(w; \theta, \tau)$.

When $\hat{\theta}$ is obtained from an M-estimation problem, its uniform consistency follows from the uniform convergence of the population objective function that is uniform over $P \in \mathcal{P}$. This latter convergence can be established by using the uniform law of large numbers that are uniform in $P \in \mathcal{P}$ or even using the stronger result of uniform central limit theorem that is uniform in $P \in \mathcal{P}$. (See Giné and Zinn (1991) and Sheehy and Wellner (1992)). The consistency of $\hat{\tau}$ uniform over $P \in \mathcal{P}$ can also be established by controlling the bias and

ACCEPTED MANUSCRIPT

variance part uniformly in P. In order to control the variance part, one may again 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 (6) 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 uniformity in $P \in \mathcal{P}$ is concerned with the $o_P(1)$ term in the asymptotic linear representation. 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 (6), we keep this condition at a high level for the sake of brevity and flexibility.

Condition (iii) is fulfilled by an appropriate choice of a weight function. The use of the weight function is convenient as it enables us to allow $\mathcal{X}$ to be unbounded. Condition (iii) is stronger than that of Horváth, Kokoszka, and Zitikis (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) 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 (iv) is a locally uniform L_2-continuity condition. (See Chen, Linton, and van Keilegom (2003)).

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

$$C(x_1, x_2) \equiv Cov(V_{x_1}(W_i;\theta_0,\tau_0), V_{x_2}(W_i;\theta_0,\tau_0)), \tag{7}$$

where

$$\begin{aligned} V_x(w;\theta,\tau) &\equiv h_x^{\Delta}(w;\theta,\tau) + \psi_x^{\Delta}(w;\theta,\tau) \text{ and} \\ h_x^{\Delta}(w;\theta,\tau) &\equiv h_x(\varphi_0(w;\theta,\tau)) - h_x(\varphi_1(w;\theta,\tau)). \end{aligned} \tag{8}$$

The asymptotic critical values are based on this Gaussian process ν. Define a contact set $B(0) \equiv \{x \in \mathcal{X} : q(x)D_{01}(x) = 0\}$. Let $\mathcal{P}_0$ be the collection of probabilities that satisfy H_0. We also define $\mathcal{P}_{00} \equiv \{P \in \mathcal{P}_0 : \int_{B(0)} w(x)dx > 0\}$.

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

ACCEPTED MANUSCRIPT

$N \to \infty$,

$$T_N \quad \to_d \quad \begin{cases} \int_{B(0)} \max\{\nu(x), 0\}^2 w(x)dx, \text{ if } P \in \mathcal{P}_{00} \\ 0, \text{ if } P \in \mathcal{P}_0 \backslash \mathcal{P}_{00}. \end{cases}$$

For each fixed $P \in \mathcal{P}_{00}$, the limiting distribution $\int_{B(0)} \max\{\nu(x), 0\}^2 w(x)dx$ of T_N is nondegenerate, but for each $P \in \mathcal{P}_0 \backslash \mathcal{P}_{00}$, it is degenerate. We call $\mathcal{P}_{00}$ the set of boundary points and $\mathcal{P}_0 \backslash \mathcal{P}_{00}$ the set of interior points. Hence the limiting distribution of T_N is discontinuous and the test statistic T_N is not locally uniformly regular in $\mathcal{P}_0$ in the sense of Bickel, Klassen, Ritov, and Wellner (1993). 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)). The statement of Theorem 1 is not uniform over $P \in \mathcal{P}_0$. Clearly when we confine our attention to the case $P \in \mathcal{P}_{00}$ such that $\int_{B(0)} w(x)dx > \varepsilon > 0$ for a fixed $\varepsilon > 0$, it is possible to characterize the limit result in Theorem 1 with uniformity. However, we do not need this at this point. 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:

$$\text{limsup}_{N\to\infty} \text{sup}_{P\in\mathcal{P}_0} \mathbf{E}_P \varphi_\alpha \leq \alpha, \text{ and} \tag{9}$$

$$\text{limsup}_{N\to\infty} \text{sup}_{P\in\mathcal{P}_0'} |\mathbf{E}_P \varphi_\alpha - \alpha| = 0. \tag{10}$$

(ii) When a test φ_α satisfies (10), we say that the test is *asymptotically similar on* $\mathcal{P}_0'$.

The Gaussian process ν in Theorem 1 depends on the data generating process in the null hypothesis, precluding the use of first order asymptotic critical values in practice. Barret and Donald (2003) suggested a bootstrap procedure and LMW suggested a subsampling approach. Both studies have not paid attention to the issue of uniformity in the convergence of tests.

One might consider a standard 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.

$$F_0(x) = F_1(x) \text{ for all } x \in \mathcal{X}. \tag{11}$$

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 hypoth-

ACCEPTED MANUSCRIPT

esis, 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 are 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 the following simple example in the finite dimensional case. (For a related result in the context of comparing predictabilities of forecasting models, see Hansen (2005)).

Example 1 : Suppose that the observations $\{X_i = (X_{0i}, X_{1i}) : i = 1, \ldots, N\}$ are mutually independently and identically distributed with unknown mean $\mu = (\mu_0, \mu_1)$ and known variance $\Sigma = diag(1, 1)$. Let the hypotheses of interest be given by:

$$H_0 : \mu_0 \leq 0 \text{ and } \mu_1 \leq 0 \;\text{ vs. }\; H_1 : \mu_0 > 0 \text{ or } \mu_1 > 0. \tag{12}$$

The "boundary" of the null hypothesis is given by $\mathcal{B}_{BD} = \{(\mu_0, \mu_1) : \mu_0 \leq 0 \text{ and } \mu_1 \leq 0\} \cap \{(\mu_0, \mu_1) : \mu_0 = 0 \text{ or } \mu_1 = 0\}$, while the "least favorable case (LFC)" is given by $\mathcal{B}_{LF} = \{(\mu_0, \mu_1) : \mu_0 = 0 \text{ and } \mu_1 = 0\} \subset \mathcal{B}_{BD}$. To test (12), one may consider the following t-statistic:

$$T_N = \max\{N^{1/2}\overline{X}_0, N^{1/2}\overline{X}_1\},$$

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

$$T_N \xrightarrow{d} \begin{cases} \max\{Z_0, Z_1\} & \text{if} \quad \mu = (0,0) \\ Z_0 & \text{if} \quad \mu_0 = 0, \mu_1 < 0 \\ Z_1 & \text{if} \quad \mu_0 < 0, \mu_1 = 0 \end{cases},$$

where Z_0 and Z_1 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 $P(\max\{Z_0, Z_1\} > z^*_\alpha) = \alpha$ 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 $\delta > 0$,

$$H_N : \mu_0 = \frac{\delta}{\sqrt{N}} \text{ and } \mu_1 < 0.$$

ACCEPTED MANUSCRIPT

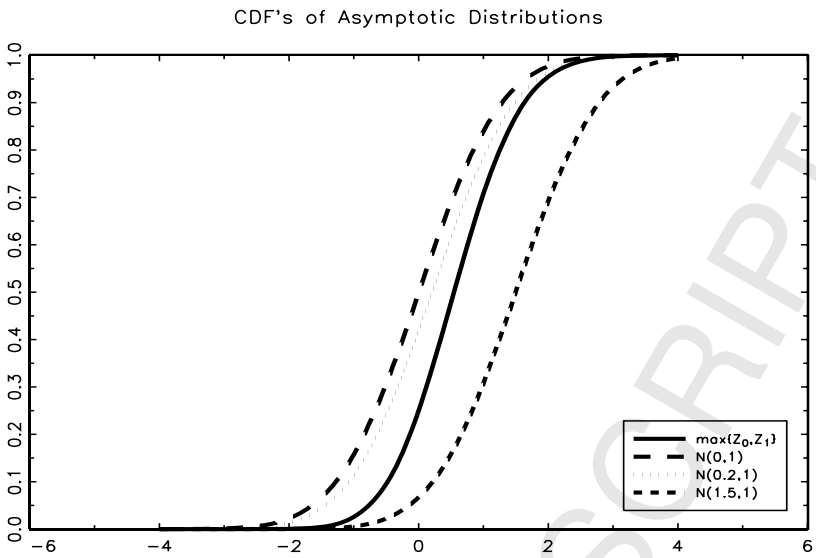

Figure 1: The CDFs of Asymptotic Distributions of the Test Statistic T_N under the Null ($N(0,1)$) and the Alternative Hypotheses ($N(0.2,1)$ and $N(1.5,1)$). The results show that the test is asymptocally biased against H_N for $\delta = 0.2$.

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 these 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 the cdfs of $\max\{Z_0, Z_1\}$ and $N(\delta, 1)$ for $\delta = 0.0$, 0.2, and 1.5. Clearly, the distribution of $\max\{Z_0, Z_1\}$ first-order stochastic dominates that of $N(0.2, 1)$, i.e., T_N is asymptotically 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 $\{W_{i,b}^*\}_{i=1}^N$, $b = 1, \ldots, B$, 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 mimics the distribution of $\hat{\theta}$ and $\hat{\tau}$. For example, when θ_0 and τ_0 are identified through a moment condition:

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

ACCEPTED MANUSCRIPT

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

$$\bar{D}_{01,b}^*(x) \equiv \frac{1}{N}\sum_{i=1}^{N}\left\{h_x(\tilde{X}_{0i,b}^*) - h_x(\tilde{X}_{1i,b}^*) - \frac{1}{N}\sum_{i=1}^{N}\{h_x(\hat{X}_{0i}) - h_x(\hat{X}_{1i})\}\right\},\ b = 1, 2, \ldots, B,$$

where $\hat{X}_{ki} = X_{ki}(\hat{\theta}, \hat{\tau})$. The quantity $\bar{D}_{01,b}^*(x)$ denotes the bootstrap counterpart of $\bar{D}_{01}(x)$. Take a sequence $c_N \to 0$ and $c_N\sqrt{N} \to \infty$, and define

$$\hat{B} \equiv \left\{x \in \mathcal{X} : q(x)|\bar{D}_{01}(x)| < c_N\right\}. \tag{13}$$

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 Zitikis (2006).

We propose the following bootstrap test statistic:

$$T_{N,b}^* \equiv \begin{cases} \int_{\hat{B}} \max\left\{q(x)\sqrt{N}\bar{D}_{01,b}^*(x), 0\right\}^2 w(x)dx, \text{ if } \int_{\hat{B}} w(x)dx > 0 \\ \\ \int_{\mathcal{X}} \max\left\{q(x)\sqrt{N}\bar{D}_{01,b}^*(x), 0\right\}^2 w(x)dx, \text{ if } \int_{\hat{B}} w(x)dx = 0. \end{cases} \tag{14}$$

The bootstrap critical values are obtained by $c_{\alpha,N,B}^* \equiv \inf\{t : B^{-1}\Sigma_{b=1}^{B} 1\{T_{N,b}^* \leq t\} \geq 1 - \alpha\}$, yielding an α-level bootstrap test: $\varphi_\alpha \equiv 1\{T_N > c_{\alpha,N,B}^*\}$.

One of the main steps 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 em-

ACCEPTED MANUSCRIPT

pirical processes was established by Giné and Zinn (1990) who considered a nonparametric bootstrap procedure. See also a nice lecture note on 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 $\{W_i\}_{i=1}^N$.

Assumption 4 : For $\psi_{x,k}(\cdot)$ in Assumption 3(ii), for any $\varepsilon > 0$,

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

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

Assumption 4 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)). This condition 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 Abrevaya and Huang (2005) for a bootstrap inconsistency result for a case where the estimators are not asymptotically linear.

We consider the asymptotic size properties of the bootstrap test based on (14). In particular, we pay attention to the control of asymptotic rejection probabilities uniform in $P \in \mathcal{P}$. We introduce the following regularity conditions for Gaussian processes.

Definition 2 : A Gaussian process ν is *regular on* $A \subset \mathcal{X}$ if for any $\alpha \in (0, 1/2]$, there exists $\bar{\varepsilon} > 0$ depending only on α such that

$$P\left\{\int_A \max\{\nu(x), 0\}^2 w(x)dx < \bar{\varepsilon}\right\} < 1 - \alpha \tag{15}$$

and for any $c > 0$,

$$\text{limsup}_{\eta\downarrow 0}\text{sup}_{P\in\mathcal{P}_0}P\left\{\left|\int_A \max\{\nu(x), 0\}^2 w(x)dx - c\right| \leq \eta\right\} = 0. \tag{16}$$

The regularity condition of the limiting Gaussian process ν is a weak condition. Condition (15) excludes the situation where the distribution of $\int_A \max\{\nu(x), 0\}^2 w(x)dx$ has a large point mass at zero to the extent that its $(1-\alpha)$-th quantile is zero. Condition (16) is a uniform version of the continuity of the distribution function of $\int_A \max\{\nu(x), 0\}^2 w(x)dx$. This condition ensures that all the points $c > 0$ constitute continuity points for the distribution of $\int_A \max\{\nu(x), 0\}^2 w(x)dx$ uniformly in probabilities in $\mathcal{P}_0$.

ACCEPTED MANUSCRIPT

We characterize the set of probabilities under which our bootstrap test has an asymptotically exact size. The characterization involves regularity on r-enlargements of the contact sets, $B(r) \equiv \{x \in \mathcal{X} : q(x)|D_{01}(x)| \leq r\}$. Note that the set $B(r)$ can be viewed as an r-enlargement of the contact set $B(0) = \{x \in \mathcal{X} : q(x)|D_{01}(x)| = 0\}$ with respect to the pseudo norm $||x||_{(1)} = q(x)|D_{01}(x)|$.

Definition 3 : (a) For each $\varepsilon > 0$, let $\mathcal{P}_0(\varepsilon)$ be the collection of probabilities in $\mathcal{P}_0$ under which ν in Theorem 1 is regular on B_N for each $N \geq 1$, where

$$B_N = \begin{cases} B((1-\varepsilon)c_N) & \text{if } \int_{B((1+\varepsilon)c_N)} w(x)dx > 0 \text{ and} \\ \mathcal{X} & \text{if } \int_{B((1+\varepsilon)c_N)} w(x)dx = 0. \end{cases}$$

(b) Given $r_N \to 0$, let $\mathcal{P}_{00}(\varepsilon, \{r_N\})$ be the collection of probabilities in $\mathcal{P}_0(\varepsilon)$ under which for each $N > 1/\varepsilon$, ν in Theorem 1 is regular on $B(N^{-1/2}r_N)$,

$$\int_{B((1-\varepsilon)c_N)} w(x)dx > 0 \text{ and } \int_{B((1+\varepsilon)c_N)\backslash B(N^{-1/2}r_N)} w(x)dx \leq r_N. \tag{17}$$

Definition 3 contains classifications of $\mathcal{P}_0$. The regularity of ν in (a) ensures that the bootstrap critical values are not arbitrarily close to zero. Note that the condition encompasses the case where the set $B((1+\varepsilon)c_N)$ becomes empty. In this case, ν is not regular on $B((1+\varepsilon)c_N)$ and hence the condition requires regularity on $\mathcal{X}$ instead.

The test has asymptotic size equal to α only if $B((1-\varepsilon)c_N)$ is nonempty in the limit. This notion is incorporated by regularity on $B(N^{-1/2}r_N)$ in (b) and the first condition of (17). To see the role of the second condition in (17), let $q(x) = 1$ and $\mathcal{X}$ is bounded. Then, the asymptotic similarity of the test is checked by comparing the distribution of

$$\sqrt{N}\bar{D}_{01}(x) = \sqrt{N}\left[\bar{D}_{01}(x) - D_{01}(x)\right] + \sqrt{N}D_{01}(x)$$

with that of the recentered bootstrap statistic $\sqrt{N}\bar{D}^*_{01}(x)$. Since for all $x \in \mathcal{X}\backslash B((1+\varepsilon)c_N)$, $\sqrt{N}D_{01}(x) \to -\infty$, these x's do not play a role in determining the limiting distribution of the test statistic. Hence we consider only those x's in $B((1+\varepsilon)c_N)$. When $x \in B(N^{-1/2}r_N)$, $\sqrt{N}D_{01}(x) \to 0$ and hence the distribution of $\sqrt{N}\bar{D}_{01}(x)$ is roughly equal to that of $\sqrt{N}\bar{D}^*_{01}(x)$ in the limit. This is not guaranteed, however, when $x \in B((1+\varepsilon)c_N)\backslash B(N^{-1/2}r_N)$. Therefore, the second condition in (17) ensures that the distributions of T_N and $T^*_{N,b}$ coincide in the limit by requiring that the $w(x)dx$-measure of $B((1+\varepsilon)c_N)\backslash B(N^{-1/2}r_N)$ is negligible as $N \to \infty$. This requirement excludes Pitman sequences such that $D_{01}(x) = \delta(x)/\sqrt{N}$, $\delta(x) \in (-\infty, 0)$ from $\mathcal{P}_{00}(\varepsilon, \{r_N\})$. In other words,

ACCEPTED MANUSCRIPT

our test is not asymptotically similar uniformly over these Pitman sequences because the limiting distribution of the test depends on the shift δ.

Theorem 2 : (i) *Suppose that the conditions of* Theorem 1 *and* Assumption 4 *hold and fix any* $\alpha \in (0, 1/2]$. *Then for each* $\varepsilon > 0$,

$$\text{limsup}_{N\to\infty}\text{sup}_{P\in\mathcal{P}_0(\varepsilon)}P\left\{T_N > c^*_{\alpha,N,\infty}\right\} \leq \alpha.$$

(ii) *Furthermore, for each decreasing sequence* $r_N \to 0$ *and* $\varepsilon > 0$,

$$\text{limsup}_{N\to\infty}\text{sup}_{P\in\mathcal{P}_{00}(\varepsilon,\{r_N\})}\left|P\left\{T_N > c^*_{\alpha,N,\infty}\right\} - \alpha\right| = 0.$$

The first result of Theorem 2 says that the bootstrap tests have asymptotically correct sizes uniformly over $P \in \mathcal{P}_0(\varepsilon)$. The second result tells us that the bootstrap tests are asymptotically similar on the subsets of the boundary $\mathcal{P}_{00}(\varepsilon, \{r_N\})$ for any decreasing $r_N \to 0$. 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 $P \in \mathcal{P}_0(\varepsilon)$.

To appreciate the situation where the asymptotic similarity arises, assume that ν is regular on $B(N^{-1/2}r_N)$ as in the definition of $\mathcal{P}_{00}(\varepsilon, \{r_N\})$ and consider the following case: for a fixed function $\delta(x) \in [-\infty, 0]$,

$$D_{01}(x) = \delta(x)/\sqrt{N}. \tag{18}$$

When $\delta(x) = 0$ for all $x \in A$ such that $\int_A w(x)dx > 0$ and $\delta(x) = -\infty$ for all $x \in \mathcal{X}\backslash A$. Then, clearly this case satisfies (17) and hence belongs to $\mathcal{P}_{00}(\varepsilon, \{r_N\})$ for any $r_N \to 0$. When $\delta(x) \in (-\infty, -\varepsilon)$ for all $x \in \mathcal{X}$, the first condition in (17) is still satisfied because $B((1+\varepsilon)c_N) = \mathcal{X}$ from some large N on. However, the second condition in (17) fails because in this case, $B(N^{-1/2}r_N) = \varnothing$ from some large N on and $\int_{B((1+\varepsilon)c_N)\backslash B(N^{-1/2}r_N)} w(x)dx = \int_{\mathcal{X}} w(x)dx > 0$. Therefore when we confine our attention to the Pitman drift types in (18), the conditions for $\mathcal{P}_{00}(\varepsilon, \{r_N\})$ require that the drift should take either 0 values or $-\infty$ at most points of x. However, considering only Pitman drifts is quite restrictive when the object is an infinite-dimensional element. In this regard, the formulation of our asymptotic similarity is rendered so that it can also cover numerous non-Pitman sequences. For example, suppose that for some sequence $A_N \subset \mathcal{X}$ such that $\liminf_{N\to\infty} \int_{A_N} w(x)dx > 0$ and

$$D_{01}(x) = \delta_N(x),$$

with $\delta_N(x) \in [-\infty, 0]$ satisfying that $\inf_{x\in A_N} \delta_N(x)\sqrt{N} \to 0$ and $\sup_{x\in\mathcal{X}\backslash A_N} \delta_N(x) < -(1 +$

ACCEPTED MANUSCRIPT

$\varepsilon)c_N$. Then, the conditions in (17) are satisfied similarly as before and hence this case belongs to $\mathcal{P}_{00}(\varepsilon, \{r_N\})$.

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* $\int_{\mathcal{X}} \max\{q(x)D_{01}(x), 0\}^2 w(x)dx > 0$. *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 Barret 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 $k \in 0, 1$, there exist 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) $\int_{C_{01}} w(x)dx > 0$, where $C_{01} \equiv \{x \in \mathcal{X} : H_0(x) - H_1(x) = 0\}$.
(ii) $\sup_{x\in\mathcal{X}}(H_0(x) - H_1(x)) \leq 0$.
(iii) $\int_{C_{01}} \max\{\delta_0(x) - \delta_1(x), 0\}^2 w(x)dx > 0$.

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 $\int_{B(0)} w(x)dx > 0$, 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

ACCEPTED MANUSCRIPT

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 $\{X^*_{ki,b} : k = 0,1\}_{i=1}^n$, $b = 1,\ldots,B$, this bootstrap procedure alternatively considers the following bootstrap test statistics

$$T^{*LF}_{N,b} \equiv \int_{\mathcal{X}} \max\left\{q(x)\sqrt{N}\bar{D}^*_{01,b}(x), 0\right\}^2 w(x)dx. \tag{20}$$

Let the bootstrap critical values 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.

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 almost everywhere,

$$\int_{\mathcal{X}} \max\left\{\nu(x), 0\right\}^2 w(x)dx > \int_{C_{01}} \max\left\{\nu(x), 0\right\}^2 w(x)dx.$$

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. 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.

ACCEPTED MANUSCRIPT

5 Monte Carlo Experiments

In this section, we examine the finite sample performance of our tests using Monte Carlo simulations. First, we compare the size properties of our method, the subsampling method and the recentered bootstrap method as we alter the data generating process gradually away from the least favorable case. We consider the following generating process for this purpose. Let U_1 and U_2 be $U(0,1)$ random variables. We define

$$\begin{aligned} X_0 &\equiv U_1 \\ X_1 &\equiv c_0^{-1}(U - a_0)1\{0 < U_2 \leq x_0\} + U_2 1\{x_0 < U_2 < 1\}, \end{aligned}$$

where $c_0 = (x_0 - a_0)/x_0 \in (0,1)$ and $x_0 \in (0,1)$. Therefore, the distribution of X_1 is given by

$$P\{X_1 \leq t\} = \begin{cases} 1, & t \geq 1 \\ t, & x_0 < t < 1 \\ c_0 t + a_0, & -a_0/c_0 < t \leq x_0 \\ 0, & t \leq -a_0/c_0 \end{cases},$$

The cdfs, F_0 and F_1, of X_0 and X_1 are shown in Figure 2. The cdf of X_1 has a "kink" at $X_1 = x_0$ and the slope of the cdf changes from c_0 to 1 at the kink point x_0. The location of x_0 determines the size of the contact set $B(0)$ over which the cdf's of X_0 and X_1 are equal. As x_0 gets larger, given $c_0 > 0$, the contact set gets smaller and hence the data generating process moves away from the least favorable case into the interior of the null hypothesis. The value of c_0 determines the behavior of the cdf of Y outside of the contact set. As c_0 gets closer to 1, the cdf of X_1 is closer to the cdf of X_0 outside of the contact set.

In the simulations, we take $x_0 \in \{0, 0.1, 0.2, ..., 0.9\}$ and $c_0 \in \{0.2, 0.4, 0.6, 0.8\}$. The case $x_0 = 0$ corresponds to the least favorable case. The sample size was fixed to be 500. The number of Monte Carlo simulations was set to be 1,000 and the number of bootstrap replications was 400. In constructing the test statistic, we took $q(x) = 1$ and for the bootstrap test statistic, we found that using the estimated contact set with $c_N = cN^{-1/2} \log\log N$, where $c \in \{3.0, 3.2, \ldots, 4.0\}$ show the reasonable size properties comparable to the recentered bootstrap under the least favorable configuration. The rejection probabilities for the c's lay between the two rejection probabilities with $c = 3.0$ and $c = 4.0$. Hence we report only the results corresponding to $c = 3.0$ and $c = 4.0$ for simplicity.

The results are shown in Figure 3. The thick and thin solid lines represent the rejection probabilities of our test, and the dashed and dotted lines, those of the subsampling and the

ACCEPTED MANUSCRIPT

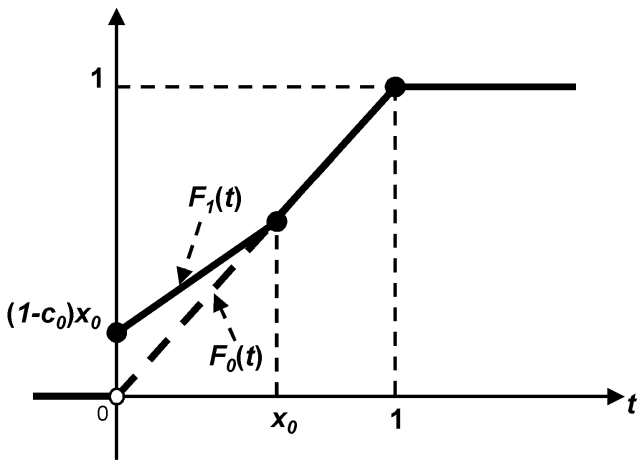

Figure 2: The CDFs, F_0 and F_1, of X_0 and X_1 under the Null Hypothesis

recentered bootstrap respectively. When c_0 is 0.8, the slope of the cdf of X_1 is close to that of X_0. In this case, the data distribution is closer to the least favorable case, and hence the discrepancy between the performances of our method and the recentered bootstrap is not much. Nevertheless, when we increase x_0, moving the data distribution into the interior, our test shows less conservative rejection probabilities than the other methods. This contrast becomes conspicuous in the case where c_0 is 0.2. In this case, as we move into the interior with increasing x_0, our test shows fairly reasonable size properties even until x_0 reaches 0.6 when other tests quickly becomes conservative. Both in the case of $x_0 = 0.4$ and $x_0 = 0.2$, the rejection probability of the subsampling is reasonable as compared to the recentered bootstrap, which is expected from the result of LMW. However, our test still improves on the subsampling method.

Let us turn to the power properties of our method. As for the power properties, we consider the following data generating process. Let $X_1 = U$ where U is a random variable distributed Uniform(0,1). Then, we define

$$X_0 = (U - ab_1)1\{ab_1 \leq U \leq x\} + (U + ab_2)1\{x < U \leq 1 - ab_2\}$$

for $a \in (0, 1)$. As a becomes closer to zero, the distribution of X_0 becomes closer to the uniform distribution. Then the distribution function of X_0 becomes

$$P\{X_0 \leq t\} = t + a\delta(t),$$

ACCEPTED MANUSCRIPT

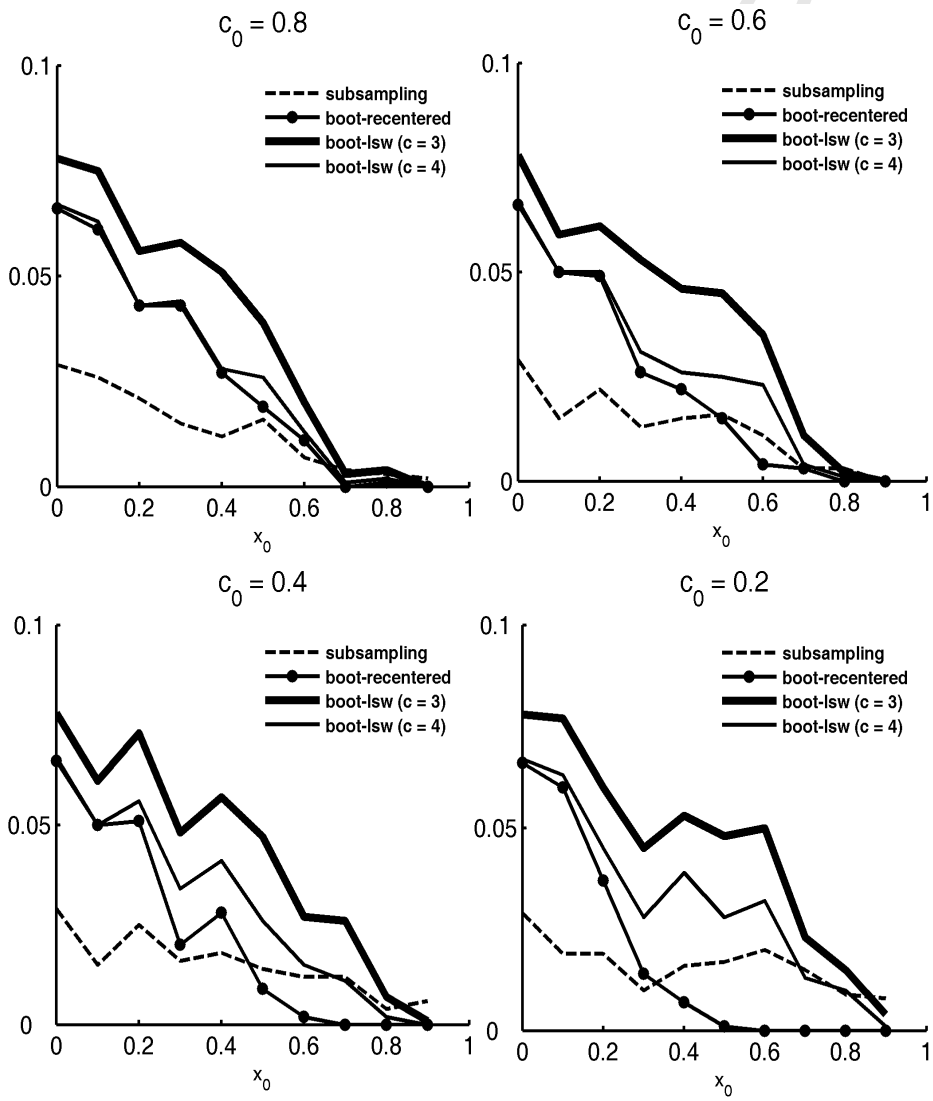

Figure 3: The Rejection Probabilities under the Null Hypothesis: : $\alpha = 0.05$.

ACCEPTED MANUSCRIPT

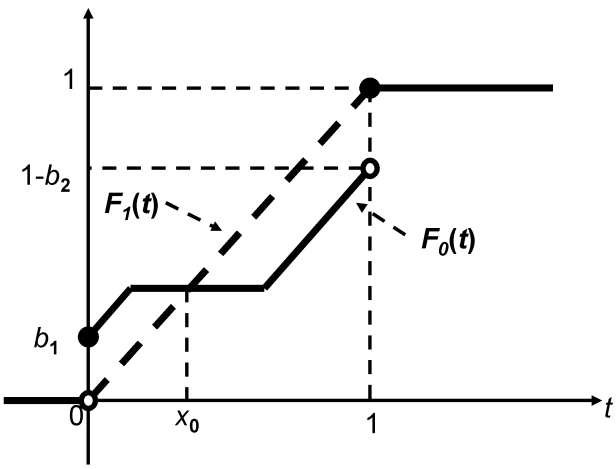

Figure 4: The CDFs of X_0 and X_1 under the Alternatives

where

$$\delta(t) = \begin{cases} 0, & t \leq 0 \\ b_1, & 0 < t \leq x_0 - b_1 \\ x_0, & x_0 - b_1 < t \leq x_0 + b_2 \\ -b_2, & x_0 + b_2 < t \leq 1 \\ 0, & t > 1. \end{cases} \tag{22}$$

The shapes of $F_0(t) = P\{X_0 \leq t\}$ and $F_1(t) = P\{X_1 \leq t\}$ are depicted in Figure 4.

The scale a plays the role of the "distance" of the data distribution from the null hypothesis. When a is large, the data distribution becomes further away from the null hypothesis and when $a = 0$, the data distribution belongs to the null hypothesis under the least favorable configuration. The number x_0 controls how fast the data distribution moves away from the null hypothesis as we increase a.

The results are shown in Figure 5. When $x_0 = 0.3$ and 0.5, our test significantly dominates the recentered bootstrap test which relies on the least favorable configuration. This is anticipated from our theoretical results on local power (Theorem 4). Compared with the subsampling test, our test performs better, in particular when a is greater than 0.6. When $x_0 = 0.7$, the data generating process moves quickly away from the least favorable configuration as a increases. In this case, there is not big difference between our bootstrap and the recentered bootstrap as expected because both tests are consistent. However, even in this case, our test still dominates the other tests. In sum, the simulation results show that

ACCEPTED MANUSCRIPT

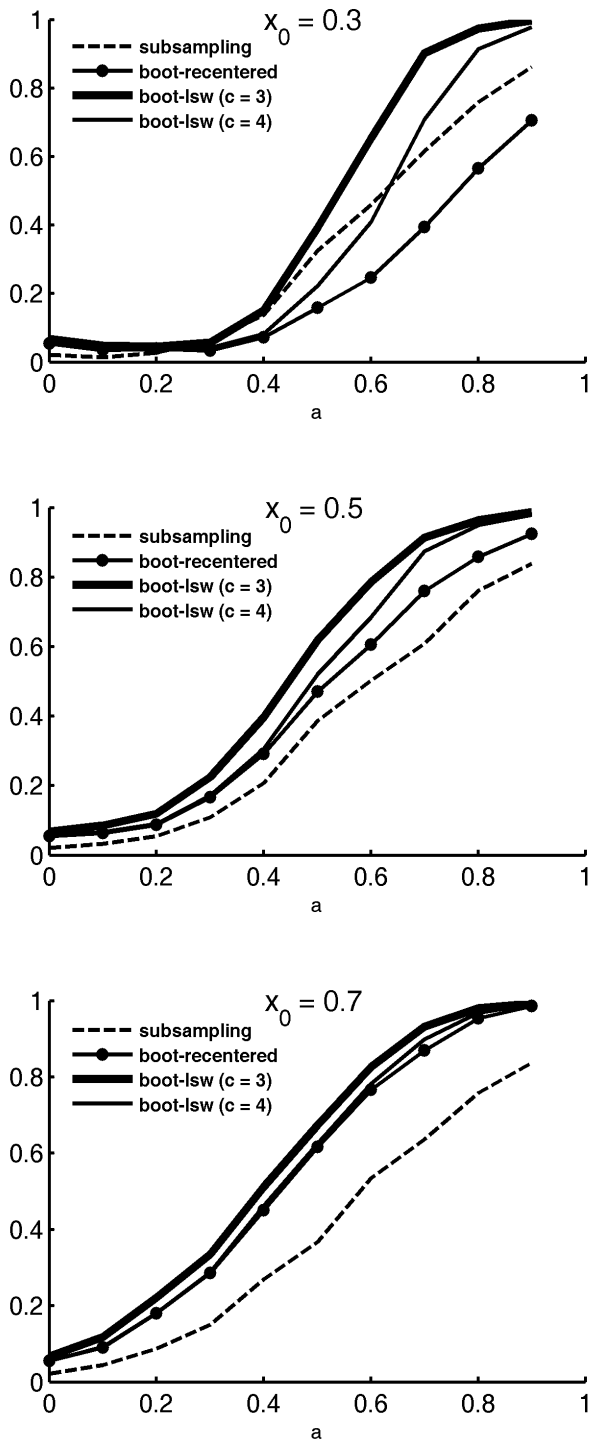

Figure 5: The Rejection Probabilities under the Alternatives: $\alpha = 0.05$.

ACCEPTED MANUSCRIPT

our method overall performs well, dominating the two other existing methods under various designs of alternative hypotheses.

6 Conclusion

This paper proposes a new method of 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 under regularity conditions. 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 some existing methods for quite modest sample sizes.

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

7 Appendix

7.1 Uniform Continuous Mapping Theorem

In this section, we establish a uniform continuous mapping theorem that applies to sequences that are weakly convergent uniformly in P. We introduce some notations. Let $\mathbb{N} = \{1, 2, \cdots\}$ be a set of positive integers and $\mathbf{S} = \mathbf{R}^{d_W}$. Define $\mathcal{P}$ to be a collection of probabilities on the Borel σ-field $\mathcal{S}$ of $\mathbf{S}$. Let $W_i : \mathbf{S}^{\mathbb{N}} \to \mathbf{S}$ be the coordinate functions, and $\{W_i\}_{i=1}^{\infty}$ are i.i.d. with respect to the product measure $P^{\mathbb{N}}$ for every $P \in \mathcal{P}$. Suppose that U_N is a sequence of random variables on the probability space $(\mathbf{S}^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, P^{\mathbb{N}})$. Then, we say that $U_N = O_P(1)$ uniformly in $P \in \mathcal{P}$ if for any $\varepsilon > 0$, there exist $M > 0$ and $N_0 > 0$ such that $\sup_{P\in\mathcal{P}} P\{|U_N| > M\} < \varepsilon$ for all $N > N_0$. Similarly, we say that $U_N = o_P(1)$ uniformly in $P \in \mathcal{P}$ if for any $\varepsilon > 0$, $\sup_{P\in\mathcal{P}} P\{|U_N| > \varepsilon\} \to 0$. Sometimes, we use short-hand notations, $O_{\mathcal{P}}(1)$ and $o_{\mathcal{P}}(1)$.

Let us introduce uniform Donsker classes. The brief exposition here follows Giné and Zinn (1991) who obtained pioneering results regarding this notion. See also Sheehy and Wellner (1992). For a given class of $\mathcal{S}$-measurable functions $\mathcal{F}$ on $\mathbf{S}$, let $l_\infty(\mathcal{F})$ be the collection of bounded maps on $\mathcal{F}$, equipped with the sup norm $||\cdot||_{\mathcal{F}} : ||\nu||_{\mathcal{F}} = \sup_{f\in\mathcal{F}}|\nu(f)|$. The focus of main interest is the following empirical process as an $l_\infty(\mathcal{F})$-valued random element on the probability space $(\mathbf{S}^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, P^{\mathbb{N}})$:

$$\nu_N^P \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\delta_{W_i} - P),$$

where δ_{W_i} is point mass at W_i. For a given $P \in \mathcal{P}$ and a class of functions $\mathcal{F}$ on $\mathbf{S}$, let G_P be a Brownian bridge process on $\mathcal{F}$, a Gaussian process whose covariance function is given by $\mathbf{E}G_P(f)G_P(g) = \mathbf{E}_P(fg) - \mathbf{E}_P f \mathbf{E}_P g$, and let $\mathcal{L}_{\mathcal{F}}(G_P)$ be the distribution of G_P. We say that a class $\mathcal{F}$ is *uniformly pregaussian* if G_P has a version with ρ_P-uniformly continuous sample paths, $\sup_{P\in\mathcal{P}} \mathbf{E}||G_P||_{\mathcal{F}} < \infty$ and $\lim_{\delta\to 0}\sup_{P\in\mathcal{P}}\mathbf{E}||G_P||_{\mathcal{F}(\delta,\rho_P)} = 0$,

ACCEPTED MANUSCRIPT

where $\mathcal{F}(\delta, \rho_P) \equiv \{f - g : f, g \in \mathcal{F}, \rho_P(f, g) \leq \delta\}$ and $\rho_P^2(f, g) \equiv \mathbf{E}_P(f - g)^2 - \{\mathbf{E}_P(f - g)\}^2$. We introduce the following class of functionals on $l_\infty(\mathcal{F})$:

$$BL^{\mathcal{F}} \equiv \left\{H : l_\infty(\mathcal{F}) \to \mathbf{R} : ||H||_\infty \leq 1 \text{ and } \sup_{x,y \in l_\infty(\mathcal{F})} |H(x) - H(y)|/||x - y||_{\mathcal{F}} \leq 1\right\}.$$

Let μ_1 and μ_2 be two probability measures on subsigma fields of Borel subsets of $l_\infty(\mathcal{F})$. Then, using the dual Lipschitz bounded class $BL^{\mathcal{F}}$, we define the following metric

$$d_{BL}(\mu_1, \mu_2) \equiv \sup_{H \in BL^{\mathcal{F}}} \left| \int^* H d\mu_1 - \int^* H d\mu_2 \right|,$$

where $\int^* \cdot d\mu$ indicates the outer integral with respect to μ. Then, we say that a class $\mathcal{F}$ is a *uniform Donsker class* in $P \in \mathcal{P}$ if $\mathcal{F}$ is uniformly pregaussian and

$$\lim_{N \to \infty} \sup_{P \in \mathcal{P}} d_{BL}(\mathcal{L}_{P,\mathcal{F}}(\nu_N^P), \mathcal{L}_{\mathcal{F}}(G_P)) = 0,$$

where $\mathcal{L}_{P,\mathcal{F}}(\nu_N^P)$ denotes the distribution of ν_N^P under $P^{\mathbb{N}}$. In this case, we say that ν_N^P weakly converges to G_P in $l_\infty(\mathcal{F})$ uniformly in $P \in \mathcal{P}$. We conclude this section by presenting a uniform continuous mapping theorem.

Lemma A1 : *Let ν_N^P be an $l_\infty(\mathcal{F})$-valued empirical process that weakly converges to G_P uniformly in $P \in \mathcal{P}$. For each $N = 1, 2, \cdots$ and $P \in \mathcal{P}$, let $f_{N,P}$ be a (possibly stochastic) continuous functional on $l_\infty(\mathcal{F})$ living on $(\mathbf{S}^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, P^{\mathbb{N}})$ such that for some (possibly stochastic) number $C_{N,P} \in \mathbf{R}$,*

$$|f_{N,P}(\nu_1) - f_{N,P}(\nu_2)| \leq C_{N,P}||\nu_1 - \nu_2||_{\mathcal{F}}, \ P\text{-a.e., for each } \nu_1, \nu_2 \in l_\infty(\mathcal{F}). \tag{23}$$

Furthermore, assume that there exists a set $A \subset \mathbf{R}$ such that

$$\lim_{\eta \to 0} \limsup_{N \to \infty} \sup_{P \in \mathcal{P}} P\{|f_{N,P}(G_P) - c| \leq C_{N,P}\eta\} = 0, \ \text{for each } c \in A. \tag{24}$$

Then, for each $c \in A \subset \mathbf{R}$,

$$\limsup_{N \to \infty} \sup_{P \in \mathcal{P}} \left| P\{f_{N,P}(\nu_N^P) \leq c\} - P\{f_{N,P}(G_P) \leq c\} \right| = 0. \tag{25}$$

The lemma formulates a continuous mapping theorem in terms of distribution functions when the mapping depends on N. The approximation is uniform in $P \in \mathcal{P}$. For this uniformity, we use Lipschitz continuity on the continuous mapping. The Lipschitz coefficient $C_{N,P}$ is allowed to change with N and depend on P. Condition (24) is related to the fact that the continuous mapping theorem holds only for P-continuity sets. This condition eliminates points c at which $f_{N,P}(G_P)$ has a point mass.

Proof of Lemma A1 : Throughout the proof, we simply write $f_N = f_{N,P}$, and $C_N = C_{N,P}$. Fix any arbitrarily small number $\eta_1 > 0$ and take m_0 such that for all $m > m_0$,

$$\limsup_{N \to \infty} \sup_{P \in \mathcal{P}} P\{|f_N(G_P) - c| \leq C_N/m\} \leq \eta_1. \tag{26}$$

Condition (24) ensures that we can choose such m_0 independent of P. By the assumed uniform weak convergence of ν_N^P, it is satisfied that for all sequences of bounded Lipschitz functionals $g_N \in BL^{\mathcal{F}}$ possibly

ACCEPTED MANUSCRIPT

depending on P,

$$\begin{aligned}\sup_{P\in\mathcal{P}}\left|\mathbf{E}_P g_N(\nu_N^P)-\mathbf{E}_P g_N(G_P)\right| &= \sup_{P\in\mathcal{P}}\left|\int g_N(\nu_N^P)dP^{\mathbb{N}}-\int g_N(G_P)dP\right| \\ &\leq \sup_{P\in\mathcal{P}}\sup_{H\in BL^{\mathcal{F}}}\left|\int H(\nu_N^P)dP^{\mathbb{N}}-\int H(G_P)dP\right|\to 0.\end{aligned} \tag{27}$$

Take $F\equiv\{r\in\mathbf{R}:r\leq c\}$ and define $F_N\equiv f_N^{-1}(F)$. The set F_N is closed because f_N is continuous. Introduce for $\nu_1\in F_N$,

$$\begin{aligned}l_{m,N}(\nu_1) &\equiv (m\times\inf_{\nu_2\in F_N^c}||\nu_1-\nu_2||_\infty)\wedge 1,\\ u_{m,N}(\nu_1) &\equiv 1-(m\times\inf_{\nu_2\in F_N}||\nu_1-\nu_2||_\infty)\wedge 1.\end{aligned}$$

Then, $l_{m,N}$ and $u_{m,N}$ are bounded and Lipschitz continuous functionals with Lipschitz coefficients bounded by m. Now, observe that

$$\begin{aligned}&\limsup_N\sup_{P\in\mathcal{P}}\left|P(\nu_N^P\in F_N)-P(G_P\in F_N)\right|\\ \leq\ &\limsup_N\sup_{P\in\mathcal{P}}\left|\mathbf{E}_P u_{m,N}(\nu_N^P)-\mathbf{E}_P l_{m,N}(G_P)\right|.\end{aligned}$$

We bound the last supremum by

$$\begin{aligned}&\sup_{P\in\mathcal{P}}\left|\mathbf{E}u_{m,N}(\nu_N^P)-\mathbf{E}u_{m,N}(G_P)\right|+\sup_{P\in\mathcal{P}}\left|\mathbf{E}l_{m,N}(\nu_N^P)-\mathbf{E}l_{m,N}(G_P)\right|\\ &+\sup_{P\in\mathcal{P}}\left|\mathbf{E}u_{m,N}(G_P)-\mathbf{E}l_{m,N}(G_P)\right|.\end{aligned} \tag{28}$$

For every m, the first two terms vanish as $N\to\infty$ by the previous result in (27). As for the last term, note that the absolute value is not zero only when $G_P\in J_{m,N}$ where $J_{m,N}\equiv\{\nu_1\in l_\infty(\mathcal{F}):\inf_{\nu_2\in F_N^c}||\nu_1-\nu_2||_{\mathcal{F}}<1/m$ and $\inf_{\nu_2\in F_N}||\nu_1-\nu_2||_{\mathcal{F}}<1/m\}$. Observe that

$$\begin{aligned}&1\{\inf_{\nu_2\in F_N^c}||\nu_1-\nu_2||_{\mathcal{F}}<1/m\}=1\{\inf_{\nu_2:f_N(\nu_2)>c}||\nu_1-\nu_2||_{\mathcal{F}}<1/m\}\\ \leq\ &1\{\inf_{\nu_2:f_N(\nu_1)>c-C_N||\nu_1-\nu_2||_{\mathcal{F}}}||\nu_1-\nu_2||_{\mathcal{F}}<1/m\}.\end{aligned} \tag{29}$$

The set $\{\nu_2:f_N(\nu_1)>c-C_N||\nu_1-\nu_2||_{\mathcal{F}}\}$ is increasing in $||\nu_1-\nu_2||_{\mathcal{F}}$. If every ν_2 such that $f_N(\nu_1)>c-C_N||\nu_1-\nu_2||_{\mathcal{F}}$ satisfies $||\nu_1-\nu_2||_{\mathcal{F}}\geq 1/m$, $\inf_{\nu_2:f_N(\nu_1)>c-C_N||\nu_1-\nu_2||_{\mathcal{F}}}||\nu_1-\nu_2||_{\mathcal{F}}\geq 1/m$ rendering the last indicator in (29) to be zero. Hence the event in this indicator implies that among ν_2's such that $f_N(\nu_1)>c-C_N||\nu_1-\nu_2||_{\mathcal{F}}$, we can find ν_2 such that $||\nu_1-\nu_2||_{\mathcal{F}}<1/m$. In this case, $f_N(\nu_1)>c-C_N||\nu_1-\nu_2||_{\mathcal{F}}>c-C_N/m$. Hence the last indicator in (29) is bounded by

$$1\{f_N(\nu_1)>c-C_N/m\}.$$

Similarly, $1\{\inf_{\nu_2\in F_N}||\nu_1-\nu_2||_{\mathcal{F}}<1/m\}$ is equal to

$$\begin{aligned}1\{\inf_{\nu_2:f_N(\nu_2)\leq c}||\nu_1-\nu_2||_{\mathcal{F}} &< 1/m\}\leq 1\{\inf_{\nu_2:f_N(\nu_1)\leq c+C_N||\nu_1-\nu_2||_{\mathcal{F}}}||\nu_1-\nu_2||_{\mathcal{F}}<1/m\}\\ &\leq 1\{f_N(\nu_1)\leq c+C_N/m\}.\end{aligned}$$

ACCEPTED MANUSCRIPT

Therefore, the event of $G_P \in J_{m,N}$ is contained in the event that

$$f_N(G_P) - c \leq C_N/m \text{ and } f_N(G_P) - c > -C_N/m$$

or that $|f_N(G_P) - c| \leq C_N/m$. By (26), the $\text{limsup}_{N\to\infty}$ of the probability of the last event is bounded by η_1. As the choice of η_1 was arbitrary, we conclude that $\text{limsup}_{N\to\infty}\sup_{P\in\mathcal{P}} |\mathbf{E}u_{m,N}(G_P) - \mathbf{E}l_{m,N}(G_P)| = 0$. ■

7.2 Bootstrap Uniform Donsker Classes

In this section, we introduce a bootstrap version of uniform Donsker classes for bootstrap empirical processes following Giné and Zinn (1990) and Giné (1997). These papers were originally concerned with a pointwise consistency of the bootstrap empirical measure. However, modifying the results of Giné and Zinn (1990, 1991), we will show that when $\mathcal{F}$ is a uniform Donsker class for $\mathcal{P}$, the bootstrap empirical measure is consistent uniformly in $P \in \mathcal{P}$.

Suppose that $\{W_i\}_{i=1}^{\infty}$ is i.i.d. living on a probability space $(\mathbf{S}^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, P^{\mathbb{N}})$. Given $\{W_i\}_{i=1}^{\infty}$, define the empirical measure:

$$P_N(\omega) \equiv \frac{1}{N}\sum_{i=1}^{N} \delta_{W_i(\omega)},\ \omega \in \mathbf{S}^{\mathbb{N}}.$$

Let $\hat{P}_N(\omega)$ be the empirical measure based on $\{\hat{W}_i(\omega)\}_{i=1}^{N}$ which is an i.i.d. draw from $P_N(\omega)$. Then, the main object of interest is the limit behavior of the bootstrap empirical measure:

$$\nu_N^{\hat{P}}(\omega) \equiv \sqrt{N}\left\{\hat{P}_N(\omega) - P_N(\omega)\right\}.$$

We view $\nu_N^{\hat{P}}(\omega)$ as an $l_\infty(\mathcal{F})$-valued random elements conditional on ω, and let $\mathcal{L}(\nu_n^{\hat{P}}(\omega))$ be the conditional distribution of $\nu_n^{\hat{P}}(\omega)$ given $\omega \in \mathbf{S}^{\mathbb{N}}$. We introduce the metric d_{BL} as before. For a given collection of probabilities $\mathcal{P}$, consider a class of measurable functions $\mathcal{F}$ such that $\sup_{P\in\mathcal{P}} \mathbf{E}_P[|f(W)|^2] < \infty$ and for each $P \in \mathcal{P}$, let G_P be a Brownian bridge process associated with $\mathcal{F}$. We say that $\mathcal{F}$ is *bootstrap uniform Donsker* if $\mathcal{F}$ is uniformly pregaussian and for any $\varepsilon > 0$,

$$\lim_{N\to\infty}\sup_{P\in\mathcal{P}} P\left\{d_{BL}(\mathcal{L}(\nu_n^{\hat{P}}(\omega)), \mathcal{L}(G_P)) > \varepsilon\right\} = 0.$$

This definition is a uniform version of Definition 2.1 of Giné (1997). In the following we show that when $\mathcal{F}$ is a uniform Donsker class, it is bootstrap uniform Donsker. The result follows by slightly modifying the proof of Theorem 2.2. of Giné (1997). For the sake of transparency and future references, we formalize this result and provide the proof here.

Lemma A2 : *If $\mathcal{F}$ is a uniform Donsker class, then $\mathcal{F}$ is also bootstrap uniform Donsker.*

Proof of Lemma A2 : Let $\mathbf{E}^*$ denote the expectation with respect to the bootstrap distribution $\mathcal{L}(\nu_N^{\hat{P}}(\omega))$ conditional on ω. Assume that $\mathcal{F}$ is a uniform Donsker class. By Theorem 2.1 of Sheehy and Wellner (1992), $(\mathcal{F}, \rho_P)$ is totally bounded uniformly in $P \in \mathcal{P}$. Hence given $\varepsilon > 0$, there is a map $\pi_\varepsilon : \mathcal{F} \to \mathcal{F}$ which takes $N(\varepsilon) < \infty$ values and $\rho_P(\pi_\varepsilon f, f) < \varepsilon$ and $N(\varepsilon)$ does not depend on P. Then, for any $H \in BL^{\mathcal{F}}$, we consider:

$$\begin{aligned}\left|\mathbf{E}^* H(\nu_N^{\hat{P}}) - \mathbf{E}H(G_P)\right| \quad \leq \quad & \left|\mathbf{E}^* H(\nu_N^{\hat{P}}) - \mathbf{E}H(\nu_N^{\hat{P}} \circ \pi_\varepsilon)\right| + \left|\mathbf{E}^* H(\nu_N^{\hat{P}} \circ \pi_\varepsilon) - \mathbf{E}H(G_P \circ \pi_\varepsilon)\right| \\ & + \left|\mathbf{E}^* H(G_P \circ \pi_\varepsilon) - \mathbf{E}H(G_P)\right|.\end{aligned} \tag{30}$$

ACCEPTED MANUSCRIPT

The uniform pregaussianity of $\mathcal{F}$ implies that $G_P \circ \pi_\varepsilon$ weakly converges to G_P in $l_\infty(\mathcal{F})$ uniformly in $P \in \mathcal{P}$ as $\varepsilon \to 0$. Therefore, the last term vanishes as $\varepsilon \to 0$ uniformly in $P \in \mathcal{P}$. For each fixed ε, the second term vanishes by the bootstrap CLT in $\mathbf{R}^{N(\varepsilon)}$. The bootstrap CLT in $\mathbf{R}^{N(\varepsilon)}$ uniform in $P \in \mathcal{P}$ can be proved using a finite dimensional CLT uniform in P which is given in Lemma 2.1 of Giné and Zinn (1991). We are left with the first part on the right-hand side of (30). In order to show that it converges to zero uniformly in $P \in \mathcal{P}$, it suffices to prove that $\sup_{P\in\mathcal{P}} P\{\mathbf{E}^*||\nu_N^{\hat{P}}||_{\mathcal{F}(\delta,\rho_P)} > \varepsilon\} \to 0$, as $N \to \infty$ and $\delta \to 0$ for any $\varepsilon > 0$ or that $\mathbf{E}[\mathbf{E}^*||\nu_N^{\hat{P}}||_{\mathcal{F}(\delta,\rho_P)}] \to 0$ as $N \to \infty$ and $\delta \to 0$, where $\mathcal{F}(\delta,\rho_P) = \{f - g \in \mathcal{F} : \rho_P(f,g) \le \delta,\ f,g \in \mathcal{F}\}$. To show the latter convergence, we follow the steps of the proof of Theorem 2.2 of Giné and Zinn (1991) using Le Cam's poissonization lemma to deduce the following:

$$\begin{aligned}
\mathbf{E}[\mathbf{E}^*||\nu_N^{\hat{P}}||_{\mathcal{F}(\delta,\rho_P)}] &= \mathbf{E}\mathbf{E}^*\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\delta_{\hat{W}_i} - P_N\right)\right\|_{\mathcal{F}(\delta,\rho_P)} \le \frac{e}{e-1}\mathbf{E}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N(N_i-1)\left(\delta_{W_i} - P_N\right)\right\|_{\mathcal{F}(\delta,\rho_P)} \\
&\le \frac{e}{e-1}\left\{\mathbf{E}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N(N_i-1)\delta_{W_i}\right\|_{\mathcal{F}(\delta,\rho_P)} + \mathbf{E}\left[\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N(N_i-1)\right| \|P_N\|_{\mathcal{F}(\delta,\rho_P)}\right]\right\},
\end{aligned}$$

where $e = \exp(1)$ and $\{N_i\}_{i=1}^N$ is i.i.d. Poisson random variables with parameter 1 independent of $\{W_i\}_{i=1}^N$. Since $\mathcal{F}$ is a uniform Donsker class, the last two expectations converge to zero uniformly in $P \in \mathcal{P}$ as $N \to \infty$ and $\delta \to 0$. ■

7.3 The Proofs of the Main Results

For any class of functions $\mathcal{F}$ with a pseudo metric d, let $N_{[]}(\varepsilon,\mathcal{F},d)$ denote the bracketing number of $\mathcal{F}$ with respect to d, i.e. the smallest number of ε- brackets that are needed to cover the space $\mathcal{F}$ (e.g. van der Vaart and Wellner (1996)).

Lemma B1 : *For a bounded map $V(x) : \mathcal{X} \to [-M, M]$, $M \in (0,\infty)$, let us introduce a pseudo metric $d_V(x,x') \equiv |V(x) - V(x')|$. Then, for each $\varepsilon \in (0,1]$,*

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

Proof of Lemma B1 : 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 $\mathcal{I}_j$, $j = 1,\ldots,N$, so that the sets $V_j \equiv \{x \in \mathcal{X} : |V(x) - v_j| \le \varepsilon/2\}$, $j = 1,\ldots,N$, 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 B2 : *Let $\mathcal{F} \equiv \{h_x(\varphi_k(\cdot;\theta,\tau)) : (x,\theta,\tau) \in \mathcal{X} \times B_{\Theta\times\mathcal{T}}(\delta)\}$. Then, for each $\varepsilon \in (0,1]$,*

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

Proof of Lemma B2 : Let $\mathcal{H} \equiv \{h_x : x \in \mathcal{X}\}$ and $\gamma_x(y) \equiv 1\{y \le x\}$. Simply write $\varphi = \varphi_k$. Define $\Phi \equiv \{\varphi(\cdot;\theta,\tau) : (\theta,\tau) \in B_{\Theta\times\mathcal{T}}(\delta)\}$. Suppose $s = 1$. Then by the local uniform L_2-continuity condition in Assumption 2(ii)(A)(c), we have

$$\log N_{[]}(\varepsilon,\Phi,||\cdot||_{P,2}) \le \log N(C\varepsilon^{1/s_2}, B_{\Theta\times\mathcal{T}}(\delta), ||\cdot|| + ||\cdot||_\infty) \le C\varepsilon^{-d/s_2}. \tag{31}$$

ACCEPTED MANUSCRIPT

Choose ε-brackets $(\varphi_j, \Delta_{1,j})_{j=1}^{N_1}$ of Φ such that $\int \Delta_{1,j}^2 dP \leq \varepsilon^2$. 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.

We let $V(x) \equiv xq(x)$ and $h(x) \equiv (\log x)^{s_2/d}$. Fix $\varepsilon > 0$. For some $\delta > 0$, $|V(x)x^\delta| < \infty$ as $|x| \to \infty$ by Assumption 3(iii). Hence for some $C > 0$, $V(x) < C/h(x)$ for all $x > 1$. Put $x = h^{-1}(2C/\varepsilon)$ where $h^{-1}(y) \equiv \inf\{x : h(x) \geq y : x \in \mathcal{X}\}$, 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

$$\begin{aligned} &\sup_{\varphi\in\Phi_j}\sup_{x\in\mathcal{X}_m} \left|\gamma_x(\varphi(w))q(x) - \gamma_{x_m}(\varphi_j(w))q(x_m)\right| \\ &\leq \sup_{\varphi\in\Phi_j}\sup_{x\in\mathcal{X}_m} \left|\gamma_x(\varphi(w))q(x) - \gamma_{x_m}(\varphi_j(w))q(x)\right| \\ &+\sup_{\varphi\in\Phi_j}\sup_{x\in\mathcal{X}_m} \left|\gamma_{x_m}(\varphi_j(w))q(x) - \gamma_{x_m}(\varphi_j(w))q(x_m)\right| = \Delta_{m,j}^*(w), \text{ say.} \end{aligned} \tag{32}$$

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

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

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 and $\sup_{x\in\mathcal{X}_0} q(x)^2 \leq \varepsilon^2/4$ as we saw before. We conclude that $||\Delta_{k,j}^*||_{P,2} \leq C\varepsilon^{1/2}$. Hence by taking appropriate constants C, and using (31),

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

Suppose $s > 1$. Then $(x - \varphi)^{s-1}\gamma_x(\varphi)q(x)$ is Lipschitz continuous in φ with the coefficient bounded by

ACCEPTED MANUSCRIPT

$C|x-\varphi|^{s-2}q(x)$. Therefore, using Assumption 1(iii), we deduce that

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

Observe that the constants above do not depend on the choice of the measure P. Combined with (31), we obtain the wanted result. ■

Proof of Theorem 1 : For h_x^Δ defined in (8), we write

$$\begin{aligned} & \sqrt{N}q(x)\bar{D}_{01}(x,\hat{\theta},\hat{\tau}) - \sqrt{N}\mathbf{E}h_x^\Delta(W_i;\theta_0,\tau_0) \\ = \quad & \frac{1}{\sqrt{N}}\sum_{i=1}^N \left\{h_x^\Delta(W_i;\theta_0,\tau_0) - \mathbf{E}h_x^\Delta(W_i;\theta_0,\tau_0)\right\} + \sqrt{N}(\Gamma_{0,P}-\Gamma_{1,P})(x)[\hat{\theta}-\theta_0,\hat{\tau}-\tau_0] + \zeta_{1N} + \zeta_{2N}, \end{aligned}$$

where

$$\begin{aligned} \zeta_{1N} &\equiv \sqrt{N}q(x)\left\{\bar{D}_{01}(x,\hat{\theta},\hat{\tau}) - \bar{D}_{01}(x,\theta_0,\tau_0) - (D_{01}(x,\hat{\theta},\hat{\tau}) - D_{01}(x,\theta_0,\tau_0))\right\} \text{ and} \\ \zeta_{2N} &\equiv \sqrt{N}q(x)(D_{01}(x,\hat{\theta},\hat{\tau}) - D_{01}(x,\theta_0,\tau_0)) - \sqrt{N}(\Gamma_{0,P}-\Gamma_{1,P})(x)[\hat{\theta}-\theta_0,\hat{\tau}-\tau_0]. \end{aligned}$$

By Assumption 3(ii), $\zeta_{2N} = o_{\mathcal{P}}(1)$. We now show that $\zeta_{1N} = o_{\mathcal{P}}(1)$. Fix any decreasing sequence $\delta_N \to 0$ and let $\mathcal{H} \equiv \{h_x^\Delta(\cdot;\theta,\tau) - h_x^\Delta(\cdot;\theta_0,\tau_0) : (x,\theta,\tau) \in \mathcal{X}\times B_{\Theta\times\mathcal{T}}(\delta_N)\}$. The bracketing entropy of this class at $\varepsilon \in (0,1]$ is bounded by $C\varepsilon^{-d\lambda/s_2}$ by Lemma B2. It is easy to show that the $L_2(P)$-norm of its envelope is $O(\delta_N^{\lambda s_2/2}) = o(1)$. Hence by using the maximal inequality and using the fact that $d\lambda/s_2 < 2$, we obtain $\zeta_{1N} = o_{\mathcal{P}}(1)$. Hence using Assumption 3(ii) and noting that $\mathbf{E}h_x^\Delta(W_i;\theta_0,\tau_0) = q(x)D_{01}(x)$,

$$\sqrt{N}q(x)\{\bar{D}_{01}(x,\hat{\theta},\hat{\tau}) - D_{01}(x)\} = \eta_N(x) + o_{\mathcal{P}}(1), \text{ uniformly in } x \in B_{\Theta\times\mathcal{T}}(\delta_N), \tag{33}$$

where

$$\eta_N(x) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^N \left\{(h_x^\Delta + \psi_x^\Delta)(W_i;\theta_0,\tau_0) - \mathbf{E}\left[(h_x^\Delta + \psi_x^\Delta)(W_i;\theta_0,\tau_0)\right]\right\}. \tag{34}$$

We turn to $\eta_N(x)$. Define $\mathcal{H}_0 \equiv \{h_x^\Delta(\cdot) : x \in \mathcal{X}\}$ and $\Psi \equiv \{\psi_x^\Delta(\cdot) : x \in \mathcal{X}\}$. Consider the class of functions $\mathcal{F} \equiv \{h+\psi : (h,\psi) \in \mathcal{H}_0 \times \Psi\}$. Using Assumption 3(iv), Lemma B1 and following the proof of Lemma B2 above, we can show that $\sup_{P\in\mathcal{P}} \log N_{[]}(\varepsilon, \mathcal{F}, L_2(P)) < C\log\varepsilon$. By Theorem 2.3 of Sheehy and Wellner (1991) (See also Theorem 2.8.4 of van der Vaart and Wellner (1996).), $\mathcal{F}$ is a uniform Donsker class. The computation of the covariance kernel of the limiting Gaussian process is straightforward. Therefore, η_N weakly converges to ν uniformly in $P \in \mathcal{P}$. ■

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 $l_\infty(\mathcal{X})$ be the space of real bounded functions on $\mathcal{X}$. Let $V_x(w;\theta,\tau)$ be as defined in (8) and denote

$$\nu^*(x;\theta,\tau) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^N \left\{V_x(W_{i,b}^*;\theta,\tau) - \frac{1}{N}\sum_{i=1}^N V_x(W_i;\theta,\tau)\right\}. \tag{35}$$

Lemma B3 : *Suppose that the assumptions of* Theorem 2 *hold.*
(i) $\sqrt{N}\bar{D}_{01,b}^*(x) = \nu^*(x,\hat{\theta},\hat{\tau}) + o_{P^*}(1)$, *in P uniformly in $P \in \mathcal{P}$.*

ACCEPTED MANUSCRIPT

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

Proof of Lemma B3 : (i) We write

$$\begin{aligned}\sqrt{N}\bar{D}^*_{01,b}(x) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N h_x^\Delta(W_i^*;\hat{\theta}^*,\hat{\tau}^*) - \frac{1}{\sqrt{N}}\sum_{i=1}^N h_x^\Delta(W_i^*;\hat{\theta},\hat{\tau}) \\ &\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^N h_x^\Delta(W_i^*;\hat{\theta},\hat{\tau}) - \frac{1}{\sqrt{N}}\sum_{i=1}^N h_x^\Delta(W_i;\hat{\theta},\hat{\tau}) \\ &= \hat{\Gamma}_{0,P}(x) - \hat{\Gamma}_{1,P}(x) + \frac{1}{\sqrt{N}}\sum_{i=1}^N h_x^\Delta(W_i^*;\hat{\theta},\hat{\tau}) - \frac{1}{\sqrt{N}}\sum_{i=1}^N h_x^\Delta(W_i;\hat{\theta},\hat{\tau}),\end{aligned}$$

where $\hat{\Gamma}_{k,P}(x)$ is as defined in Assumption 4. By Assumption 4, the leading term $\hat{\Gamma}_{0,P}(x) - \hat{\Gamma}_{1,P}(x)$ is written as

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \left\{\psi_x^\Delta(W_i^*;\hat{\theta},\hat{\tau}) - \frac{1}{N}\sum_{i=1}^N \psi_x^\Delta(W_i;\hat{\theta},\hat{\tau})\right\} + o_{P_*}(1). \tag{36}$$

Hence we obtain the wanted result.

(ii) Let $\mathcal{H}(\delta) \equiv \{V_x(\cdot;\theta,\tau) : (x,\theta,\tau) \in \mathcal{X} \times B_{\Theta\times\mathcal{T}}(\delta)\}$ and let H be an envelope of $\mathcal{H}(\delta)$ such that $\sup_{P\in\mathcal{P}} ||H||_{P,2+\varepsilon} < \infty$ for some $\varepsilon > 0$. We later show the following:

$$\int_0^1 \sup_{P\in\mathcal{P}} \sqrt{1 + \log N_{[]}(\varepsilon||H||_{P,2}, \mathcal{H}(\delta), ||\cdot||_{P,2})}\, d\varepsilon ||H||_{P,2} < \infty \tag{37}$$

which implies Pollard's entropy condition. Then by Proposition 3.1 of Giné and Zinn (1991), $\mathcal{H}$ is finitely uniform pregaussian and hence by Corollary 2.7 of Giné and Zinn (1991), it is a uniform Donsker class. The bootstrap uniform CLT for the process $\{\nu^*(x;\theta,\tau) : (x,\theta,\tau) \in \mathcal{X}\times B_{\Theta\times\mathcal{T}}(\delta)\}$ follows from Lemma A2 above. Similarly, by considering $V_x(\cdot;\theta,\tau) - V_x(\cdot;\theta_0,\tau_0)$ in place of $V_x(\cdot;\theta,\tau)$, we can show that

$$\sup_{(x,\theta,\tau)\in\mathcal{X}\times B_{\Theta\times\mathcal{T}}(\delta)} |\nu^*(x;\theta,\tau) - \nu^*(x;\theta_0,\tau_0)| = o_{P^*}(1)$$

in P uniformly in $P \in \mathcal{P}$. The bootstrap uniform CLT applied to $\nu^*(x;\theta_0,\tau_0)$ gives us the wanted result.

Let us show (37). Certainly by Assumptions 1(iii) and 3(ii), we can take H such that $\sup_{P\in\mathcal{P}} ||H||_{P,2} < \infty$. Therefore, in proving (37), we put $||H||_{P,2} = 1$ without loss of generality. We split the proof of (37) into that for $\mathcal{H}_1(\delta) \equiv \{h_x^\Delta(\cdot;\theta,\tau) : (x,\theta,\tau) \in \mathcal{X}\times B_{\Theta\times\mathcal{T}}(\delta)\}$ and that for $\mathcal{H}_2(\delta) \equiv \{\psi_x^\Delta(\cdot;\theta,\tau) : (x,\theta,\tau) \in \mathcal{X}\times B_{\Theta\times\mathcal{T}}(\delta)\}$. As we saw in the proof of Theorem 1 the bracketing entropy of $\mathcal{H}_1(\delta)$ is bounded by $C\varepsilon^{-d\lambda/s_2}$ by Lemma B2. Introduce a pseudo metric $d_V(x,x') \equiv |V(x) - V(x')|$. By the local uniform L_2-continuity condition for the elements in $\mathcal{H}_2(\delta)$ (Assumption 3(iv)), we deduce that

$$\begin{aligned}\sup_{P\in\mathcal{P}} \log N_{[]}(\varepsilon, \mathcal{H}_2(\delta), ||\cdot||_{P,2}) &\le C\sup_{P\in\mathcal{P}} \log N(\varepsilon^{1/s_1}, \mathcal{X}, d_V) \\ &\quad + C\sup_{P\in\mathcal{P}} \log N_{[]}(\varepsilon^{1/s_1}, B_{\Theta\times\mathcal{T}}(\delta), ||\cdot|| + ||\cdot||_\infty) \\ &\le -C\log(\varepsilon) + C\varepsilon^{-d/s_1},\end{aligned} \tag{38}$$

by Lemma B1. By applying Theorem 6 of Andrews (1994) and using the condition that $d/s_1 < 2$ and $d\lambda/s_2 < 2$, we obtain (37). ■

ACCEPTED MANUSCRIPT

Define $\nu^*(x;\theta,\tau)$ as in (35) and let $\nu_0^*(x) \equiv \nu^*(x;\theta_0,\tau_0)$ for short. Define also $\bar{\nu}(x) \equiv q(x)\sqrt{N}\{\bar{D}_{01}(x) - D_{01}(x)\}$. Introduce a functional: for $B_N \subset \mathcal{X}$,

$$f(v;B_N) \equiv \int_{B_N} \max\{v(x),0\}^2 w(x)dx.$$

Let

$$\begin{aligned}
\bar{c}_{\alpha,N}(B_N) &\equiv \inf\{c : P\{f(\bar{\nu};B_N) \le c\} \ge 1-\alpha\} \\
\bar{c}^*_{\alpha,N,\infty}(B_N) &\equiv \inf\{c : P\{f(\nu_0^*;B_N) \le c|\mathcal{G}_N\} \ge 1-\alpha\} \\
c_{\alpha,\infty}(B_N) &\equiv \inf\{c : P\{f(\nu;B_N) \le c\} \ge 1-\alpha\},
\end{aligned}$$

where ν is the Gaussian process in Theorem 1.

Lemma B4 : *Suppose that ν is regular on $B_N \subset \mathcal{X}$ for all $N \ge 1$, under probabilities in $\mathcal{P}' \subset \mathcal{P}$. Then, for each $\alpha \in (0, \frac{1}{2}]$,*

$$\begin{aligned}
\bar{c}_{\alpha,N}(B_N) &= c_{\alpha,\infty}(B_N) + o(1) \\
&= \bar{c}^*_{\alpha,N,\infty}(B_N) + o_{\mathcal{P}'}(1).
\end{aligned} \tag{39}$$

Proof of Lemma B4 : The Gaussian process ν is tight uniformly over $P \in \mathcal{P}$. Hence for any $\eta > 0$, we can take M independent of $P \in \mathcal{P}$ such that $\sup_{P\in\mathcal{P}} P\{\sup_{x\in\mathcal{X}}|\nu(x)| > M\} < \eta$. Since $\bar{\nu}$ weakly converges to ν uniformly in $P \in \mathcal{P}$, for any $\eta > 0$, we can also choose M such that $\text{limsup}_{N\to\infty} \sup_{P\in\mathcal{P}} P\{\sup_{x\in\mathcal{X}}|\bar{\nu}(x)| > M\} < \eta$. Similarly for $\bar{\nu}^*$, applying Lemma B3, for any η, we can also take M such that

$$\text{limsup}_{N\to\infty} \sup_{P\in\mathcal{P}} P\left\{P\{\sup_{x\in\mathcal{X}}|\nu_0^*(x)| > M|\mathcal{G}_N\} < \eta\right\} = 1.$$

Hence throughout the proof, we confine our attention to the event that $\sup_{x\in\mathcal{X}}|\nu(x)| \le M$, $\sup_{x\in\mathcal{X}}|\bar{\nu}(x)| \le M$, and $\sup_{x\in\mathcal{X}}|\nu_0^*(x)| \le M$ conditional on $\mathcal{G}_N$. This means that it suffices to consider $l_\infty^M(\mathcal{X}) = \{v \in l_\infty(\mathcal{X}) : \sup_x |v(x)| \le M\}$ for potential realizations of ν, $\bar{\nu}$ and ν_0^*.

We show only the second equality of (39). The first equality follows similarly by the uniform CLT in the proof of Theorem 1. Write simply $\bar{c}^*_{\alpha,N,\infty}$ and $c_{\alpha,\infty}$, suppressing the argument of B_N. By regularity of ν, we can take $\varepsilon > 0$ such that

$$\sup_{P\in\mathcal{P}'} P\left\{\int_{B_N} \max\{\nu(x),0\}^2 w(x)dx \le \varepsilon\right\} \le 1-\alpha.$$

Hence $c_{\alpha,\infty} \ge \varepsilon$. Fix $\varepsilon_1 \in (0,\varepsilon/2)$ and let E_N be the event that $\left|\bar{c}^*_{\alpha,N,\infty} - c_{\alpha,\infty}\right| > \varepsilon_1$. In this event either $\bar{c}^*_{\alpha,N,\infty} - c_{\alpha,\infty} > \varepsilon_1$ or $c_{\alpha,\infty} - \bar{c}^*_{\alpha,N,\infty} > \varepsilon_1$. Suppose that the first case holds. Then, define $f_1 : l_\infty^M(\mathcal{X}) \to \mathbf{R}$ by

$$f_1(v) \equiv f(v) - c_{\alpha,\infty} - \varepsilon_1.$$

We would like to apply Lemma A1 to $f_1(\nu^*)$ and $f_1(\nu)$. It is easy to check that

$$|f_1(v_1) - f_1(v_2)| \le 2M \times \int_{\mathcal{X}} w(x)dx \times \sup_{x\in\mathcal{X}}|v_1(x) - v_2(x)| \le C\sup_{x\in\mathcal{X}}|v_1(x) - v_2(x)|.$$

ACCEPTED MANUSCRIPT

Furthermore, note that since $c_{\alpha,\infty}+\varepsilon_1>\varepsilon>0$,

$$\begin{aligned}\sup_{P\in\mathcal{P}_0(\varepsilon)}P\left\{|f_1(\nu)|\le C\eta\right\} &= \sup_{P\in\mathcal{P}_0(\varepsilon)}P\left\{\left|\int_{B_N}\max\left\{\nu(x),0\right\}^2 w(x)dx-c_{\alpha,\infty}-\varepsilon_1\right|\le C\eta\right\}\\ &\to 0\end{aligned}$$

as $\eta\to 0$ by Condition (16). Hence we can apply Lemma A1 with $A=\{0\}$ so that for any $\varepsilon>0$,

$$\sup_{P\in\mathcal{P}'}P\left\{\left|P\left\{f_1(\nu^*)\le 0|\mathcal{G}_N\right\}-P\left\{f_1(\nu)\le 0\right\}\right|>\varepsilon\right\}=o(1).$$

Therefore, recalling that $c_{\alpha,\infty}+\varepsilon_1<\bar{c}^*_{\alpha,N,\infty}$,

$$\alpha\ge P\{f(\nu)>c_{\alpha,\infty}\}\ge P\{f(\nu^*)>c_{\alpha,\infty}|\mathcal{G}_N\}+o_{\mathcal{P}'}(1)\ge P\{f(\nu^*)>\bar{c}^*_{\alpha,N,\infty}-\varepsilon_1|\mathcal{G}_N\}+o_{\mathcal{P}'}(1),$$

contradicting the definition of $\bar{c}^*_{\alpha,N,\infty}$.

Suppose that the second case that $c_{\alpha,\infty}-\bar{c}^*_{\alpha,N,\infty}>\varepsilon_1$ holds. Then, we define $f_2: l^M_\infty(\mathcal{X})\to\mathbf{R}$ by

$$f_2(v)\equiv f(v)-c_{\alpha,\infty}+\varepsilon_1.$$

We can check the condition for Lemma A1 similarly as before, using the fact that $c_{\alpha,\infty}-\varepsilon_1>\varepsilon/2>0$. Now, by applying Lemma A1 with $A=\{0\}$,

$$\sup_{P\in\mathcal{P}'}P\left\{\left|P\left\{f_2(\nu^*)\le 0|\mathcal{G}_N\right\}-P\left\{f_2(\nu)\le 0\right\}\right|>\varepsilon\right\}=o(1).$$

Now, since $c_{\alpha,\infty}-\bar{c}^*_{\alpha,N,\infty}>\varepsilon_1$,

$$\alpha\ge P\{f(\nu^*)>\bar{c}^*_{\alpha,N,\infty}|\mathcal{G}_N\}\ge P\{f(\nu^*)>c_{\alpha,\infty}-\varepsilon_1|\mathcal{G}_N\}=P\{f(\nu)>c_{\alpha,\infty}-\varepsilon_1\}+o_{\mathcal{P}'}(1),$$

contradicting the definition of $c_{\alpha,\infty}$. Hence $\sup_{P\in\mathcal{P}'}PE_N\to 0$. Since $\varepsilon_1>0$ was arbitrary, the proof is complete. ■

Suppose that a_N is any sequence of real numbers and $a_{N'}$ with $\{N'\}\subset\mathbb{N}$ is a subsequence. We define

$$\begin{aligned}\limsup_{\{N'\}}a_N &= \lim_{k\to\infty}\sup\{a_N: N\in\{N'\}, N\ge k\} \text{ and}\\ \liminf_{\{N'\}}a_N &= \lim_{k\to\infty}\inf\{a_N: N\in\{N'\}, N\ge k\}.\end{aligned}$$

Lemma B5 : *Let $\{a_{N'_1}\}$ and $\{a_{N'_2}\}$ be two subsequences of $\{a_N\}$,* i.e., $\{N'_1\}\subset\mathbb{N}$ *and* $\{N'_2\}\subset\mathbb{N}$. *Furthermore, assume that for some $a\in\mathbf{R}$,*

$$\limsup_{\{N'_1\}}a_N=\liminf_{\{N'_1\}}a_N=\liminf_{\{N'_2\}}a_N=\limsup_{\{N'_2\}}a_N=a.$$

Then, we have

$$\limsup_{\{N'_1\}\cup\{N'_2\}}a_N=\liminf_{\{N'_1\}\cup\{N'_2\}}a_N.$$

ACCEPTED MANUSCRIPT

Proof of Lemma B5 : Observe that

$$\begin{aligned}
\text{limsup}_{\{N_1'\}\cup\{N_2'\}} a_N &= \lim_{k\to\infty} \max\left(\sup\{a_N : N \in \{N_1'\}, N \geq k\}, \sup\{a_N : N \in \{N_2'\}, N \geq k\}\right) \\
&= \max\left(\lim_{k\to\infty}\sup\{a_N : N \in \{N_1'\}, N \geq k\}, \lim_{k\to\infty}\sup\{a_N : N \in \{N_2'\}, N \geq k\}\right) \\
&= \max\left(\text{limsup}_{\{N_1'\}} a_N, \text{limsup}_{\{N_2'\}} a_N\right) = \text{limsup}_{\{N_1'\}} a_N = \text{liminf}_{\{N_1'\}} a_N \\
&= \min\left(\text{liminf}_{\{N_1'\}} a_N, \text{liminf}_{\{N_2'\}} a_N\right) = \text{liminf}_{\{N_1'\}\cup\{N_2'\}} a_N.
\end{aligned}$$

The second and the last equalities follow because the mappings 'max' and 'min' are continuous. ■

Proof of Theorem 2 : Recall the definition $\bar{\nu}(x) \equiv q(x)\sqrt{N}\{\bar{D}_{01}(x) - D_{01}(x)\}$ and let

$$\bar{\nu}^*(x) \equiv q(x)\sqrt{N}\bar{D}^*_{01,b}(x).$$

We can invoke Lemma B3 and follow the proof of Lemma B4. Hence it suffices to consider $l_\infty^M(\mathcal{X}) = \{v \in l_\infty(\mathcal{X}) : \sup_x |v(x)| \leq M\}$ for potential realizations of ν, $\bar{\nu}$ and $\bar{\nu}^*$. For any sequence of events A_N and collections of probabilities $\mathcal{P}'$, we say that "A_N arises $wp \to_{\mathcal{P}'} 1$" if $\inf_{P\in\mathcal{P}'} P\{A_N\} \to 1$.

Recall the definition: $\nu_0^*(x) \equiv \nu^*(x;\theta_0,\tau_0)$. For any set $A \in \mathcal{X}$, let us denote

$$\begin{aligned}
\tilde{a}^*(A) &\equiv \int_A \max\{\bar{\nu}^*(x), 0\}^2 w(x)dx, \\
\hat{a}^*(A) &\equiv \int_A \max\{\nu^*(x;\hat{\theta},\hat{\tau}), 0\}^2 w(x)dx, \\
a^*(A) &\equiv \int_A \max\{\nu_0^*(x), 0\}^2 w(x)dx, \text{ and} \\
a(A) &\equiv \int_A \max\{\nu(x), 0\}^2 w(x)dx.
\end{aligned}$$

We will show the following at the end of the proof:

Claim 1: $P\{B((1-\varepsilon)c_N) \subset \hat{B} \subset B((1+\varepsilon)c_N)\} \to 1$, uniformly in $P \in \mathcal{P}$.
Claim 2: For any sequence of subsets $B_N \subset \mathcal{X}$, $|\tilde{a}^*(B_N) - a^*(B_N)| = o_{P^*}(1)$ in P, uniformly in $P \in \mathcal{P}$.

Define $\pi_N^* \equiv \int_{\mathcal{X}} \max\{q(x)\sqrt{N}\bar{D}^*_{01,b}(x), 0\}^2 w(x)dx$ and write $T^*_{N,b} = \tilde{a}^*(\hat{B})\hat{1}_N + \pi_N^*(1-\hat{1}_N)$, where $1_N \equiv \{\int_{B((1-\varepsilon)c_N)} w(x)dx > 0\}$, $\bar{1}_N \equiv \{\int_{B((1+\varepsilon)c_N)} w(x)dx > 0\}$, and $\hat{1}_N \equiv \{\int_{\hat{B}} w(x)dx > 0\}$. Let

$$\begin{aligned}
T^{-(1)}_{N,b} &\equiv \tilde{a}^*(B((1-\varepsilon)c_N))1_N + \pi_N^*(1-1_N), \\
T^{-(2)}_{N,b} &\equiv \tilde{a}^*(B((1-\varepsilon)c_N))\bar{1}_N + \pi_N^*(1-\bar{1}_N), \\
T^{+(1)}_{N,b} &\equiv \tilde{a}^*(B((1+\varepsilon)c_N))1_N + \pi_N^*(1-1_N), \text{ and} \\
T^{+(2)}_{N,b} &\equiv \tilde{a}^*(B((1+\varepsilon)c_N))\bar{1}_N + \pi_N^*(1-\bar{1}_N).
\end{aligned}$$

By Claim 1, $wp \to_{\mathcal{P}} 1$, the conditional probability given $\mathcal{G}_N$ of

$$\min\{T^{-(1)}_{N,b}, T^{-(2)}_{N,b}\} \leq T^*_{N,b} \leq \max\{T^{+(1)}_{N,b}, T^{+(2)}_{N,b}\} \tag{40}$$

is equal to one.

ACCEPTED MANUSCRIPT

Now, let us turn to the proof of the theorem. We simply write $\mathcal{P}_{00}(\varepsilon) = \mathcal{P}_{00}(\varepsilon, \{r_N\})$. We show that for each $P \in \mathcal{P}_0(\varepsilon)$,

$$P\left\{T_N > c^*_{\alpha,N,\infty}\right\} \leq \alpha + \varepsilon_{1N}(P), \tag{41}$$

where $\sup_{P\in\mathcal{P}_0(\varepsilon)} \varepsilon_{1N}(P) = o(1)$, and that for each $P \in \mathcal{P}_{00}(\varepsilon)$,

$$\left|P\left\{T_N > c^*_{\alpha,N,\infty}\right\} - \alpha\right| = \varepsilon_{2N}(P) \tag{42}$$

where $\sup_{P\in\mathcal{P}_{00}(\varepsilon)} \varepsilon_{2N}(P) = o(1)$.

(i) : Given each $P \in \mathcal{P}_0(\varepsilon)$, and for $\varepsilon_{1N}(P)$ in (41), for any subsequence $\{\varepsilon_{1N'}\}$ of $\{\varepsilon_{1N}\}$, we can always divide the subsequence into further subsequences of two types $\{\varepsilon_{1N_1''}\}$ and $\{\varepsilon_{1N_2''}\}$ such that $\int_{B((1+\varepsilon)c_{N_1''})} w(x)dx = 0$ (Type 1) and $\int_{B((1+\varepsilon)c_{N_2''})} w(x)dx > 0$ (Type 2). Hence, by Lemma B5, it suffices to assume that without loss of generality, we have infinitely many such N_1'''s and N_2'''s, and show (41) with $\sup_{P\in\mathcal{P}_0(\varepsilon)} \varepsilon_{1N_j''}(P) = o(1)$ for both types of subsequences, $j = 1, 2$.

(Type 1 Subsequence) : Select N's such that $\int_{B((1+\varepsilon)c_N)} w(x)dx = 0$ and confine our attention only to such sequences. Note that $0 = \bar{1}_N \geq 1_N$. Hence $\pi^*_N \leq T^*_{N,b}$, $wp \to_{\mathcal{P}_0(\varepsilon)} 1$. This implies that

$$c^*_{\alpha,N,\infty} \geq \tilde{c}^*_{\alpha,N,\infty},\ wp \to_{\mathcal{P}_0(\varepsilon)} 1, \tag{43}$$

where $\tilde{c}^*_{\alpha,N,\infty}$ denotes the $(1-\alpha)$-th quantile of the conditional distribution of π^*_N given $\mathcal{G}_N$.

Now, observe that

$$\begin{aligned} T_N &= \int_{\mathcal{X}} \max\{\sqrt{N}q(x)\bar{D}_{01}(x), 0\}^2 w(x)dx \\ &= \int_{\mathcal{X}\setminus B((1-\varepsilon)c_N)} \max\{\sqrt{N}q(x)(\bar{D}_{01}(x) - D_{01}(x) + D_{01}(x)), 0\}^2 w(x)dx \\ &\quad + \int_{B((1-\varepsilon)c_N)} \max\{\sqrt{N}q(x)(\bar{D}_{01}(x) - D_{01}(x) + D_{01}(x)), 0\}^2 w(x)dx. \end{aligned} \tag{44}$$

As for the first term, observe that for any sequence $r'_N \to 0$,

$$\begin{aligned} & P\left\{\int_{\mathcal{X}\setminus B((1-\varepsilon)c_N)} \max\{\sqrt{N}q(x)(\bar{D}_{01}(x) - D_{01}(x) + D_{01}(x)), 0\}^2 w(x)dx > r'_N\right\} \\ \leq\ & P\left\{\int_{\mathcal{X}\setminus B((1-\varepsilon)c_N)} \max\{\sqrt{N}q(x)(\bar{D}_{01}(x) - D_{01}(x)) - \sqrt{N}(1-\varepsilon)c_N, 0\}^2 w(x)dx > r'_N\right\} \\ \leq\ & P\left\{\sup_{x\in\mathcal{X}} \sqrt{N}q(x)(\bar{D}_{01}(x) - D_{01}(x)) > \sqrt{N}(1-\varepsilon)c_N\right\} \to 0. \end{aligned}$$

The convergence to 0 in the last term is uniform in $P \in \mathcal{P}$ because $\sqrt{N}q(x)(\bar{D}_{01}(x) - D_{01}(x))$ is asymptotically tight uniformly over $P \in \mathcal{P}$ as shown in the proof of Theorem 1. The first inequality follows because when $x \in \mathcal{X}\setminus B((1-\varepsilon)c_N)$, $q(x)|D_{01}(x)| > (1-\varepsilon)c_N$. Since we are under the null hypothesis, $q(x)D_{01}(x) < -(1-\varepsilon)c_N$, which yields the first inequality in the above display. Therefore, for any sequence $r'_N \to 0$,

$$\int_{\mathcal{X}} \max\{\sqrt{N}q(x)\bar{D}_{01}(x), 0\}^2 w(x)dx = \int_{B((1-\varepsilon)c_N)} \max\{\sqrt{N}q(x)\bar{D}_{01}(x), 0\}^2 w(x)dx + O_P(r'_N)$$

ACCEPTED MANUSCRIPT

where $O_P(r'_N)$ is uniform over $P \in \mathcal{P}$. Since $D_{01}(x) \leq 0$, the last integral is bounded by

$$\begin{aligned} & \int_{B((1-\varepsilon)c_N)} \max\{\sqrt{N}q(x)\{\bar{D}_{01}(x) - D_{01}(x)\}, 0\}^2 w(x)dx \\ \leq \; & \sup_{x\in\mathcal{X}} \left|\sqrt{N}q(x)\{\bar{D}_{01}(x) - D_{01}(x)\}\right| \times \int_{B((1-\varepsilon)c_N)} w(x)dx = o_{\mathcal{P}}(1), \end{aligned} \tag{45}$$

because $\int_{B((1-\varepsilon)c_N)} w(x)dx \leq \int_{B((1+\varepsilon)c_N)} w(x)dx = 0$. We conclude that $T_N = o_{\mathcal{P}}(1)$.

On the other hand, let $\tilde{c}_\alpha$ be the $(1-\alpha)$-th quantile of $\int_{\mathcal{X}} \max\{\nu(x), 0\}^2 w(x)dx$. Since ν is regular on $\mathcal{X}$, we can apply Lemma B4 to deduce that

$$\tilde{c}^*_{\alpha,N,\infty} = \tilde{c}_\alpha + o_{\mathcal{P}_0(\varepsilon)}(1) \text{ and } \tilde{c}_\alpha \geq \eta, \tag{46}$$

for some small $\eta > 0$ that does not depend on P. Hence by (43),

$$\begin{aligned} P\left\{T_N > c^*_{\alpha,N,\infty}\right\} &\leq P\left\{T_N > \tilde{c}^*_{\alpha,N,\infty}\right\} + o(1) \\ &= P\left\{T_N > \tilde{c}_\alpha\right\} + o(1). \end{aligned}$$

The last probability converges to zero because $T_N = o_{\mathcal{P}}(1)$ from (45) and because $\tilde{c}_\alpha \geq \eta > 0$ from (46). This establishes (41) for Type 1 subsequences.

(Type 2 Subsequence) : Select N's such that $\int_{B((1+\varepsilon)c_N)} w(x)dx > 0$ and confine our attention only to such sequences. In this case, one can divide the subsequence into further subsequences of two types, either $\int_{B((1-\varepsilon)c_N)} w(x)dx > 0$ or $\int_{B((1-\varepsilon)c_N)} w(x)dx = 0$. Consider the first case. Then, $wp \to_{\mathcal{P}_0(\varepsilon)} 1$, $a^*(B((1-\varepsilon)c_N)) \leq T^*_{N,b}$. Let $\bar{c}^*_{\alpha,N,\infty}$ and $c_{\alpha,\infty}$ be as defined prior to Lemma B4 with $B_N = B((1-\varepsilon)c_N)$. Since ν is regular on $B((1-\varepsilon)c_N))$, we apply Lemma B4 to deduce that

$$c^*_{\alpha,N,\infty} \geq \bar{c}^*_{\alpha,N,\infty} + o_{\mathcal{P}_0(\varepsilon)}(1) = c_{\alpha,\infty} + o_{\mathcal{P}_0(\varepsilon)}(1). \tag{47}$$

On the other hand, using (44) and following the steps thereafter, we obtain that for any sequence $r'_N \to 0$,

$$\begin{aligned} T_N &= \int_{B((1-\varepsilon)c_N)} \max\left\{\sqrt{N}q(x)(\bar{D}_{01} - D_{01})(x) + \sqrt{N}q(x)D_{01}(x), 0\right\}^2 w(x)dx + O_{\mathcal{P}}(r'_N) \\ &\leq \int_{B((1-\varepsilon)c_N)} \max\left\{\sqrt{N}q(x)(\bar{D}_{01} - D_{01})(x), 0\right\}^2 w(x)dx + O_{\mathcal{P}}(r'_N). \end{aligned}$$

By Lemma B4 and (47),

$$\begin{aligned} P\left\{T_N > c^*_{\alpha,N,\infty}\right\} &\leq P\left\{T_N > c_{\alpha,\infty} + o_{\mathcal{P}_0(\varepsilon)}(1)\right\} \\ &\leq P\left\{\int_{B((1-\varepsilon)c_N)} \max\left\{\nu(x), 0\right\}^2 w(x)dx > c_{\alpha,\infty} + o_{\mathcal{P}_0(\varepsilon)}(1)\right\} + o(1) \\ &\leq \alpha + o(1). \end{aligned}$$

Now, suppose that $\int_{B((1-\varepsilon)c_N)} w(x)dx = 0$. In this case, the Gaussian process ν is not regular on $B((1-\varepsilon)c_N)$. Hence this case is excluded by the definition of $\mathcal{P}_0(\varepsilon)$.

(ii) : Since $\int_{B((1-\varepsilon)c_N)} w(x)dx > 0$ for all $N > 1/\varepsilon$, we have $\int_{B((1+\varepsilon)c_N)} w(x)dx > 0$ for all $N > 1/\varepsilon$, implying

ACCEPTED MANUSCRIPT

that $1_N = \bar{1}_N = 1$. From (40),

$$\tilde{a}^*(B((1-\varepsilon)c_N)) \leq T^*_{N,b} \leq \tilde{a}^*(B((1+\varepsilon)c_N)),\ wp \to_{\mathcal{P}_{00}(\varepsilon)} 1.$$

Therefore,

$$\begin{aligned} \left|T^*_{N,b} - \tilde{a}^*(B((1-\varepsilon)c_N))\right| &\leq \left|\tilde{a}^*(B((1+\varepsilon)c_N)) - \tilde{a}^*(B((1-\varepsilon)c_N))\right| \\ &\leq \left|a^*(B((1+\varepsilon)c_N)) - a^*(B((1-\varepsilon)c_N))\right| + o_{P^*}(1) \\ &= o_{P^*}(1), \end{aligned}$$

uniformly in $P \in \mathcal{P}_{00}(\varepsilon)$, by Claim 2 and (17). This implies that

$$T^*_{N,b} = \tilde{a}^*(B((1-\varepsilon)c_N)) + o_{P^*}(1) = a^*(B(N^{-1/2}r_N)) + o_{P^*}(1) \text{ in } P, \text{ uniformly in } P \in \mathcal{P}_{00}(\varepsilon),$$

by Claim 2 and (17) again. Hence $T^*_{N,b} = f(\nu^*_0) + o_{P^*}(1)$ in P, uniformly in $P \in \mathcal{P}_{00}(\varepsilon)$, where

$$f(v) \equiv \int_{B(N^{-1/2}r_N)} \max\{v(x), 0\}^2 w(x)dx.$$

Let $\bar{c}^*_{\alpha,N,\infty} \equiv \inf\{c : P\{f(\nu^*_0) \leq c|\mathcal{G}_N\} \geq 1-\alpha\}$ and $c_{\alpha,\infty} \equiv \inf\{c : P\{f(\nu) \leq c\} \geq 1-\alpha\}$. By Lemma B4, we have

$$\bar{c}^*_{\alpha,N,\infty} = c_{\alpha,\infty} + o_{\mathcal{P}_{00}(\varepsilon)}(1), \tag{48}$$

due to the regularity of ν on $B(N^{-1/2}r_N)$. We turn to the original test statistic T_N. Using (44) and the steps thereafter, we obtain that for any sequence $r'_N \to 0$,

$$\begin{aligned} T_N &= \int_{B((1-\varepsilon)c_N)} \max\left\{\sqrt{N}q(x)(\bar{D}_{01} - D_{01})(x) + \sqrt{N}q(x)D_{01}(x), 0\right\}^2 w(x)dx + O_{\mathcal{P}}(r'_N) \\ &= \int_{B((1-\varepsilon)c_N)\setminus B(N^{-1/2}r_N)} \max\left\{\sqrt{N}q(x)(\bar{D}_{01} - D_{01})(x) + \sqrt{N}q(x)D_{01}(x), 0\right\}^2 w(x)dx \\ &\quad + \int_{B(N^{-1/2}r_N)} \max\left\{\sqrt{N}q(x)(\bar{D}_{01} - D_{01})(x), 0\right\}^2 w(x)dx + O_{\mathcal{P}}(\max\{r_N^2, r'_N\}). \end{aligned}$$

The third to the last term is bounded by

$$\int_{B((1-\varepsilon)c_N)\setminus B(N^{-1/2}r_N)} \max\left\{\sqrt{N}q(x)(\bar{D}_{01} - D_{01})(x), 0\right\}^2 w(x)dx = O_{\mathcal{P}_{00}(\varepsilon)}(r_N) = o_{\mathcal{P}_{00}(\varepsilon)}(1),$$

by (17) and the uniform asymptotic tightness of $\sqrt{N}q(\bar{D}_{01} - D_{01})$. We conclude that $T_N = f(\bar{\nu}) + o_{\mathcal{P}_{00}(\varepsilon)}(1)$. Let the $(1-\alpha)$-th quantile of the distribution of $f(\bar{\nu})$ be $\bar{c}_{\alpha,N}$. Then by applying Lemma B4, we obtain that

$$\bar{c}_{\alpha,N} = c_{\alpha,\infty} + o_{\mathcal{P}_{00}(\varepsilon)}(1).$$

Combined with (48), this implies that

$$P\left\{T_N \leq \bar{c}^*_{\alpha,N,\infty}\right\} = P\left\{T_N \leq c_{\alpha,\infty}\right\} + o(1) = P\left\{T_N \leq \bar{c}_{\alpha,N}\right\} + o(1) = 1 - \alpha + o(1),$$

uniformly in $P \in \mathcal{P}_{00}(\varepsilon)$. Therefore, we obtain the wanted result.

ACCEPTED MANUSCRIPT

(Proof of Claim 1): Since the empirical process $\sqrt{N}q(x)\{D_k(x) - \bar{D}_k(x)\}$ is asymptotically tight for each $k = 0, 1$, $P\{\sup_x(q(x)|D_0(x) - \bar{D}_0(x)| + q(x)|D_1(x) - \bar{D}_1(x)|) > \varepsilon c_N\} \to 0$ by the choice of $c_N \to 0$ and $c_N\sqrt{N} \to \infty$. For any $x \in B((1-\varepsilon)c_N)$, by the triangular inequality,

$$q(x)\left|\bar{D}_{01}(x)\right| \leq (1-\varepsilon)c_N + q(x)\left|D_0(x) - \bar{D}_0(x)\right| + q(x)\left|D_1(x) - \bar{D}_1(x)\right| \leq c_N,$$

with probability approaching one. Thus we deduce that $P\{B((1-\varepsilon)c_N) \subset \hat{B}\} \to 1$. Now, for any $x \in \hat{B}$, by the triangular inequality,

$$q(x)\left|D_{01}(x)\right| \leq c_N + q(x)\left|D_0(x) - \bar{D}_0(x)\right| + q(x)\left|D_1(x) - \bar{D}_1(x)\right| \leq (1+\varepsilon)c_N,$$

with probability approaching one. Therefore, $P\{\hat{B} \subset B((1+\varepsilon)c_N)\} \to 1$. 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): Take $\delta_N = CN^{-1/4+\varepsilon} \to 0$ for $\varepsilon > 0$ and consider only those events such that $||\hat{\theta} - \theta_0|| + ||\hat{\tau} - \tau_0||_\infty \leq \delta_N$. By Lemma B3(i), we have $|\tilde{a}^*(B_N) - \hat{a}^*(B_N)| = o_{P^*}(1)$. By using $a^2 - b^2 = (a+b)(a-b)$ and the Cauchy-Schwartz inequality, the above term is bounded by

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

because both the processes $\nu^*(\cdot,\hat{\theta},\hat{\tau})$ and $\nu^*(\cdot,\theta_0,\tau_0)$ are asymptotically tight by the weak convergence result of Lemma B3(ii). Using Fubini's theorem, the last expectation is written as

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

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

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

where $CN^{-1/2+1/p} = o(1)$. Using Le Cam's poissonization lemma and following the proof of Theorem 2.2 of Giné (1997), we obtain that for $e = \exp(1)$,

$$\begin{aligned}
&\mathbf{E}\left(\mathbf{E}[\sup_{(\theta,\tau)\in B_{\Theta\times\mathcal{T}}(\delta_N)}|\nu^*(x,\theta,\tau) - \nu^*(x,\theta_0,\tau_0)||\mathcal{G}_N]\right) \qquad (49) \\
\leq\ & \frac{e}{e-1}\mathbf{E}\left[\sup_{(\theta,\tau)\in B_{\Theta\times\mathcal{T}}(\delta_N)}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N (N_i - 1)\{V_x(W_i,\theta,\tau) - V_x(W_i,\theta_0,\tau_0)\}\right|\right] \\
&+ \frac{e}{e-1}\mathbf{E}\left[\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N (N_i - 1)\right| \times \sup_{(\theta,\tau)\in B_{\Theta\times\mathcal{T}}(\delta_N)}\left|\frac{1}{N}\sum_{i=1}^N \{V_x(W_i,\theta,\tau) - V_x(W_i,\theta_0,\tau_0)\}\right|\right],
\end{aligned}$$

ACCEPTED MANUSCRIPT

where $V_x(W_i, \theta, \tau)$ is defined in (8) and $\{N_i\}$ is i.i.d. Poisson random variables of parameter 1 and independent of $\{W_i\}_{i=1}^N$.

Redefine $\mathcal{H}(\delta)$ as in the proof of Lemma B3(ii) with $V_x(\cdot, \theta, \tau)$ replaced by $V_x(\cdot, \theta, \tau) - V_x(\cdot, \theta_0, \tau_0)$. Then by Assumption 2(ii) and Assumption 4, we can take an envelope of $\mathcal{H}(\delta_N)$, whose $L_2(P)$-norm is bounded by $C(\delta_N^{\lambda s_2/2} + \delta_N^{s_1})$. Using the maximal inequality, we can obtain that the expectation in (49) is $o_P(1)$. Therefore,

$$\mathbf{E}\left[|\hat{a}^*(B_N) - a^*(B_N)| \, |\mathcal{G}_N\right] = o_P(1) \times \int_{B_N} w(x)dx = o_P(1)$$

because w is integrable. ■

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 $T^{*LF}_{N,b}$ is recentered, it converges in distribution (conditional on $\mathcal{G}_N$) to the distribution of $\int_{\mathcal{X}} \max\{\nu(x), 0\}^2 w(x)dx$, while the distribution of the bootstrap test statistic $T^*_{N,b}$ converges to that of $\int_{C_{01}} \max\{\nu(x), 0\}^2 w(x)dx$. Note that

$$\int_{\mathcal{X}} \max\{\nu(x), 0\}^2 w(x)dx \geq \int_{C_{01}} \max\{\nu(x), 0\}^2 w(x)dx$$

because $C_{01} \subset \mathcal{X}$. Hence $c^{*LF}_{\alpha,N,\infty} \geq c^*_{\alpha,N,\infty} + o_P(1)$. This implies that under the local alternatives,

$$\begin{aligned} \lim_{N\to\infty} P_N\left\{T_N > c^*_{\alpha,N,\infty}\right\} &= P\left\{\int_{C_{01}} \max\left\{\nu(x) + \delta(x), 0\right\}^2 w(x)dx > c^*_{\alpha,N,\infty}\right\} \\ &\geq P\left\{\int_{C_{01}} \max\left\{\nu(x) + \delta(x), 0\right\}^2 w(x)dx > c^{*LF}_{\alpha,N,\infty}\right\}. \end{aligned} \tag{50}$$

Now, when $\int_{\mathcal{X}} \max\left\{\nu(x), 0\right\}^2 w(x)dx > \int_{C_{01}} \max\left\{\nu(x), 0\right\}^2 w(x)dx$ almost everywhere, $c^{*LF}_{\alpha,N,\infty} > c^*_{\alpha,N,\infty} + o_P(1)$. Hence the inequality in (50is strict. ■

ACCEPTED MANUSCRIPT

Acknowledgements

We would like to thank Donald Andrews, Yanqin Fan, Hidehiko Ichimura, Taisuke Otsu, Peter Phillips, and participants at several conferences and seminars for useful suggestions and helpful comments. We would also like to thank two referees who gave us valuable comments. Linton thanks the ESRC and Leverhulme foundations for financial support. This paper was partly written while Linton was a Universidad Carlos III de Madrid-Banco Santander Chair of Excellence, and he thanks them for financial support. This paper was partly written while Whang was visiting the Cowles Foundation at Yale University, whose hospitality is gratefully acknowledged.

References

Abrevaya, J. and J. Huang, 2005, On the bootstrap of the maximum score estimator. Econometrica 73, 1175-1204.

Anderson, G. J., 1996, Nonparametric tests of stochastic dominance in income distributions. Econometrica 1183-1193.

Andrews, D. W. K., 1994, Empirical process methods in econometrics, in: R. F. Engle and D. L. McFadden, (Eds.), Handbook of Econometrics, Vol. IV, Amsterdam: North-Holland.

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.

Andrews, D. W. K. and G. Soares, 2007, Inference for parameters defined by moment inequalities using generalized moment selection. Cowles Foundation Discussion Paper 1631.

Arcones, M. A. and E. Giné, 1992, On the bootstrap of M-estimators and other statistical functionals, in: R. LePage and L. Billard, (Eds.), Exploring the Limits of Bootstrap, Wiley, pp.13-47.

Barret, G. and S. Donald, 2003, Consistent tests for stochastic dominance. Econometrica 71,71-104.

Bennett, C. J., 2007, New consistent integral-type tests for stochastic dominance. Working paper.

Bickel, P. J. , A.J. Klaassen, Y. Ritov, and J. A. Wellner, 1993, Efficient and Adaptive Estimation for Semiparametric Models, Springer Verlag, New York.

ACCEPTED MANUSCRIPT

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.

Bugni, F., 2008, Bootstrap inference in partially identified models. Working paper.

Chen, X., O. Linton, and I. van Keilegom, 2003, Estimation of semiparametric models when the criterion function is not smooth. Econometrica 71, 1591-1608.

Chernozhukov, V., H. Hong, and E. Tamer, 2007, Estimation and confidence regions for parameter sets in econometric models. Econometrica 75, 1243-1284.

Davidson, R. and J. Y. Duclos, 2000, Statistical inference for stochastic dominance and measurement for the poverty and inequality. Econometrica 68, 1435-1464.

Fan, Y. and S. Park, 2007, Confidence sets for some partially identified models. Working Paper.

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.

Giné, E. and J. Zinn, 1990, Bootstrapping general empirical measures. Annals of Probability 18, 851-869.

Giné, E. and J. Zinn, 1991, Gaussian characterization of uniform Donsker classes of functions. Annals of Probability 19, 758-782.

Hall, P. and J. L. Horowitz, 1996, Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64, 891-916.

Hansen, P., 2003, Asymptotic tests of composite hypotheses. Working Paper.

Hansen, P., 2005, A test for superior predictive ability. Journal of Buiness, Economics, and Statistics 23, 365-380.

Härdle, W. and E. Mammen, 1993, Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926-1947.

Heckman, J. J., Ichimura, H. and P. Todd, 1997, Matching as an econometric evaluation estimator: theory and evidence on its performance applied to the JTPA program, Part I. theory and methods, Review of Economic Studies 64, 605-654.

Horváth, L., P. Kokoszka, and R. Zitikis, 2006, Testing for stochastic dominance using the weighted McFadden-type statistic. Journal of Econometrics 133, 191-205.

ACCEPTED MANUSCRIPT

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.

Klecan, L., R. McFadden, and D. McFadden, 1991, A robust test for stochastic dominance. Working paper.

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.

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.

McFadden, D., 1989, Testing for stochastic dominance, in: in T. B. Fomby and T. K. Seo, Studies in the Economics of Uncertainty: In Honor of Josef Hadar, New York, Berlin, London and Tokyo: Springer.

Mikusheva, A., 2007, Uniform inference in autoregressive models. Econometrica 75, 1411-1452.

Moon, R. H. and F. Schorfheide, 2006, Boosting your instruments: estimation with overidentifying inequality moment conditions. Working Paper.

Newey, W. K., J. L. Powell, and J. R. Walker, 1990, Semiparametric sample selection models: some empirical results. American Economic Review 80, 324-328.

Sheehy, A. and J. A. Wellner, 1992, Uniform Donsker classes of functions. Annals of Probability 20, 1983-2030.

van der Vaart, A. W. and J. A. Wellner, 1996, Weak Convergence and Empirical Processes, Springer-Verlag, New York.

Wellner, J. A. and Y. Zhan, 1996, Bootstrapping Z-estimators. Working Paper.

Whang, Y. J., 2000, Consistent bootstrap tests of parametric regression functions. Journal of Econometrics 98, 27-46.