Journal of Business & Economic Statistics

Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/ubes20

Conditional Stochastic Dominance Testing

Miguel A. Delgado [a] & Juan Carlos Escanciano [b]

[a] Universidad Carlos III de Madrid, Getafe, 28903, Spain

[b] Indiana University, Bloomington, IN, 47405

Accepted author version posted online: 28 Aug 2012.Version of record first published: 28 Jan 2013.

To cite this article: Miguel A. Delgado & Juan Carlos Escanciano (2013): Conditional Stochastic Dominance Testing, Journal of Business & Economic Statistics, 31:1, 16-28

To link to this article: http://dx.doi.org/10.1080/07350015.2012.723556

Conditional Stochastic Dominance Testing

Miguel A. Delgado
Universidad Carlos III de Madrid, Getafe-28903, Spain (*miguelangel.delgado@uc3m.es*)

Juan Carlos Escanciano
Indiana University, Bloomington, IN 47405 (*jescanci@indiana.edu*)

This article proposes bootstrap-based stochastic dominance tests for nonparametric conditional distributions and their moments. We exploit the fact that a conditional distribution dominates the other if and only if the difference between the marginal joint distributions is monotonic in the explanatory variable at each value of the dependent variable. The proposed test statistic compares restricted and unrestricted estimators of the difference between the joint distributions, and it can be implemented under minimal smoothness requirements on the underlying nonparametric curves and without resorting to smooth estimation. The finite sample properties of the proposed test are examined by means of a Monte Carlo study. We illustrate the test by studying the impact on postintervention earnings of the National Supported Work Demonstration, a randomized labor training program carried out in the 1970s.

KEY WORDS: Conditional inequality restrictions; Least concave majorant; Nonparametric testing; Treatment effects.

1. INTRODUCTION

Stochastic dominance plays a major role in applied research, particularly in economics. It has been used to rank investment strategies, to measure income and poverty inequality, or to assess treatment effects, social programs, or policies. The earliest proposal of Smirnov (1939) in the classical two-sample problem has been followed by numerous extensions to different concepts of stochastic dominance under alternative data-generating process assumptions; see, for example, McFadden (1989), Anderson (1996), Davidson and Duclos (2000), Barrett and Donald (2003), Linton, Maasoumi, and Whang (2005), or Scaillet and Topaloglou (2010), among others. This literature has been confined, however, to unconditional stochastic dominance testing, and although there are some proposals that can accommodate covariate heterogeneity, these tests are only consistent under rather strong independence assumptions between regression errors and covariates. This article proposes consistent tests for conditional stochastic dominance and other conditional moment inequalities under mild regularity conditions on the underlying data-generating process and without requiring smoothed estimates.

Related to testing conditional stochastic dominance is a great deal of literature on two-sided tests for the equality of nonparametric regression curves. Some of these tests compare smooth estimators of the nonparametric curves, like Härdle and Marron (1990), Hall and Hart (1990), or King, Hart, and Wehrly (1991). Others avoid smooth estimation of conditional moments by comparing estimates of their integrals, like Delgado (1993) or Ferreira and Stute (2004). The literature on one-sided tests of conditional moment restrictions is by contrast rather scarce, and more recent. Tests for nonpositiveness of conditional moments can be based on the positive part of a smoothed estimator, as suggested by Hall and Yatchew (2005) or Lee and Whang (2009). A related idea was implemented by Linton, Song, and Whang (2010), who used the positive part of the difference between sample distributions to test stochastic dominance. One can avoid using smoothers by noticing that a conditional moment is nonpositive if and only if its integral is monotonically nonincreasing. This fact was exploited by Kim (2008) and Andrews and Shi (2010) for constructing confidence intervals of parameters partially identified by means of conditional moment inequalities. See also Khan and Tamer (2009) for an application to censored regression. So, as Andrews and Shi (2010) suggested, a test of monotonicity on the integrated curve can be used for testing the inequality restrictions.

Our approach is also based on integrated moments but relies on a different methodology to that of the aforementioned works. We first characterize the problem of testing for monotonicity of the integrated moment as one of testing for concavity, by integrating one more time. Then, instead of a Wald-type test statistic, as in Kim (2008) or Andrews and Shi (2010), we consider a likelihood ratio (LR)-type approach, comparing restricted and unrestricted estimates of the double-integrated conditional moment. Our approach is then more related to classical LR tests for parameter inequality restrictions, see Dykstra and Robertson (1982, 1983), Robertson, Wright, and Dykstra (1988), Wolak (1989), or Kodde and Palm (1986). However, unlike in this classical literature, our null hypothesis is nonparametric, that is, involves infinite restrictions. The restricted estimator of the integrated conditional moment is in fact an isotonic estimator, which does not use smoothers. See Barlow et al. (1972) for a comprehensive account of results on isotonic estimation, and see Durot (2003) and Delgado and Escanciano (2012) for applications of the isotonic regression principles to conditional moment monotonicity testing. The proposed conditional stochastic dominance test is easy to implement using available algorithms for nonparametric isotonic estimation. Also, it can be implemented under fairly weak assumptions on the underlying data-generating process, and it is fully data-driven, without requiring user-chosen parameters such as bandwidths.

Journal of Business & Economic Statistics
January 2013, Vol. 31, No. 1
DOI: 10.1080/07350015.2012.723556

In this article, we focus on the first-order conditional stochastic dominance testing problem in a one-sample setting. Under the null, the difference between the two conditional distributions, or their moments, is nonpositive/nonnegative. The null hypothesis is satisfied if and only if the difference between the corresponding unconditional joint distribution functions is monotonic with respect to the explanatory variable. Thus, our tests consist of comparing restricted and unrestricted estimates of the difference between the joint distribution functions. The limiting distribution of the test statistic is nonpivotal in the least favorable case (l.f.c.), that is, the case under the null closest to the alternative, but critical values can be consistently estimated with the assistance of a bootstrap procedure as shown below.

The test statistic designed for testing conditional stochastic dominance is easily adapted to testing inequality restrictions on other conditional moments, possibly indexed by unknown parameters that must be estimated. Likewise, higher-order stochastic dominance can be easily accommodated. Our testing procedure is particularly well suited for the evaluation of treatment programs. We apply the testing method to the National Supported Work (NSW) Demonstration program, a randomized labor training program carried out in the 1970s, which has been employed for illustrating different proposals for treatment effect evaluation ever since the landmark article by Lalonde (1986). In this application, we find evidence against a nonnegative average treatment effect conditional on age when the whole age distribution is included, and we show that this rejection is mainly due to young individuals between 17 and 21 years old. For these young individuals, the job training program was not beneficial. Unconditional methods are unable to uncover this age heterogeneity in treatment effects. This feature of the data is also missed by methods using smoothers, because of their lack of precision in the tails of the age distribution, where there are few observations. Hence, this application highlights the merits of the proposed methodology—the conditional aspect and the gains in precision derived from estimating integrals rather than derivatives.

We have organized the article as follows. In the next section, we present the testing procedure. Section 3 is devoted to applications of the basic framework to situations of particular practical relevance. We consider testing inequality restrictions on conditional moments, possibly indexed by unknown parameters, which are illustrated with an application to testing conditional treatment effects in social programs. We also discuss the application of the testing procedure when conditioning on a vector of covariates. A Monte Carlo study in Section 4 investigates the finite sample properties of the proposal. We also report in this section an application of the procedure to the NSW study. In Section 5, we conclude and suggest extensions for future research. Mathematical proofs are gathered in an Appendix at the end of the article.

2. CONDITIONAL STOCHASTIC DOMINANCE TESTING

Henceforth, all the random variables are defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Any generic random vector $\boldsymbol{\xi}$ takes values in $\mathcal{X}_{\boldsymbol{\xi}}$, $F_{\boldsymbol{\xi}}$ denotes its cumulative distribution function (cdf), and for each pair of random vectors $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ on $(\Omega, \mathcal{A}, \mathbb{P})$, $F_{\boldsymbol{\xi}_1|\boldsymbol{\xi}_2}$ denotes the conditional cdf of $\boldsymbol{\xi}_1$ given $\boldsymbol{\xi}_2$, that is,

$$F_{(\xi_1,\xi_2)}(t_1, t_2) = \int_{-\infty}^{t_2} F_{\xi_1|\xi_2}(t_1, \bar{t}_2)\, F_{\xi_2}(d\bar{t}_2).$$

Given an $\mathbb{R}^3$-*valued* random vector (Y_1, Y_2, X) and sets $\mathcal{W}_Y \subseteq \mathcal{X}_{Y_1} \cap \mathcal{X}_{Y_2}$ and $\mathcal{W}_X \subseteq \mathcal{X}_X$, such that $\mathcal{W}_Y \times \mathcal{W}_Y \times \mathcal{W}_X \subseteq \mathcal{X}_{(Y_1,Y_2,X)}$, we consider the hypothesis

$$H_0 : F_{Y_1|X} \leq F_{Y_2|X} \text{ a.s. in the set } \mathcal{W}_Y \times \mathcal{W}_X. \quad (1)$$

The alternative hypothesis H_1 is the negation of H_0. We allow, but do not require, that $\mathcal{W}_Y \times \mathcal{W}_Y \times \mathcal{W}_X \equiv \mathcal{X}_{(Y_1,Y_2,X)}$. The discussion is centered on the case where X is univariate and F_X is continuous. When X is discrete, the conditional distribution can be estimated $\sqrt{n}$-consistently, and H_0 can be tested using simple modifications of existing unconditional methods. In Section 4, we consider the implementation when X is multivariate, where some of the components, but not all, can be discrete.

Note that H_0 is satisfied if and only if the difference between the joint distributions,

$$D(y, x) \equiv (F_{(Y_1,X)} - F_{(Y_2,X)})(y, x) = \int_{-\infty}^{x} (F_{Y_1|X} - F_{Y_2|X})(y, \bar{x})\, F_X(d\bar{x}),$$

is nonincreasing in $x \in \mathcal{W}_X$, for each $y \in \mathcal{W}_Y$. In turn, since the quantile function F_X^{-1} is nondecreasing, a necessary and sufficient condition for (1) is that

$$C(y, u) \equiv \int_0^u D\left(y, F_X^{-1}(\bar{u})\right) d\bar{u}$$

is concave in $u \in \mathcal{U}_X \equiv F_X(\mathcal{W}_X)$, for each $y \in \mathcal{W}_Y$. That is, the null hypothesis is satisfied if and only if the integrated curve D is monotonically nonincreasing in $x \in \mathcal{W}_X$, for each $y \in \mathcal{W}_Y$. Then, the monotonicity of D is satisfied if and only if its integrated curve, C, is concave with respect to its second argument.

Therefore, H_0 can be characterized by the least concave majorant (l.c.m.) operator $\mathcal{T}$, which is defined as follows in this bivariate context. Let $\mathcal{C}$ be the space of concave functions on $[0, 1]$. For any generic measurable function $g : \mathcal{W}_Y \times \mathcal{U}_X \to \mathbb{R}$, $\mathcal{T}g(y, \cdot)$ is the function satisfying the following two properties for each $y \in \mathcal{W}_Y$: (1) $\mathcal{T}g(y, \cdot) \in \mathcal{C}$ and (2) if there exists $h \in \mathcal{C}$ with $h \geq g(y, \cdot)$, then $h \geq \mathcal{T}g(y, \cdot)$. Henceforth, $\mathcal{T}g$ denotes the function resulting from applying the operator $\mathcal{T}$ to the function $g(y, \cdot)$ for each $y \in \mathcal{W}_Y$. Obviously, for a concave function g on $[0, 1]$, $\mathcal{T}g = g$. Thus, H_0 can be rewritten as an equality restriction,

$$H_0 : \mathcal{T}C - C = 0, \text{ a.s. in the set } \mathcal{W}_Y \times \mathcal{U}_X.$$

This suggests, using as test statistic, some functional of an estimator of $\mathcal{T}C - C$. Let $\mathcal{Z}_n \equiv \{(Y_{1i}, Y_{2i}, X_i)\}_{i=1}^n$ be independent and identically distributed (iid) observations of $\mathbf{Z} \equiv (Y_1, Y_2, X)$. Henceforth, for a given generic sample $\{\boldsymbol{\xi}_i\}_{i=1}^n$ of a possibly multivariate random vector $\boldsymbol{\xi}$, let $F_{\boldsymbol{\xi}n}$ denote its corresponding empirical cdf and $F_{\boldsymbol{\xi}n}^{-1}$ its corresponding empirical quantile. A natural estimator of C is

$$C_n(y, u) \equiv \int_0^u D_n\left(y, F_{Xn}^{-1}(\bar{u})\right) d\bar{u}, \quad (y, u) \in \mathcal{W}_Y \times \mathcal{U}_X,$$

where

$$D_n(y,x) \equiv (F_{(Y_1,X)n} - F_{(Y_2,X)n})(y,x), \quad (y,x) \in \mathcal{W}_Y \times \mathcal{W}_X.$$

Notice that $D_n(F_{Yn}^{-1}(v), F_{Xn}^{-1}(u))$, $(v,u) \in [0,1]^2$ is the sample analog of the difference between the copula functions of (Y_1, X) and (Y_2, X), $D(F_Y^{-1}(v), F_X^{-1}(u))$, which has been considered by Rémillard and Scaillet (2009) and Bücher and Dette (2010) for copula equality testing among others.

The test statistic is the sup-distance between $\mathcal{T}C_n$ and C_n, that is,

$$\eta_n \equiv \sqrt{n} \sup_{(y,u)\in\mathcal{W}_Y\times\mathcal{U}_{Xn}} (\mathcal{T}C_n - C_n)(y,u), \tag{2}$$

where $\mathcal{U}_{Xn} \equiv F_{Xn}(\mathcal{W}_X)$ is the sample analog of $\mathcal{U}_X$. Of course, other distances could be used. Notice that

$$\hat{\eta}_n = \sqrt{n} \sup_{(y,u)\in\mathcal{W}_Y\times\mathcal{U}_{Xn}} \int_0^u \left(D_n^0 - D_n\right)\left(y, F_{Xn}^{-1}(\bar{u})\right) d\bar{u},$$

where $D_n^0(y, F_{Xn}^{-1}(u))$ is the slope of $\mathcal{T}C_n(y,u)$ for y fixed. Thus, $\hat{\eta}_n$ is in fact a distance between a restricted and an unrestricted estimator of the difference between the joint distribution functions.

2.1 Computation of the Test Statistic

Note that, for $(y,u) \in \mathcal{W}_Y \times \mathcal{U}_{Xn}$,

$$C_n(y,u) = \frac{1}{n}\sum_{i=1}^{n}(1_{\{Y_{1i}\le y\}} - 1_{\{Y_{2i}\le y\}})(u - F_{Xn}(X_i)) \times 1_{\{F_{Xn}(X_i)\le u\}}. \tag{3}$$

Therefore, it is evident from (3) that $C_n(y,\cdot)$ is, for each $y \in \mathcal{W}_Y$, piecewise linear with knots in $\mathcal{U}_{Xn}$, as is $\mathcal{T}C_n(y,\cdot)$. For each $y \in \mathcal{W}_Y$, we can always write

$$C_n\left(y, \frac{l}{n}\right) = \frac{1}{n}\sum_{j=1}^{l} r_{nj}(y), \; l = 1, \dots, n,$$

for a suitable sequence $\{r_{nj}(y)\}_{j=1}^n$ of increments of $C_n(y,\cdot)$, with $r_{n1}(y) \equiv 0$. In particular, when there are no ties in $\{X_i\}_{i=1}^n$, the function $r_{nj}(y)$ is given by

$$r_{nj}(y) \equiv \frac{1}{n}\sum_{i=1}^{j-1}(1_{\{Y_{1[i:n]}\le y\}} - 1_{\{Y_{2[i:n]}\le y\}}), \; j = 2, \dots, n, \tag{4}$$

where $\{Y_{j[i:n]}\}_{i=1}^n$, $j = 1, 2$, are the Y_j*-concomitants* of the order statistics $\{X_{i:n}\}_{i=1}^n$, that is, $Y_{j[i:n]} = Y_{jk}$ if $X_{i:n} = X_k$, $j = 1, 2$, with $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$.

The knots of $\mathcal{T}C_n(y,.)$, for each $y \in \mathcal{W}_Y$, are easily located applying the Pooled Adjacent Violators Algorithm (PAVA) proposed by Barlow et al. (1972). The input for the algorithm must be $\{r_{ni}(y)\}_{i=1}^n$, which can be easily computed recursively according to (4) when there are no X ties, or simply by computing the increments of $C_n(y,\cdot)$ in the general case. See Cran (1980) and Bril et al. (1984) for FORTRAN implementations and de Leeuw, Hornik, and Mair (2009) for R routines. Moreover, the maximum difference of $(\mathcal{T}C_n - C_n)(y,\cdot)$, with $y \in \mathcal{W}_Y$ fixed, is attained at one of the points in $\mathcal{U}_{Xn}$, restricting the supremum to a maximum on a finite number of points for each $n \ge 1$. Furthermore, $C_n(y,\cdot)$, and hence $\mathcal{T}C_n(y,\cdot)$, takes on the same values when y is between consecutive order statistics of the pooled sample $\{Y_{1i}, Y_{2i}\}_{i=1}^n$, which shows that $\sup_{y\in\mathcal{W}_Y}$ can also be computed as a maximum. Hence, we can simply write

$$\eta_n = \sqrt{n} \max_{(y,u)\in(\mathcal{U}_{Yn},\mathcal{U}_{Xn})} (\mathcal{T}C_n - C_n)(y,u), \tag{5}$$

where $\mathcal{U}_{Yn} \equiv \{Y_{ki} : Y_{ki} \in \mathcal{W}_Y, \; 1 \le i \le n, \; k = 1, 2\}$. Matlab subroutines for computing η_n are available from the authors upon request.

2.2 Asymptotic Distribution

We discuss now the asymptotic distribution of η_n under the l.f.c., which corresponds to (1) under equality. The limiting distribution follows from the functional central limit theorem applied to $\sqrt{n}C_n$ and the continuous mapping theorem. But it must be proved first that considering the empirical distribution function F_{Xn} in C_n and in the estimated set $\mathcal{U}_{Xn}$, rather than the genuine F_X, does not have any effect on the asymptotic distribution of the test statistic under the l.f.c. In the Appendix, we characterize the limiting distribution of η_n and prove that, under H_0,

$$\lim_{n\to\infty} \mathbb{P}\{\eta_n > c_\alpha\} \le \alpha,$$

where

$$c_\alpha = \inf\left\{c \in [0,\infty) : \lim_{n\to\infty} \mathbb{P}\{\eta_n > c\} \le \alpha \text{ in the l.f.c.}\right\}.$$

However, c_α is hard to estimate directly from the sample. We propose estimating c_α by means of a multiplier-type bootstrap. See Chapter 2.9 in van der Vaart and Wellner (1996). The asymptotic critical value c_α is estimated by

$$c_{n\alpha}^* \equiv \inf\{c \in [0,\infty) : \mathbb{P}_n^*(\eta_n^* > c) \le \alpha\},$$

where $\mathbb{P}_n^*$ means bootstrap probability, that is, conditional on the sample $\mathcal{Z}_n$,

$$\eta_n^* \equiv \sqrt{n} \max_{(y,u)\in(\mathcal{U}_{yn},\mathcal{U}_{Xn})} (\mathcal{T}C_n^* - C_n^*)(y,u)$$

and, for each $(y,u) \in \mathcal{W}_Y \times \mathcal{U}_{Xn}$,

$$C_n^*(y,u) \equiv \frac{1}{n}\sum_{i=1}^{n}(1_{\{Y_{1i}\le y\}} - 1_{\{Y_{2i}\le y\}})(u - F_{Xn}(X_i)) \times 1_{\{F_{Xn}(X_i)\le u\}} V_i.$$

The random variables $\mathcal{V}_n \equiv \{V_i\}_{i=1}^n$ are iid, independently generated from the sample $\mathcal{Z}_n$, according to a random variable V with bounded support, mean zero, and variance one. This type of multiplicative bootstrap has been used in many problems involving empirical processes with a nonpivotal asymptotic distribution. See, for instance, Delgado and González-Manteiga (2001) or Scaillet (2005). In practice, $c_{n\alpha}^*$ is approximated as accurately as desired by $\eta_{n[B(1-\alpha)]}^*$, the $[B(1-\alpha)]$th order statistic computed from B replicates $\{\eta_{nj}^*\}_{j=1}^B$ of η_n^*. Equivalently, the test can be implemented using the bootstrap p-value $p_n^* = \mathbb{P}_n^*(\eta_n^* > \eta_n)$, which is also approximated by Monte Carlo. Our bootstrap test rejects H_0 at the αth nominal level, $\alpha \in (0,1)$, when $\eta_n > c_{n\alpha}^*$, or equivalently $p_n^* < \alpha$. The next theorem states that the bootstrap test is consistent and has the right asymptotic size.

Theorem 1. Assume that F_X is continuous and $\{V_i\}_{i=1}^n$ are iid, independent of the sample $\mathcal{Z}_n$, bounded, and with mean zero and variance one. Then, for each $\alpha \in [0, 1]$,

1. under H_0, $\lim_{n\to\infty} \mathbb{P}(\eta_n > c_{n\alpha}^*) \le \alpha$, with equality under the l.f.c.;
2. under H_1, $\lim_{n\to\infty} \mathbb{P}(\eta_n > c_{n\alpha}^*) = 1$.

Our methodology is directly applicable to testing second-order or, more generally, jth-order conditional stochastic dominance, $j \ge 2$, simply replacing the empirical process C_n by

$$C_{n,j}(y,u) \equiv \frac{1}{n}\sum_{i=1}^n (1_{\{Y_{1i}\le y\}} - 1_{\{Y_{2i}\le y\}})(u - F_{Xn}(X_i))^j \times 1_{\{F_{Xn}(X_i)\le u\}}, \ j \ge 2.$$

See, for example, McFadden (1989) for discussion of higher-order stochastic dominance.

The test is also applicable to testing inequality restrictions of general conditional moments, possibly indexed by parameters, and it can be accommodated to situations with multiple covariates. These applications are discussed in the next section.

3. SOME APPLICATIONS OF THE BASIC FRAMEWORK

3.1 Conditional Moment Inequalities With Unknown Parameters

We apply the basic framework to testing inequality restrictions on general conditional moments of functions of the observable variables, which may be indexed by unknown parameters. That is, given a random vector $\mathbf{Z}$ and a measurable function $m_{\boldsymbol\theta} : \mathcal{X}_{\mathbf{Z}} \to \mathbb{R}$ indexed by a vector of parameters $\boldsymbol\theta \in \Theta$, where $\Theta \subseteq \mathbb{R}^k$ is a parameter space, the null hypothesis of interest is

$$H_0 : \mathbb{E}(m_{\boldsymbol\theta_0}(\mathbf{Z})|X = x) \le 0 \text{ for all } x \in \mathcal{W}_X \text{ and some } \boldsymbol\theta_0 \in \Theta. \tag{6}$$

Many applications fall under this setting. When $\mathbf{Z} = (Y_1, Y_2, X)$ and $m_{\boldsymbol\theta}(\mathbf{Z}) = Y_1 - Y_2$, Equation (6) is the hypothesis that a regression function dominates another. A version of the null hypothesis (6) is natural in treatment program evaluation. Let D be an indicator of participation in the program, that is, $D = 1$ if the individual participates in the treatment and $D = 0$ otherwise. Denote the observed outcome by $Y = Y(1)D + Y(0)(1 - D)$, where $Y(1)$ and $Y(0)$ are the potential outcomes of the individual in the treatment and control groups, respectively. We assume unconfoundedness or selection on observables, that is, $Y(1)$ and $Y(0)$ are independent of D, conditional on the covariate X. The hypothesis of interest is that the treatment is beneficial for individuals with $x \in \mathcal{W}_X$, that is,

$$\mathbb{E}(Y(0)|X = x) \le \mathbb{E}(Y(1)|X = x), \qquad \forall x \in \mathcal{W}_X. \tag{7}$$

Let $q(x) \equiv \mathbb{E}(D|X = x)$ be the propensity score, and assume that $q \in (0, 1)$ a.s. In applied work, it is usually assumed that $q(x) = q_{\boldsymbol\theta_0}(x)$ for some $\boldsymbol\theta_0 \in \Theta \subset \mathbb{R}^p$, where $q_{\boldsymbol\theta}$ is some cdf indexed by a vector of parameters $\boldsymbol\theta$; for example, a probit or a logit specification. Under these circumstances, using the fact that

$$\mathbb{E}((q_{\boldsymbol\theta_0}(X) - D)Y|X = x) = \{\mathbb{E}(Y(0)|X = x) - \mathbb{E}(Y(1)|X = x)\}q_{\boldsymbol\theta_0}(x)(1 - q_{\boldsymbol\theta_0}(x)),$$

the hypothesis in (7) can be rewritten as H_0 in (6) with $\mathbf{Z} = (Y, D, X)$ and $m_{\boldsymbol\theta}(\mathbf{Z}) = (q_{\boldsymbol\theta}(x) - D)Y$. Lee and Whang (2009) and Hsu (2011) implemented different tests for (7) based on smooth estimates of $E(Y(0) - Y(1)|X = x)$.

When $\boldsymbol\theta_0$ is known, the basic framework presented in the previous section is directly applicable without changes. For any generic function $m : \mathcal{X}_{\mathbf{Z}} \to \mathbb{R}^{d_m}$, we consider the test statistic

$$\bar\eta_{m,n} \equiv \sqrt{n} \max_{u\in\mathcal{U}_{Xn}} \left(\mathcal{T}\bar C_{m,n} - \bar C_{m,n}\right)(u),$$

where

$$\bar C_{m,n}(u) \equiv \frac{1}{n}\sum_{i=1}^n m(\mathbf{Z}_i)(u - F_{Xn}(X_i))1_{\{F_{Xn}(X_i)\le u\}}, \ u \in [0,1], \tag{8}$$

estimates $\bar C_m(u) \equiv \mathbb{E}(m(\mathbf{Z})(u - F_X(X))1_{\{F_X(X)\le u\}})$. When $\boldsymbol\theta_0$ is known, tests based on $\bar\eta_{m_{\boldsymbol\theta_0},n}$ are justified using the same arguments as in Theorem 1. Naturally, the stochastic dominance hypothesis between treatment and control groups conditional on the covariate X can be implemented by using $\mathbf{Z} = (Y, X, D)$ and $m(\mathbf{Z}) = (q_{\boldsymbol\theta}(x) - D)1_{\{Y\le y\}}$, which is also indexed by $y \in \mathcal{X}_Y$. A test for unconditional stochastic dominance has recently been proposed by Donald and Hsu (2011) based on the difference between the marginal distribution estimators of $Y(0)$ and $Y(1)$.

In many applications of practical relevance, the moment function $m_{\boldsymbol\theta_0}$ involves an unknown parameter $\boldsymbol\theta_0$. It happens when comparing productivity indexes that are residuals of some production function estimate, see, for example, Delgado, Fariñas, and Ruano (2002). It also happens when testing treatment effects with an unknown propensity score. In randomized experiments, D is independent of X, and hence, $q(x)$ is constant, say $q(x) \equiv \theta_0$. In this case, the parameter θ_0 can be estimated by its sample analog $\theta_n = n^{-1}\sum_{i=1}^n D_i$, which is the relative frequency of participants in the treatment. When dealing with non-experimental data, that is, if D and X are not mean-independent, q can be modeled by means of a discrete choice model depending on some unobserved latent variable, leading to $q = q_{\boldsymbol\theta_0}$ for some unknown $\boldsymbol\theta_0 \in \Theta \subset \mathbb{R}^p$.

Given iid observations $\{\mathbf{Z}_i\}_{i=1}^n$ of $\mathbf{Z}$, we assume that an $\sqrt{n}$-consistent estimator of $\boldsymbol\theta_0$ is available, which satisfies the following assumption.

Assumption E. The estimator $\boldsymbol\theta_n$ is strongly consistent for $\boldsymbol\theta_0$ and satisfies the following linear expansion:

$$\sqrt{n}(\boldsymbol\theta_n - \boldsymbol\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{l}_{\boldsymbol\theta_0}(\mathbf{Z}_i) + o_{\mathbb{P}}(1),$$

where $\mathbf{l}_{\boldsymbol\theta}(\cdot)$ is such that (1) $\mathbb{E}(\mathbf{l}_{\boldsymbol\theta_0}(\mathbf{Z})) = 0$ and $\mathbf{L}_{\boldsymbol\theta_0} \equiv \mathbb{E}(\mathbf{l}_{\boldsymbol\theta_0}(\mathbf{Z})\mathbf{l}_{\boldsymbol\theta_0}(\mathbf{Z})')$ exists and is positive definite and (2) $\lim_{\delta\to 0}\mathbb{E}(\sup_{\boldsymbol\theta\in\Theta_0, |\boldsymbol\theta-\boldsymbol\theta_0|\le\delta} |\mathbf{l}_{\boldsymbol\theta}(\mathbf{Z}) - \mathbf{l}_{\boldsymbol\theta_0}(\mathbf{Z})|^2) = 0$, where Θ_0 is a neighborhood of $\boldsymbol\theta_0$, $\Theta_0 \subset \Theta$.

We also need some smoothness on $m_{\boldsymbol\theta}$. Define $\dot{\mathbf{m}}_{\boldsymbol\theta} \equiv \partial m_{\boldsymbol\theta}/\partial\boldsymbol\theta$ a.s.

Assumption S. The moment function $m_{\boldsymbol{\theta}}$ is a.s. continuously differentiable in a neighborhood of $\boldsymbol{\theta}_0$, $\Theta_0 \subset \Theta$, with $\mathbb{E}(|m_{\boldsymbol{\theta}_0}(\mathbf{Z})|^2) < \infty$ and $\mathbb{E}(\sup_{\boldsymbol{\theta}\in\Theta_0} |\dot{\mathbf{m}}_{\boldsymbol{\theta}}(\mathbf{Z})|) < \infty$.

These assumptions are fulfilled under mild moment conditions when, for example, $m_{\boldsymbol{\theta}}(\mathbf{Z}) = \varepsilon_{1\boldsymbol{\theta}_1}(\mathbf{Z}) - \varepsilon_{2\boldsymbol{\theta}_2}(\mathbf{Z})$ with $\varepsilon_{i\boldsymbol{\theta}_i} : \mathcal{X}_{\mathbf{Z}} \times \Theta \to \mathbb{R}$, $i = 1, 2$, known functions and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$. For example, the ε_i''s may be the productivity indexes estimated as least-squares residuals of a Cobb–Douglas production function. These assumptions are also fulfilled in randomized experiments by $\theta_n = n^{-1}\sum_{i=1}^n D_i$, with $l_\theta(\mathbf{Z}) = (D-\theta)$ and $\dot{m}_\theta(\mathbf{Z}) = Y$, provided $0 < \theta_0 < 1$ and $\mathbb{E}(Y^2) < \infty$.

Under these two assumptions and the l.f.c., we show in the Appendix that $\bar{C}_{m_{\theta_n},n}$, defined as in (8), has the uniform in $u \in \mathcal{U}_X$ representation

$$\bar{C}_{m_{\theta_n},n}(u) = \frac{1}{n}\sum_{i=1}^n \{m_{\boldsymbol{\theta}_0}(\mathbf{Z}_i)(u - F_X(X_i))1_{\{F_X(X_i)\le u\}} + \mathbf{l}_{\boldsymbol{\theta}_0}'(\mathbf{Z}_i)\bar{\mathbf{C}}_{\dot{\mathbf{m}}_{\boldsymbol{\theta}}}(u)\} + o_{\mathbb{P}}(n^{-1/2}). \quad (9)$$

This uniform expansion suggests a simple bootstrap approximation based on

$$\bar{C}^*_{m_{\theta_n},n}(u) = \frac{1}{n}\sum_{i=1}^n \{m_{\boldsymbol{\theta}_n}(\mathbf{Z}_i)((u - F_{Xn}(X_i))1_{\{F_{Xn}(X_i)\le u\}} + \mathbf{l}_{\boldsymbol{\theta}_n}'(\mathbf{Z}_i)\bar{\mathbf{C}}_{\dot{\mathbf{m}}_{\boldsymbol{\theta}_{n,n}}}(u))\}V_i,$$

where $\{V_i\}_{i=1}^n$ are iid generated as indicated in Theorem 1. Let $\bar{\eta}^*_{m_{\theta_n,n}}$ be the bootstrap test statistic based on $\bar{C}^*_{m_{\theta_n},n}$, and denote by $\bar{c}^*_{\alpha,n}$ the corresponding bootstrap critical value. Our next result is the analog of Theorem 1 in the current setting.

Theorem 2. Let the assumptions of Theorem 1, E and S hold. Then,

1. under H_0, $\lim_{n\to\infty}\mathbb{P}(\bar{\eta}_{m_{\theta_n,n}} > \bar{c}^*_{\alpha,n}) \le \alpha$, with equality under the l.f.c.;
2. under H_1, $\lim_{n\to\infty}\mathbb{P}(\bar{\eta}_{m_{\theta_n,n}} > \bar{c}^*_{\alpha,n}) = 1$.

3.2 Multiple Covariates

In this section, we consider testing H_0 with $\mathbf{X}$ a d-dimensional covariate. We discuss two approaches. The first approach is based on the fact that the null hypothesis implies that for *all* $\boldsymbol{\beta} \in \mathbb{S}^d \equiv \{\boldsymbol{\beta} \in \mathbb{R}^d : \boldsymbol{\beta}'\boldsymbol{\beta} = 1\}$,

$$F_{Y_1|\boldsymbol{\beta}'\mathbf{X}}\left(y, \boldsymbol{\beta}'x\right) \le F_{Y_2|\boldsymbol{\beta}'\mathbf{X}}\left(y, \boldsymbol{\beta}'x\right) \text{ for all } (y, x) \in \mathcal{W}_Y \times \mathcal{W}_{\mathbf{X}}. \quad (10)$$

Escanciano (2006) considered a similar approach for the problem of testing the lack of fit of a regression model, and Kim (2008) has also used this approach for inferences under conditional moment inequalities. For each fixed $\boldsymbol{\beta} \in \mathbb{S}^d$, let $\hat{\eta}_n(\boldsymbol{\beta})$ denote the test statistic in (5) using the sample $\{Y_{1i}, Y_{2i}, \boldsymbol{\beta}'\mathbf{X}_i\}_{i=1}^n$. The test statistic for (10) is $\int_{\mathbb{S}^d}\hat{\eta}_n(\boldsymbol{\beta})d\boldsymbol{\beta}$. In applications, computing the integral can be a cumbersome task. For that reason, we propose the Monte Carlo approximation $\hat{\eta}_{n,m} \equiv m^{-1}\sum_{j=1}^m \hat{\eta}_n(\boldsymbol{\beta}_j)$, where $\{\boldsymbol{\beta}_j\}_{j=1}^m$ is a sequence of iid variables from a uniform distribution in $\mathbb{S}^d$, with $m \to \infty$ as $n \to \infty$. The sequence $\{\boldsymbol{\beta}_j\}_{j=1}^m$ can be easily generated from a d-dimensional vector of standard normals, scaled by its norm.

Alternatively, the researcher may be interested in particular choices of $\boldsymbol{\beta}_j$. For instance, $\boldsymbol{\beta}_j = (1, 0, \ldots, 0) \in \mathbb{S}^d$ leads to a test focusing on the conditional distributions of Y_k, $k = 1, 2$, given the first component of $\mathbf{X}$.

The limit distribution of $\hat{\eta}_{n,m}$ under the l.f.c. can be approximated by the bootstrap distribution of $m^{-1}\sum_{j=1}^m \hat{\eta}_n^*(\boldsymbol{\beta}_j)$, where $\hat{\eta}_n^*(\boldsymbol{\beta}_j)$ is the bootstrap approximation suggested in Section 2, using the same sequence $\mathcal{V}_n$ for $j = 1, \ldots, m$. The validity of the resulting bootstrap test follows from combining the empirical processes tools in Escanciano (2006) with our results of Section 2 in a routine fashion.

Alternatively, following a traditional approach in multivariate modeling, see the projection pursuit idea of Friedman and Tukey (1974), we could consider the composite hypothesis,

$$H_0 : F_{Y_1|\boldsymbol{\beta}_0'\mathbf{X}}(y, \boldsymbol{\beta}_0'x) \le F_{Y_2|\boldsymbol{\beta}_0'\mathbf{X}}(y, \boldsymbol{\beta}_0'x) \text{ for all } (y, x) \in \mathcal{W}_Y \times \mathcal{W}_{\mathbf{X}}, \quad (11)$$

where $\boldsymbol{\beta}_0$ is an unknown d-dimensional parameter, $\boldsymbol{\beta}_0 \in \Theta \subset \mathbb{R}^d$. For instance, such a situation arises in treatment effects when the conditional distribution of (Y, D) given $\mathbf{X}$ satisfies a single-index restriction, that is, $F_{(Y,D)|\mathbf{X}}(y, d) = F_{(Y,D)|\boldsymbol{\beta}_0'\mathbf{X}}(y, d)$ for some $\boldsymbol{\beta}_0 \in \Theta \subset \mathbb{R}^d$. A test for the composite hypothesis can be constructed based on $\hat{\eta}_n(\boldsymbol{\beta}_n)$, where $\boldsymbol{\beta}_n$ is a consistent estimator of $\boldsymbol{\beta}_0$ obtained from the single-index restriction, for example, by average derivatives or semiparametric least-square methods. The parameter $\boldsymbol{\beta}_0$ is only identified up to scale; so some normalization is in general needed. Here, it is technically convenient to normalize the first component of $\boldsymbol{\beta} \in \Theta$ to 1. In particular, we assume $\boldsymbol{\beta}_{01} = 1$. Furthermore, we also assume that this coefficient corresponds to a continuous component X_1 of $\mathbf{X} = (X_1, \mathbf{X}_{-1})$, where $\mathbf{X}_{-1} \equiv (X_2, \ldots, X_d)$. The following assumption requires smoothness for the conditional distribution of X_1 given $\mathbf{X}_{-1}$.

Assumption M. The conditional distribution of X_1 given $\mathbf{X}_{-1}$ has a (uniformly) bounded Lebesgue density. Furthermore, $\mathbb{E}(|\mathbf{X}|^2) < \infty$ and the parameter space Θ is compact.

We now show that under some mild regularity conditions, $\hat{\eta}_n(\boldsymbol{\beta}_n)$ and $\hat{\eta}_n(\boldsymbol{\beta}_0)$ have the same asymptotic distribution under the l.f.c. That is, asymptotically, the estimated parameters $\boldsymbol{\beta}_n$ do not have any effect on the limiting distribution under l.f.c. See Stute and Zhu (2005) for a related result in a different context. The bootstrap consistency of the test in this single-index model follows combining our results in Theorem 1 and the next Theorem in a routine fashion.

Theorem 3. Let Assumption M hold. Then, under the l.f.c. in (11), if $\boldsymbol{\beta}_n$ is a consistent estimator of $\boldsymbol{\beta}_0$, then

$$\hat{\eta}_n(\boldsymbol{\beta}_n) = \hat{\eta}_n(\boldsymbol{\beta}_0) + o_{\mathbb{P}}(1).$$

Theorem 3 also holds if we replace the index $\boldsymbol{\beta}_0'x$ by a general parametric index $v(\boldsymbol{\beta}_0, x)$, without significant changes in the proof. For instance, we could take $v(\boldsymbol{\beta}_0, x) = q_{\boldsymbol{\theta}_0}(x)$ and $\boldsymbol{\beta}_0 = \boldsymbol{\theta}_0$ in the treatment effect example, which is often used in applications. Furthermore, this result is also valid for more general index functions, including semiparametric or nonparametric ones, but formally proving this is beyond the scope of this article. The result in Theorem 3 is particularly convenient for ease of implementation of our test, as there is no need for reestimating the parameters $\boldsymbol{\beta}_0$ in each bootstrap iteration, or estimating the influence function of the estimator $\boldsymbol{\beta}_n$. Given data $\{\mathbf{Z}_i\}_{i=1}^n$, we estimate consistently $\boldsymbol{\beta}_0$, and then apply the test statistic of Section 2 to $\{Y_{1i}, Y_{2i}, \boldsymbol{\beta}_n'\mathbf{X}_i\}_{i=1}^n$, using the same multiplier-type bootstrap.

4. EMPIRICAL RESULTS

4.1 Monte Carlo Simulations

This section illustrates the finite sample performance of the tests by means of simulations and an application to testing treatment effects. The $\{V_i\}_{i=1}^n$ used in the bootstrap implementation are independently generated as V with $\mathbb{P}(V = 1 - \varphi) = \varphi/\sqrt{5}$ and $\mathbb{P}(V = \varphi) = 1 - \varphi/\sqrt{5}$, where $\varphi = (\sqrt{5} + 1)/2$. See Mammen (1993) for motivation on this popular choice. The bootstrap-critical values are approximated by Monte Carlo using 1000 replications and the simulations are based on 10,000 Monte Carlo experiments. We report rejection probabilities at 10%, 5%, and 1% significance levels.

We first investigate the size accuracy and power of the proposed conditional stochastic dominance tests for the following designs:

(i) $Y_1 = 1 + \varepsilon^{(1)}$; $Y_2 = 1 + \varepsilon^{(2)}$,
(ii) $Y_1 = \exp(X) + \varepsilon^{(1)}$; $Y_2 = \exp(X) + \varepsilon^{(2)}$,
(iii) $Y_1 = \sin(2\pi X) + \varepsilon^{(1)}$; $Y_2 = \sin(2\pi X) + \varepsilon^{(2)}$,
(iv) $Y_1 = 1 + \varepsilon^{(1)}$; $Y_2 = 1 + X + \varepsilon^{(2)}$,
(v) $Y_1 = \exp(X) + \varepsilon^{(1)}$; $Y_2 = \exp(X) + X + \varepsilon^{(2)}$,
(vi) $Y_1 = \sin(2\pi X) + \varepsilon^{(1)}$; $Y_2 = \sin(2\pi X) + X + \varepsilon^{(2)}$,
(vii) $Y_1 = 1 + \varepsilon^{(1)}$; $Y_2 = \sin(2\pi X) + \varepsilon^{(2)}$,
(viii) $Y_1 = \exp(X) + \varepsilon^{(1)}$; $Y_2 = \exp(X) + \sin(2\pi X) + \varepsilon^{(2)}$,
(ix) $Y_1 = \sin(2\pi X) + \varepsilon^{(1)}$; $Y_2 = 2\sin(2\pi X) + \varepsilon^{(2)}$,

where X is distributed as $U[0, 1]$, independently of the normal errors $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ that are independent, have zero mean, and variance $\sigma^2 = 1/4$. Similar designs were used by Neumeyer and Dette (2003) for testing the equality of regression functions in a two-sample context. Table 1 reports the proportion of rejections for models (i)–(ix) and sample sizes $n = 50$ and 150.

Models (i)–(iii) fall under the null hypothesis. We observe that our bootstrap test exhibits good size accuracy, even when $n = 50$. The power is moderate for $n = 50$ under alternatives (iv)–(viii), and uniformly high for any alternative with $n = 150$. The highest power is achieved for the alternative (ix), where the regression functions cross at one point.

Table 1. Rejection probabilities

n		50			150		
Model	α	10%	5%	1%	10%	5%	1%
(i)		0.099	0.042	0.006	0.104	0.050	0.008
(ii)		0.098	0.045	0.006	0.099	0.052	0.085
(iii)		0.099	0.046	0.006	0.101	0.050	0.087
(iv)		0.757	0.631	0.331	0.982	0.962	0.855
(v)		0.752	0.628	0.323	0.984	0.965	0.858
(vi)		0.749	0.630	0.323	0.982	0.963	0.855
(vii)		0.830	0.667	0.235	0.999	0.994	0.924
(viii)		0.827	0.662	0.227	0.998	0.993	0.930
(ix)		0.988	0.966	0.803	1.000	1.000	0.999

NOTES: One thousand bootstrap replications. Ten thousand Monte Carlo simulations.

Table 2. Rejection probabilities. Index model

n		50			150		
Model	α	10%	5%	1%	10%	5%	1%
(ii)		0.096	0.048	0.010	0.099	0.045	0.010
(iii)		0.089	0.044	0.013	0.102	0.043	0.013
(v)		0.996	0.981	0.745	1.000	1.000	1.000
(vi)		1.000	1.000	1.000	1.000	1.000	1.000
(viii)		0.257	0.151	0.046	0.732	0.572	0.250
(ix)		0.755	0.608	0.322	1.000	0.998	0.964

NOTES: 1,000 bootstrap replications. 10,000 Monte Carlo simulations.

In the second experiment, we study the finite sample performance of the stochastic dominance test applied to multivariate covariates with index restrictions. The designs are those of models (ii, iii) (under null) and models (v, vi, viii, and ix) (under the alternative), where X is replaced by the index $\boldsymbol{\beta}_0'\mathbf{X} \equiv X_1 + X_2 + X_3$, where X_j, $j = 1, 2$, and 3 are mutually independently distributed as $U[0, 1]$, and also independent of the normal errors $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$. The methodology is applied following a two-step approach. In the first step, the unknown parameter $\boldsymbol{\beta}_0 = (1, 1, 1)'$ is estimated from the data $\{(Y_{1i}, \mathbf{X}_i)\}_{i=1}^n$ by the minimum average variance estimator (MAVE) proposed in Xia et al. (2002) and denoted by $\boldsymbol{\beta}_n$. We implement the MAVE with a Gaussian kernel and a cross-validation method for choosing the bandwidth parameter. In a second step, the test is applied to the data $\{(Y_{1i}, Y_{2i}, \boldsymbol{\beta}_n'\mathbf{X}_i)\}_{i=1}^n$, as in the univariate case. Table 2 reports the proportion of rejections for sample sizes $n = 50$ and 150.

The size performance for models (ii) and (iii) is excellent for small sample sizes as $n = 50$. The obtained results support our asymptotic analysis. The estimation of $\boldsymbol{\beta}_0$ does not have an impact in the finite sample distribution, in agreement with the asymptotic equivalence of Theorem 3. Relative to the first set of experiments, the empirical power is higher for models (v) and (vi) and lower for (viii) and (ix), which is due to the additional uncertainty in the semiparametric estimation of the index parameter $\boldsymbol{\beta}_0$. This second set of simulations confirms our theoretical results—estimation of the nuisance parameter $\boldsymbol{\beta}_0$ does not affect the asymptotic distribution of our test under the l.f.c., but it may affect the power performance. It is remarkable that with a small sample size as $n = 50$, the asymptotic result already provides a good approximation of the finite sample distribution in a semiparametric context in which infinite-dimensional estimation is involved.

In our third experiment, we study the finite sample performance of the treatment effect test discussed in Section 3.1. We consider the design,

$$Y(0) = 1 - X + \varepsilon^{(1)}, \qquad (12)$$
$$Y(1) = 1 - c + (4c^2 - 1)X + cX^2 + \varepsilon^{(2)},$$

where X, $\varepsilon^{(1)}$, and $\varepsilon^{(2)}$ are generated as independent $U[0, 1]$ variables and c is a positive constant. The treatment

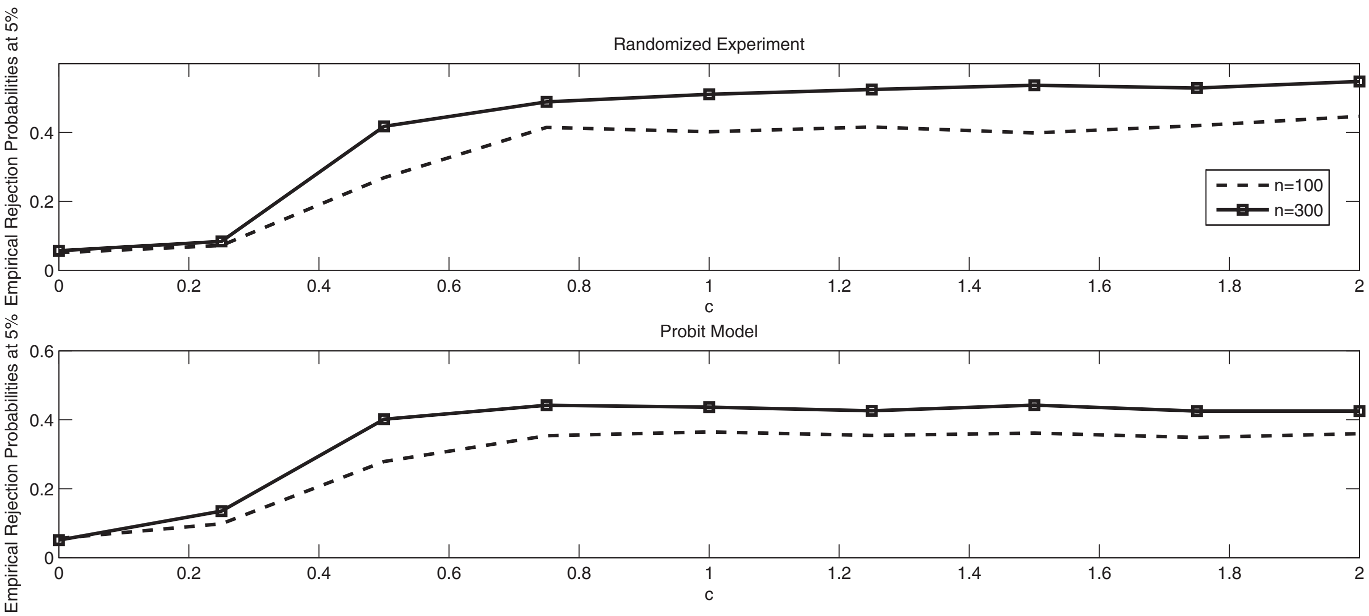

Figure 1. Five percent empirical power function for (12): randomized experiment (top panel) and probit model (bottom panel). Monte Carlo replications 10,000. $B = 1000$.

indicator is generated as $D = 1_{\{U^{(3)} \leq U^{(4)}\}}$, where $U^{(3)}$ and $U^{(4)}$ are independent copies of $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$. The observed outcome is $Y = Y(1)D + Y(0)(1 - D)$. The l.f.c. corresponds to $c = 0$ and, as c increases, the design deviates from the null in a direction somewhat similar to that observed in the empirical application in Section 4.2.

The top panel of Figure 1 reports the percentage of rejections as a function of c, for values of c from 0 to 2 at intervals of 0.25, and with $n = 100$ and 300. For $c = 0$, the size accuracy is excellent, with a proportion of rejections, when $n = 100$, of 1.1%, 5.1%, and 10.1% at 1%, 5%, and 10% of significance, respectively. The empirical power is nondecreasing in c, is low for $c = 0.25$, is high for $c \geq 0.5$, and stabilizes for $c \geq 0.75$.

In the fourth experiment, we relax the conditional mean independence between D and X, and generate data from (12) but with $D = 1_{\{\alpha_0 + \beta_0 X \leq \varepsilon\}}$, where $\boldsymbol{\theta}_0 \equiv (\alpha_0, \beta_0) = (1, 0.2)$ is assumed to be unknown, and ε follows a standard normal distribution, independently of the standard normal covariate X and the errors $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$. The propensity score is modeled by a probit model, and the parameter $\boldsymbol{\theta}_0$ is estimated by the conditional maximum likelihood estimator. The bottom panel of Figure 1 reports the percentage of rejections as a function of c, for sample sizes $n = 100$ and 300. The results for the nonrandomized experiment with a probit propensity score are qualitatively the same as for the randomized experiment.

Overall, the simulations show that the proposed bootstrap tests exhibit fairly good size accuracy and power for relatively small sample sizes, with good power across all alternatives considered.

4.2 An Application to Experimental Data

We apply the proposed testing method to studying the effectiveness of the NSW Demonstration program. The NSW was a randomized, temporary employment program carried out in the United States during the mid-1970s to help disadvantaged workers. In an influential article, Lalonde (1986) used the NSW experimental data to examine the performance of alternative statistical methods for analyzing nonexperimental data. Variations and subsamples of this dataset were later reanalyzed by Dehejia and Wahba (1999), among others. We use the original data for males in Lalonde (1986) to illustrate our procedure. For a comprehensive description of the experimental data, see Lalonde (1986) and Dehejia and Wahba (1999).

The data consist of 297 treatment group observations and 425 control group observations. Our dependent variable Y is the increment in earnings, measured in 1982 dollars, between 1978 (postintervention year) and 1975 (preintervention year). To illustrate our methods, we choose as independent variable X age. Figure 2 plots the kernel regression estimates for the period 1975–1978 with age restricted to its 10% and 90% quantiles to avoid boundary biases. We used a Gaussian kernel with bandwidth values 1 and 2 for the control and treatment groups, respectively. Cross-validation led to smaller bandwidths of 0.55 and 1.38, respectively, which imply undersmoothing. Nonparametric smoothed estimates suggest a positive treatment, especially for old workers. Parametric tests carried out by Lalonde (1986) for significance of the unconditional average treatment effect also indicated a positive effect.

The null hypothesis of nonnegative conditional mean treatment effect is considered, as in (7). The treatment was randomized, and hence, our hypothesis corresponds to (6) with $m_{\boldsymbol{\theta}_0}(\mathbf{Z}) = (\theta_0 - D)Y$, where $\theta_0 = \mathbb{E}(D)$ is consistently estimated by $\theta_n = n^{-1}\sum_{i=1}^{n} D_i$. The test statistic is implemented as in Section 3.1. In Table 3, we report the bootstrap p-values over 10,000 bootstrap replications of our test for several values of a_l in $\mathcal{W}_X = [a_l, 55]$. The value $a_l = 17$ corresponds to the full support of age in the data. Table 3 also contains the sample sizes of the control and treatment groups, n_1 and n_0, respectively.

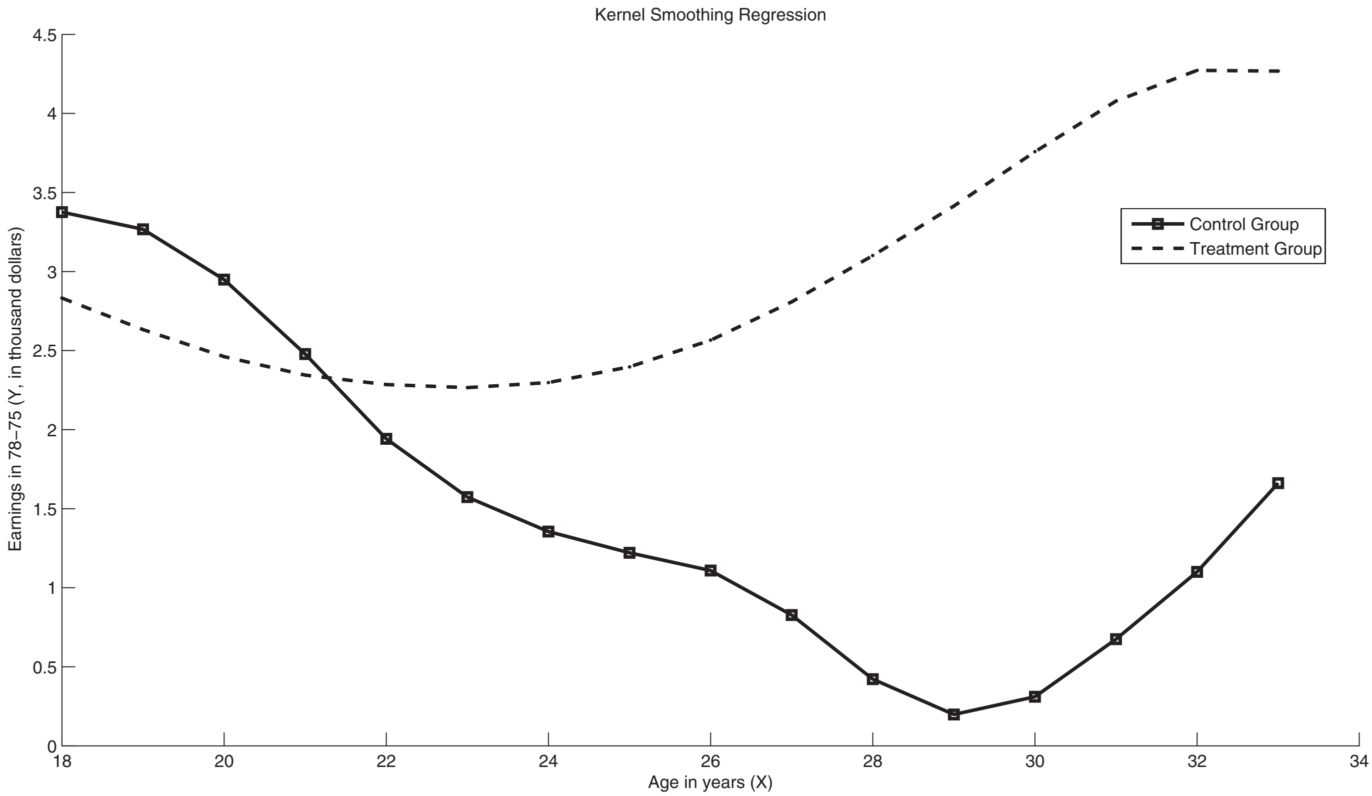

Figure 2. Nonparametric kernel estimates of the conditional means of changes in earnings between 1978 and 1975, as a function of age.

As evidenced from Table 3, our test rejects the null hypothesis of nonnegative impact of the NSW program at 5% when the whole age distribution is included ($a_l = 17$). Our results, in contrast with previous findings in the literature, provide evidence of treatment effect heterogeneity in age. Lee and Whang (2009) used the same dataset and failed to reject the null hypothesis of nonnegative conditional treatment effect using a test based on the L_1-distance of smoothed estimates of $E(Y(0) - Y(1)|X = x)$ and the space of nonpositive functions. Figure 2 and Table 3 suggest that the rejection is due to young individuals between 17 and 21 years old for whom the job training program was not beneficial, as measured by the incremental earnings between postintervention and preintervention years. This feature of the data is missed by methods using smoothers because their lack of precision in the tails of the age distribution implies a lack of power against small deviations of the null in the direction observed in this data.

Table 3. Nonparametric tests for the NSW. Bootstrap p-values

a_l	n_1	n_0	Mean age	Distribution age	Mean single-index
17	425	297	0.028	0.537	0.032
18	395	275	0.008	0.656	0.262
19	346	249	0.020	0.214	0.664
20	308	224	0.023	0.357	0.600
21	271	203	0.203	0.148	0.648
22	252	182	0.075	0.183	0.334
23	227	165	0.255	0.245	0.424
24	208	143	0.634	0.515	0.404

NOTES: 10,000 bootstrap replications. Cross-validated bandwidth.

For completeness, we have also applied the conditional stochastic dominance test for the whole distribution, that is, using $m_{\theta_0}(\mathbf{Z}) = (D - \theta_0)1_{\{Y_i \le y\}}$, which is also indexed by $y \in \mathcal{X}_Y$. The results are reported in Table 3. The test does not reject this hypothesis. That is, we reject the hypothesis that the treatment group dominates the control group in terms of the conditional means, but we cannot reject the stochastic dominance hypothesis in terms of the whole distribution. Notice that this does not lead to contradictory results. We can also arrive at the same conclusion in a pure parametric setting. For instance, when comparing confidence intervals and confidence ellipses on parameter restrictions, that is, we can reject a significance hypothesis on different single parameters, but we may be unable to reject the joint significance hypothesis on these parameters.

To check the robustness of the previous results to the inclusion of other covariates in the NSW study, we consider a single-index semiparametric specification as in Section 3.2. The covariates in the NSW study are, in addition to age, educ = years of schooling; black = 1 if black, 0 otherwise; hisp = 1 if Hispanic, 0 otherwise; married = 1 if married, 0 otherwise; and ndegr = 1 if no high school degree, 0 otherwise. We specify $\mathbb{E}(Y|\mathbf{X}) = \mathbb{E}(Y|\boldsymbol{\beta}_0'\mathbf{X})$, and estimate the parameter $\boldsymbol{\beta}_0$ by the MAVE proposed in Xia et al. (2002), which allows continuous and discrete covariates. We implement the MAVE with a Gaussian kernel and a cross-validation method for choosing the bandwidth parameter. The bootstrap p-values obtained from 10,000 replications are reported in the third column of Table 3. For a better comparison with the previous results, we consider the same subsamples, divided according to age. The null hypothesis is still rejected when considering the full range of the age distribution, but the test does not reject when considering subsamples with individuals older than 18 years old. In view of the previous results, the latter is likely to be driven by a decrease in precision because of the semiparametric smoothed estimation involved.

5. CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

This article has proposed a methodology for testing one-sided conditional moment restrictions, with two distinctive features. On the one hand, the tests can be implemented under minimal requirements on the smoothness of the underlying nonparametric curves and without resorting to smooth estimation. On the other hand, the new tests can be easily computed using the efficient PAVA algorithm, already implemented in many statistical packages. We have shown how the proposed methods can be applied to accommodate composite hypotheses of different nature and multiple covariates. Finally, we have illustrated the practical usefulness of our methods with an application to evaluating treatment effects in social programs.

Our basic results can be extended to other situations of practical interest. For instance, a straightforward extension of our results consists of allowing serial-dependent observations. This has important applications in a number of settings, see, for example, tests of superior predictive ability in Hansen (2005). The extension to time series does not pose any additional difficulties, as long as the weak convergence of the process $\sqrt{n}C_n$ holds. There is, however, an extensive literature providing sufficient conditions for weak convergence of empirical processes under weak dependence, see, for example, Linton, Maasoumi, and Whang (2005) and Scaillet and Topaloglou (2010) for applications in the context of stochastic dominance testing.

In the rest of this section, we discuss extensions of the basic framework to cases where smoothing cannot be avoided. Most notably, the conditional stochastic dominance test can also be applied when the covariate observations are different in each sample by introducing covariate-matching techniques. See, for example, Hall and Turlach (1997), Hall, Huber, and Speckman (1997), Koul and Schick (1997, 2003), Cabus (1998), Neumeyer and Dette (2003), Pardo-Fernández, van Keilegom, and González-Manteiga (2007), or Srihera and Stute (2010). These techniques use smooth estimators, typically kernels. In particular, proposals by Cabus (1998) and Neumeyer and Dette (2003), designed for testing the equality of nonparametric regression curves in a two-sample context, can be accommodated into one-sided testing by applying the methodology presented in this article.

Another important extension would consist of allowing the function $m_{\boldsymbol{\theta}}$ in (6) to be indexed by an infinite-dimensional nuisance parameter $\boldsymbol{\theta}$. For instance, this is the case in the context of nonexperimental treatment effects when the propensity score q is nonparametrically specified. When $\boldsymbol{\theta}_0$ is a nonparametric function estimated by kernels, or other smoothing techniques, the corresponding $\bar{C}_{m_{\boldsymbol{\theta}_{n,n}}}$ is asymptotically equivalent to a U-process under the l.f.c. The test can also be implemented in this case by means of a multiplier bootstrap on the Hoeffding projection, along the lines suggested by Delgado and González-Manteiga (2001). A detailed analysis of these extensions is beyond the scope of this article and is deferred to future work.

APPENDIX

Before proving the main results of the article, we first introduce some notation. For a generic set $\mathcal{G}$, let $\ell^\infty(\mathcal{G})$ be the Banach space of all uniformly bounded real functions on $\mathcal{G}$ equipped with the uniform metric $\|f\|_{\mathcal{G}} \equiv \sup_{z\in\mathcal{G}} |f(z)|$. In this article, we consider convergence in distribution of empirical processes in the metric space $(\ell^\infty(\mathcal{G}), \|\cdot\|_{\mathcal{G}})$ in the sense of J. Hoffmann-Jørgensen (see, e.g., van der Vaart and Wellner 1996). For any generic Euclidean random vector $\boldsymbol{\xi}$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $\chi_{\boldsymbol{\xi}}$ denotes its state space and $P_{\boldsymbol{\xi}}$ denotes its induced probability measure with corresponding distribution function $F_{\boldsymbol{\xi}}(\cdot) = P_{\boldsymbol{\xi}}(-\infty, \cdot]$. Given iid observations $\{\boldsymbol{\xi}_i\}_{i=1}^n$ of $\boldsymbol{\xi}$, $\mathbb{P}_{\boldsymbol{\xi}n}$ denotes the empirical measure, which assigns a mass n^{-1} to each observation, that is, $\mathbb{P}_{\boldsymbol{\xi}n} f \equiv n^{-1}\sum_{i=1}^n f(\boldsymbol{\xi}_i)$. Let $F_{\boldsymbol{\xi}n}(\cdot) \equiv \mathbb{P}_{\boldsymbol{\xi}n}(-\infty, \cdot]$ be the corresponding empirical cdf. Likewise, the expectation is denoted by $P_{\boldsymbol{\xi}} f = \int f\, dP_{\boldsymbol{\xi}}$. The empirical process evaluated at f is $\mathbb{G}_{\boldsymbol{\xi}n} f$ with $\mathbb{G}_{\boldsymbol{\xi}n} \equiv \sqrt{n}(\mathbb{P}_{\boldsymbol{\xi}n} - P_{\boldsymbol{\xi}})$. Let $\|\cdot\|_{2,P}$ be the $L_2(P)$ norm, that is, $\|f\|_{2,P}^2 = \int f^2 dP$. When P is clear from the context, we simply write $\|\cdot\|_2 \equiv \|\cdot\|_{2,P}$. Let $|\cdot|$ denote the Euclidean norm, that is, $|A|^2 = A^\top A$. For a measurable class of functions $\mathcal{G}$ from $\mathcal{X}_{\mathbf{Z}}$ to $\mathbb{R}$, let $\|\cdot\|$ be a generic pseudo-norm on $\mathcal{G}$, that is, a norm except for the property that $\|f\| = 0$ does not necessarily imply that $f \equiv 0$. Let $N(\varepsilon, \mathcal{G}, \|\cdot\|)$ be the *covering number with respect to* $\|\cdot\|$, that is, the minimal number of ε-balls with respect to $\|\cdot\|$ needed to cover $\mathcal{G}$. Given two functions $l, u \in \mathcal{G}$, the bracket $[l, u]$ is the set of functions $f \in \mathcal{G}$ such that $l \le f \le u$. An ε-bracket with respect to $\|\cdot\|$ is a bracket $[l, u]$ with $\|l - u\| \le \varepsilon$. The *covering number with bracketing* $N_{[\cdot]}(\varepsilon, \mathcal{G}, \|\cdot\|)$ is the minimal number of ε-brackets with respect to $\|\cdot\|$ needed to cover $\mathcal{G}$. Let $\mathcal{H}_B$ be the collection of all nondecreasing functions $F : \mathbb{R} \to [0, 1]$ of bounded variation less or equal than 1, and define $\Pi \equiv [-\infty, \infty] \times [0, 1]$. Finally, throughout K is a generic positive constant that may change from expression to expression.

We first state an auxiliary result from the empirical process literature. Define the generic class of measurable functions $\mathcal{G} \equiv \{\mathbf{Z} \to m(\mathbf{Z}, \boldsymbol{\theta}, h) : \boldsymbol{\theta} \in \Theta, h \in \mathcal{H}\}$, where Θ and $\mathcal{H}$ are endowed with the pseudonorms $|\cdot|_\Theta$ and $|\cdot|_{\mathcal{H}}$, respectively. The following result is part of Theorem 3 in Chen, Linton, and van Keilegom (2003).

Lemma A.1. Assume that for all $(\boldsymbol{\theta}_0, \mathbf{h}_0) \in \Theta \times \mathcal{H}$, $m(\mathbf{Z}, \boldsymbol{\theta}, \mathbf{h})$ is locally uniformly $L_2(P)$ continuous, in the sense that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}:|\boldsymbol{\theta}_0-\boldsymbol{\theta}|_\Theta<\delta,\mathbf{h}:|\mathbf{h}_0-\mathbf{h}|_{\mathcal{H}}<\delta} |m(\mathbf{Z}, \boldsymbol{\theta}, \mathbf{h}) - m(\mathbf{Z}, \boldsymbol{\theta}_0, \mathbf{h}_0)|^2\right] \le K\delta^s,$$

for all sufficiently small $\delta > 0$ and some constant $s \in (0, 2]$. Then,

$$N_{[\cdot]}(\varepsilon, \mathcal{G}, \|\cdot\|_2) \le N\left(\left(\frac{\varepsilon}{2K}\right)^{2/s}, \Theta, |\cdot|_\Theta\right) \times N\left(\left(\frac{\varepsilon}{2K}\right)^{2/s}, \mathcal{H}, |\cdot|_{\mathcal{H}}\right).$$

Proof of Theorem 1. Throughout $\mathbf{Z}_i \equiv (Y_{1i}, X_i)$, $i \ge 1$, $\bar{\mathbf{z}} \equiv (\bar{y}_1, \bar{y}_2, \bar{x}) \in \chi_{\mathbf{Z}}$. Let $\tilde{C}_n$ be defined as C_n but with F_{Xn} replaced by the true cdf F_X. Set $\Delta_n \equiv \sqrt{n}(\mathcal{T}C_n - C_n)$, and similarly define $\tilde{\Delta}_n$ with $\tilde{C}_n$ replacing C_n. The proof of Theorem 1 follows three steps: first, we prove that tests based on Δ_n and $\tilde{\Delta}_n$ are asymptotically equivalent under the l.f.c., that is,

$$\sup_{(y,u)\in\mathcal{W}_Y\times\mathcal{U}_{Xn}} \Delta_n(y,u) = \sup_{(y,u)\in\mathcal{W}_Y\times\mathcal{U}_{Xn}} \tilde{\Delta}_n(y,u) + o_{\mathbb{P}}(1). \quad (A.1)$$

Second, we prove that the supremum in $\mathcal{U}_{Xn}$ in the test statistic can be replaced by a supremum in $\mathcal{U}_X$, that is,

$$\sup_{(y,u)\in\mathcal{W}_Y\times\mathcal{U}_{Xn}} \tilde{\Delta}_n(y,u) = \sup_{(y,u)\in\mathcal{W}_Y\times\mathcal{U}_X} \tilde{\Delta}_n(y,u) + o_{\mathbb{P}}(1). \quad (A.2)$$

Finally, we prove the asymptotic behavior of the test under H_0 and H_1, not just under the l.f.c. □

We proceed with the proof of (A.1). To that end, we shall prove that $\tilde{C}_n$ and C_n are asymptotically equivalent under the l.f.c. First, define the classes of functions

$$\mathcal{G}_1 \equiv \{(\bar{y}_1,\bar{y}_2)\in\chi_{Y_1}\times\chi_{Y_2}\to\Delta_y(\bar{y}_1,\bar{y}_2) \equiv 1_{\{\bar{y}_1\le y\}} - 1_{\{\bar{y}_2\le y\}} : y\in[-\infty,\infty]\}$$

and

$$\mathcal{G}_2 \equiv \{\bar{x}\in\chi_X\to f_{u,F}(\bar{x}) \equiv (u - F(\bar{x}))\,1_{\{F(\bar{x})\le u\}} : u\in[0,1],\ F\in\mathcal{H}_B\}.$$

Define the product class $\mathcal{H}\equiv\mathcal{G}_1\cdot\mathcal{G}_2$, and notice that $\tilde{C}_n(y,u) = \mathbb{P}_{\mathbf{Z}n}h_{y,u,F_X}$, where

$$h_{y,u,F}(\bar{\mathbf{z}}) \equiv \{1_{\{\bar{y}_1\le y\}} - 1_{\{\bar{y}_2\le y\}}\}(u - F(\bar{x}))\,1_{\{F(\bar{x})\le u\}}$$

belongs to $\mathcal{H}$. We prove that $\mathcal{H}$ is $P_{\mathbf{Z}}-$ Donsker. By Example 2.10.8 in van der Vaart and Wellner (1996) and the fact that $\mathcal{G}_1$ is $P_{\mathbf{Z}}-$Donsker, it suffices to prove that $\mathcal{G}_2$ is $P_{\mathbf{Z}}-$Donsker. To that end, note that for each $(u,F)\in[0,1]\times\mathcal{H}_B$, using the triangle inequality and the simple inequality $|a_+ - b_+|^2 \le |a-b|^2$ for all $a,b\in\mathbb{R}$, where $a_+ = \max\{a,0\}$, we obtain

$$\mathbb{E}[\sup|f_{u_1,F_1}(X) - f_{u,F}(X)|^2] \le K\delta^2,$$

where the supremum is over the set $u_1\in[0,1]$ and $F_1\in\mathcal{H}_B$ such that $|u_1 - u|\le\delta$ and $\sup_{x\in\mathbb{R}}|F_1(x) - F(x)|\le\delta$, respectively. By Lemma A.1 and Theorem 19.5 in van der Vaart (1998), the class $\mathcal{G}_2$, and hence $\mathcal{H}$, is $P_{\mathbf{Z}}-$ Donsker.

Thus, by a stochastic equicontinuity argument and the Glivenko–Cantelli theorem

$$\sup_{(y,u)\in\Pi}|\mathbb{G}_{\mathbf{Z}n}h_{y,u,F_{Xn}} - \mathbb{G}_{\mathbf{Z}n}h_{y,u,F_X}| \to_p 0.$$

Furthermore, since under the l.f.c. $P_{\mathbf{Z}}h = 0$, for all $h\in\mathcal{H}$,

$$\sup_{(y,u)\in\Pi}|\mathbb{P}_{\mathbf{Z}n}h_{y,u,F_{Xn}} - \mathbb{P}_{\mathbf{Z}n}h_{y,u,F_X}| = o_{\mathbb{P}}(n^{-1/2}),$$

and hence,

$$\sup_{(y,u)\in\Pi}|C_n(y,u) - \tilde{C}_n(y,u)| = o_{\mathbb{P}}(n^{-1/2}). \quad (A.3)$$

To prove (A.1), we must show the continuity in the metric space $(\ell^\infty(\Pi), \|\cdot\|_\Pi)$ of the functional $\varphi:\ell^\infty(\Pi)\longmapsto\mathbb{R}^+$ defined as

$$\varphi(f)\equiv\sup_{(y,u)\in\Pi}(\mathcal{T}f - f)(y,u).$$

To that end, note that Lemma 2.2 in Durot and Tocquet (2003) implies that for each $f,g\in\ell^\infty(\Pi)$,

$$\sup_{u\in[0,1]}|(\mathcal{T}f - \mathcal{T}g)(y,u)| \le \sup_{u\in[0,1]}|(f-g)(y,u)| \text{ for each } y\in\mathbb{R} \text{ fixed.}$$

Since the last inequality holds for all $y\in\mathbb{R}$, for any $f,g\in\ell^\infty(\Pi)$,

$$|\varphi(f)-\varphi(g)| \le \|\mathcal{T}f - \mathcal{T}q\|_\Pi + \|f-g\|_\Pi \le 2\|f-g\|_\Pi,$$

which shows that φ is continuous with respect to $\|\cdot\|_\Pi$. Then, (A.1) follows from (A.3) and the continuity of φ.

We now prove (A.2) under the l.f.c. We have shown above that $\mathcal{H}$ is a Donsker class, that is, $\mathbb{G}_{\mathbf{Z}n}$ converges in distribution to a $P_{\mathbf{Z}}$-bridge as a random element of $(\ell^\infty(\mathcal{H}), \|\cdot\|_{\mathcal{H}})$, which in turn implies that $\tilde{C}_n(y,u) = \mathbb{P}_{\mathbf{Z}n}h_{y,u,F_X}$, and hence C_n by (A.3), converges in distribution under the l.f.c. to a tight Gaussian process C_∞ in $\ell^\infty(\Pi)$ with zero mean and covariance function

$$K(v_1,v_2)\equiv\mathbb{E}(h_{v_1,F_X}(\mathbf{Z})h_{v_2,F_X}(\mathbf{Z})),\ v_j = (y_j,u_j),\ j=1,2. \quad (A.4)$$

In particular, these arguments prove that $\tilde{\Delta}_n$ is stochastically equicontinuous in $\ell^\infty(\Pi)$ with respect to the $\|\cdot\|_2$. Hence, from the triangle inequality, the equicontinuity of $\tilde{\Delta}_n$ and the Glivenko–Cantelli theorem it holds

$$\begin{aligned}
&\left|\sup_{(y,u)\in\mathcal{W}_Y\times\mathcal{U}_{Xn}}\tilde{\Delta}_n(y,u) - \sup_{(y,u)\in\mathcal{W}_Y\times\mathcal{U}_X}\tilde{\Delta}_n(y,u)\right| \\
&= \left|\sup_{(y,x)\in\mathcal{W}_Y\times\mathcal{W}_X}\tilde{\Delta}_n(y,F_{Xn}(x)) - \sup_{(y,x)\in\mathcal{W}_Y\times\mathcal{W}_X}\tilde{\Delta}_n(y,F_X(x))\right| \\
&\le \sup_{(y,x)\in\mathcal{W}_Y\times\mathcal{W}_X}\left|\tilde{\Delta}_n(y,F_{Xn}(x)) - \tilde{\Delta}_n(y,F_X(x))\right| \\
&\le \sup_{y\in\mathcal{W}_Y,\ |u-u'|\le\delta_n}\left|\tilde{\Delta}_n(y,u) - \tilde{\Delta}_n(y,u')\right| \\
&= o_{\mathbb{P}}(1),
\end{aligned}$$

where $\delta_n\equiv\sup_{x\in\mathcal{W}_X}|F_{Xn}(x) - F_X(x)|$.

Hence, by (A.1) and (A.2) and the continuous mapping theorem, η_n converges in distribution under the l.f.c. to

$$\varphi(C_\infty)\equiv\sup_{(y,u)\in\mathcal{W}_Y\times\mathcal{U}_X}(\mathcal{T}C_\infty - C_\infty)(y,u).$$

We now study the behavior of the test, not just under the l.f.c., but under H_0 and the alternative hypothesis. To that end, we define $G_n\equiv C_n - C$. Then, by definition of the l.c.m., the function $\mathcal{T}G_n(y,\cdot) + C(y,\cdot)$ is above $C_n(y,\cdot)$ and is concave in $u\in\mathcal{U}_X$ under H_0, since both $\mathcal{T}G_n(y,\cdot)$ and $C(y,\cdot)$ are concave. Hence, $\mathcal{T}G_n + C$ is uniformly above $\mathcal{T}C_n$. Thus, under H_0,

$$\sqrt{n}(\mathcal{T}C_n - C_n)\le\sqrt{n}(\mathcal{T}G_n - G_n). \quad (A.5)$$

Under the l.f.c., $C(y,u)\equiv 0$, and hence $G_n = C_n$, so Equation (A.5) becomes an equality.

Now, the multiplier functional limit theorem (Theorem 2.9.6 in van der Vaart and Wellner 1996) and the continuous mapping theorem imply that, for all $x\ge 0$,

$$\mathbb{P}_n^*(\eta_n^* > x)\to_{\text{a.s.}} 1 - F_\varphi(x),$$

where F_φ is the cdf of $\|\sqrt{n}(\mathcal{T}G_\infty - G_\infty)\|_{\mathcal{W}_Y\times\mathcal{U}_X}$, with G_∞ being a tight Gaussian process in $\ell^\infty(\mathcal{W}_Y \times \mathcal{U}_X)$ with zero mean and covariance function (A.4). Being the cdf of a continuous mapping of a Gaussian process, F_φ is continuous, see Lifshits (1982). Hence, by Equation (A.5), under H_0,

$$\mathbb{P}(\eta_n > c^*_{n,\alpha}) \le \mathbb{P}(\|\sqrt{n}(\mathcal{T}G_n - G_n)\|_{\mathcal{W}_Y\times\mathcal{U}_X} > c^*_{n,\alpha}) = \alpha + o(1),$$

with equality under the l.f.c. Under the alternative H_1, it can be easily shown that η_n diverges to infinity, and because $c^*_{n,\alpha} = O(1)$ a.s.,

$$\mathbb{P}(\eta_n > c^*_{n,\alpha}) \to 1.$$

This completes the proof of Theorem 1.

Proof of Theorem 2. Applying a classical mean value theorem argument, uniformly in $u \in [0,1]$,

$$\bar{C}_{m_{\theta_n},n}(u) = \bar{C}_{m_{\theta_0},n}(u) + \bar{C}_{\dot{\mathbf{m}}_{\bar{\theta}_n},n}(u)'(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0), \qquad \text{(A.6)}$$

where $\bar{\boldsymbol{\theta}}_n$ is an intermediate point that satisfies $|\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \le |\boldsymbol{\theta}_n - \boldsymbol{\theta}_0|$ a.s. Define the class of functions on $\mathcal{X}_{\mathbf{Z}}$

$$\mathcal{H}_1 \equiv \{\mathbf{z} \to \dot{\mathbf{m}}_{\boldsymbol{\theta}}(\mathbf{z})(u - F(x)) \times 1_{\{F(x)\le u\}} : u \in [0,1],\ F \in \mathcal{H}_B, \boldsymbol{\theta} \in \Theta_0\}.$$

By Examples 19.7 and 19.11 in van der Vaart (1998) and by Problem 8 in van der Vaart and Wellner (1996, pg. 204), $\mathcal{H}_1$ is a Glivenko-Cantelli class under Assumption S. Thus, by Assumption E and the classical Glivenko-Cantelli theorem, uniformly in $u \in [0,1]$,

$$\bar{C}_{\dot{\mathbf{m}}_{\theta_n},n}(u) = \bar{C}_{\dot{\mathbf{m}}_{\theta}}(u) + o_{\mathbb{P}}(1). \qquad \text{(A.7)}$$

Next, define the class of functions

$$\mathcal{H}_2 \equiv \{\mathbf{z} \to q_{u,F}(\mathbf{z}) \equiv m_{\boldsymbol{\theta}_0}(\mathbf{z})(u - F(x)) \times 1_{\{F(x)\le u\}} : u \in [0,1],\ F \in \mathcal{H}_B\}.$$

Note that for all $u \in [0,1]$ and $F \in \mathcal{H}_B$,

$$E[\sup |q_{u_1,F_1}(\mathbf{Z}) - q_{u,F}(\mathbf{Z})|^2] \le K\delta^2,$$

where the supremum is over the set $u_1 \in [0,1]$ and $F_1 \in \mathcal{H}_B$ such that $|u_1 - u| \le \delta$ and $\sup_{x\in\mathbb{R}} |F_1(x) - F(x)| \le \delta$, respectively. By Lemma A1 and Theorem 19.5 in van der Vaart (1998), the class $\mathcal{H}_2$ is $P_{\mathbf{Z}}$−Donsker. Hence, by the classical Glivenko-Cantelli theorem

$$\sup_{u\in[0,1]} |\mathbb{G}_{\mathbf{Z}n} q_{u,F_{Xn}} - \mathbb{G}_{\mathbf{Z}n} q_{u,F_X}| \to_p 0.$$

Furthermore, since under the l.f.c. $P_{\mathbf{Z}}q = 0$, for all $q \in \mathcal{H}_2$,

$$\sup_{u\in[0,1]} |\bar{C}_{m_{\theta_0},n}(u) - \tilde{C}_{m_{\theta_0},n}(u)| = o_{\mathbb{P}}(n^{-1/2}), \qquad \text{(A.8)}$$

where $\tilde{C}_{m_{\theta_0},n}$ is defined as $\bar{C}_{m_{\theta_0},n}$ but with F_{Xn} replaced by the true *cdf* F_X. Then, (A.6), (A.7) and (A.8) yield (A.9) under the l.f.c.

We now prove the validity of the bootstrap approximation. Using the mean value theorem, we write

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n m_{\boldsymbol{\theta}_n}(\mathbf{Z}_i)(u - F_{Xn}(X_i))1_{\{F_{Xn}(X_i)\le u\}}V_i \\ &= \frac{1}{n}\sum_{i=1}^n m_{\boldsymbol{\theta}_0}(\mathbf{Z}_i)(u - F_{Xn}(X_i))1_{\{F_{Xn}(X_i)\le u\}}V_i \\ &\quad + (\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)'\frac{1}{n}\sum_{i=1}^n \dot{\mathbf{m}}_{\bar{\boldsymbol{\theta}}_n}(\mathbf{Z}_i)(u - F_{Xn}(X_i))1_{\{F_{Xn}(X_i)\le u\}}V_i \\ &\equiv I_{1n}(u) + I_{2n}(u), \end{aligned} \qquad \text{(A.9)}$$

where $\bar{\boldsymbol{\theta}}_n$ is an intermediate point that satisfies $|\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \le |\boldsymbol{\theta}_n - \boldsymbol{\theta}_0|$ a.s.

Noticing that the class of real-valued measurable functions on $\mathcal{X}_{\mathbf{Z}} \times \mathcal{X}_V$

$$\mathcal{H}_{1,*} \equiv \{(\mathbf{z}, v) \to \dot{\mathbf{m}}_{\boldsymbol{\theta}}(\mathbf{z})(u - F(x)) \times 1_{\{F(x)\le u\}}v : u \in [0,1],\ F \in \mathcal{H}_B, \boldsymbol{\theta} \in \Theta_0\},$$

is a Glivenko-Cantelli class, and using Assumption E, one concludes that $I_{2n}(u) = o_{\mathbb{P}^*_n}(n^{-1/2})$ a.s., uniformly in $u \in [0,1]$. Next, define the class on $\mathcal{X}_{\mathbf{Z}} \times \mathcal{X}_V$,

$$\mathcal{H}_{2,*} \equiv \{(\mathbf{z}, v) \to h_{u,F}(\mathbf{z}, v) \equiv m_{\boldsymbol{\theta}_0}(\mathbf{z})(u - F(x)) \times 1_{\{F(x)\le u\}}v : u \in [0,1],\ F \in \mathcal{H}_B\}.$$

The class $\mathcal{H}_{2,*}$ is $P_{(\mathbf{Z},V)}$−Donsker, since $\mathcal{H}_2$ is $P_{\mathbf{Z}}$−Donsker, see Theorem 2.9.2 in van der Vaart and Wellner (1996). Then, since $\mathbb{P}^*_n h = 0$ a.s., for all $h \in \mathcal{H}_{2,*}$,

$$I_{1n}(u) = \frac{1}{n}\sum_{i=1}^n m_{\boldsymbol{\theta}_0}(\mathbf{Z}_i)(u - F_X(X_i))1_{\{F_X(X_i)\le u\}}V_i + o_{\mathbb{P}^*_n}\left(n^{-1/2}\right), \text{ a.s.} \qquad \text{(A.10)}$$

On the other hand, by Assumption E and a strong uniform law of large numbers,

$$\begin{aligned} &\mathrm{var}^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{\mathbf{l}_{\boldsymbol{\theta}_n}(\mathbf{Z}_i, X_i) - \mathbf{l}_{\boldsymbol{\theta}_0}(\mathbf{Z}_i, X_i)\right\}V_i\right) \\ &= \frac{1}{n}\sum_{i=1}^n (\mathbf{l}_{\boldsymbol{\theta}_n}(\mathbf{Z}_i, X_i) - \mathbf{l}_{\boldsymbol{\theta}_0}(\mathbf{Z}_i, X_i))(\mathbf{l}_{\boldsymbol{\theta}_n}(\mathbf{Z}_i, X_i) - \mathbf{l}_{\boldsymbol{\theta}_0}(\mathbf{Z}_i, X_i))' \\ &= o(1), \text{ a.s.} \end{aligned}$$

Thus,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{l}_{\boldsymbol{\theta}_n}(\mathbf{Z}_i, X_i)V_i = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{l}_{\boldsymbol{\theta}_0}(\mathbf{Z}_i, X_i)V_i + o_{\mathbb{P}^*_n}(1), \text{ a.s.} \qquad \text{(A.11)}$$

The expansions (A.9), (A.10) and (A.11), and the multiplier central limit theorem, see Theorem 2.9.2 in van der Vaart and Wellner (1996), imply that $\bar{C}^*_{m_{\theta_n},n}$ converges weakly (almost surely) to the same weak limit as $\bar{C}_{m_{\theta_n},n}$ in $(\ell^\infty(\mathcal{U}_X), \|\cdot\|_{\mathcal{U}_X})$. From this point, the rest of the proof follows the reasoning of Theorem 1 in a routine fashion. Details are omitted. □

Proof of Theorem 3. The proof follows the same steps as that of Theorem 1. Hence, to save space, we only discuss here the differences. Let $\hat{F}_{\mathbf{X}n}$ denote the empirical cdf of $\{\boldsymbol{\beta}_n'\mathbf{X}_i\}_{i=1}^n$ and let $\hat{C}_n$ be defined as $\tilde{C}_n$ but with $\hat{F}_{\mathbf{X}n}$ replacing the true cdf $F_{\boldsymbol{\beta}_0'\mathbf{X}}$. Set $\hat{\Delta}_n \equiv \sqrt{n}(\mathcal{T}\hat{C}_n - \hat{C}_n)$. Define the class of functions

$$\mathcal{G}_3 \equiv \{\bar{\mathbf{x}} \in \chi_{\mathbf{X}} \to f_{u,F,\boldsymbol{\beta}}(\bar{\mathbf{x}}) \equiv (u - F(\boldsymbol{\beta}'\bar{\mathbf{x}})) \times 1_{\{F(\boldsymbol{\beta}'\bar{\mathbf{x}})\leq u\}} : u \in [0,1],\ F \in \mathcal{L}_B, \boldsymbol{\beta} \in \Theta\},$$

where $\mathcal{L}_B$ is the set of Liptschitz functions in $\mathcal{H}_B$, that is, for all z_1 and z in $\mathbb{R}$, with $z_1 \geq z$,

$$F(z_1) - F(z) \leq K[z_1 - z].$$

We prove that $\mathcal{G}_3$ is $P_{\mathbf{Z}}-$Donsker. To that end, note that for each $(u, F) \in [0,1] \times \mathcal{H}_B$, using the triangle inequality and the simple inequality $|a_+ - b_+|^2 \leq |a-b|^2$ for all $a, b \in \mathbb{R}$, where $a_+ = \max\{a, 0\}$, we obtain

$$\begin{aligned}\mathbb{E}\big[\sup|f_{u_1,F_1,\boldsymbol{\beta}_1}(\mathbf{X}) - f_{u,F,\boldsymbol{\beta}}(\mathbf{X})|^2\big] &\leq 2\mathbb{E}\big[\sup|F_1(\boldsymbol{\beta}_1'\mathbf{X}) - F_2(\boldsymbol{\beta}_1'\mathbf{X})|^2\big] \\ &\quad + 2\mathbb{E}\big[\sup|F_2(\boldsymbol{\beta}_1'\mathbf{X}) - F_2(\boldsymbol{\beta}_2'\mathbf{X})|^2\big] \\ &\leq K(1 + \mathbb{E}\big[|\mathbf{X}|^2\big])\delta^2,\end{aligned}$$

where the supremum is over the set $u_1 \in [0,1]$, $F_1 \in \mathcal{L}_B$, and $\boldsymbol{\beta}_1 \in \Theta$ such that $|u_1 - u| \leq \delta$, $\sup_{x\in\mathbb{R}}|F_1(x) - F(x)| \leq \delta$, and $|\boldsymbol{\beta}_1 - \boldsymbol{\beta}| \leq \delta$, respectively. By Lemma A.1, the class $\mathcal{G}_3$, and hence $\mathcal{H} \equiv \mathcal{G}_1 \cdot \mathcal{G}_3$, is $P_{\mathbf{Z}}-$Donsker.

We now prove that $\mathbb{P}(\hat{F}_{\mathbf{X}n} \in \mathcal{L}_B) \to 1$ as $n \to \infty$. First, notice that $\hat{F}_{\mathbf{X}n} \in \mathcal{H}_B$ for each $n \geq 1$. Also, by the Chebyshev inequality, for all $z_1 \geq z$ and any constant $K_1 > 0$,

$$\begin{aligned}\mathbb{P}\left(\hat{F}_{\mathbf{X}n}(z_1) - \hat{F}_{\mathbf{X}n}(z) > K_1[z_1 - z]\right) &\leq K_1^{-1}[z_1 - z]^{-1}\mathbb{E}\left[\hat{F}_{\mathbf{X}n}(z_1) - \hat{F}_{\mathbf{X}n}(z)\right] \\ &\leq K_1^{-1}[z_1 - z]^{-1}\mathbb{E}\left[\hat{s}(z_1, z)\right],\end{aligned}$$

where $\hat{s}(z_1, z) \equiv 1_{\{\boldsymbol{\beta}_n'\mathbf{X}\leq z_1\}} - 1_{\{\boldsymbol{\beta}_n'\mathbf{X}\leq z\}}$. By Assumption M, and defining $\boldsymbol{\beta}_n =: (1, \boldsymbol{\theta}_n')'$,

$$\begin{aligned}\mathbb{E}[\hat{s}_1(z_1, z)] &= \mathbb{E}\big[1_{\{z-\boldsymbol{\theta}_n'\mathbf{X}_{-1}\leq X_1\leq z_1-\boldsymbol{\theta}_n'\mathbf{X}_{-1}\}}\big] \\ &= \mathbb{E}\big[F_{X_1|\mathbf{X}_{-1}}\left(z_1 - \boldsymbol{\theta}_n'\mathbf{X}_{-1}, \mathbf{X}_{-1}\right) \\ &\qquad - F_{X_1|\mathbf{X}_{-1}}\left(z - \boldsymbol{\theta}_n'\mathbf{X}_{-1}, \mathbf{X}_{-1}\right)\big] \\ &\leq K[z_1 - z].\end{aligned}$$

Choosing K_1 sufficiently large, we obtained the desired result.

Similarly, it can be shown that $\hat{F}_{\mathbf{X}n}$ is uniformly consistent for $F_{\boldsymbol{\beta}_0'\mathbf{X}}$, since the class $\{1_{\{\boldsymbol{\beta}'\bar{\mathbf{x}}\leq z\}} : z \in \mathbb{R}, \boldsymbol{\beta} \in \Theta\}$ is Glivenko–Cantelli, the map $\boldsymbol{\beta} \in \Theta \to \mathbb{E}[1_{\{\boldsymbol{\beta}'\mathbf{X}\leq z\}}]$ is continuous under Assumption M and $\boldsymbol{\beta}_n$ is consistent for $\boldsymbol{\beta}_0$.

Thus, by a stochastic equicontinuity argument and the Glivenko–Cantelli theorem

$$\sup_{(y,u)\in\Pi}\left|\mathbb{G}_{Zn}h_{y,u,F_{\mathbf{X}n},\boldsymbol{\beta}_n} - \mathbb{G}_{Zn}h_{y,u,F_{\boldsymbol{\beta}_0'\mathbf{X}},\boldsymbol{\beta}_0}\right| \to_p 0,$$

where $h_{y,u,F,\boldsymbol{\beta}}(\bar{\mathbf{z}}) \equiv \{1_{\{\bar{y}_1\leq y\}} - 1_{\{\bar{y}_2\leq y\}}\}(u - F(\boldsymbol{\beta}'\bar{\mathbf{x}}))1_{\{}F(\boldsymbol{\beta}'\bar{\mathbf{x}}) \leq u\}$. From the arguments of Theorem 1, we conclude that under the l.f.c.

$$\sup_{(y,u)\in\mathcal{W}_Y\times\mathcal{U}_{Xn}}\hat{\Delta}_n(y,u) = \sup_{(y,u)\in\mathcal{W}_Y\times\mathcal{U}_{Xn}}\tilde{\Delta}_n(y,u) + o_{\mathbb{P}}(1).$$

From here, the same arguments of Theorem 1 lead to

$$\hat{\eta}_n(\boldsymbol{\beta}_n) = \hat{\eta}_n(\boldsymbol{\beta}_0) + o_{\mathbb{P}}(1),$$

under the l.f.c. □

ACKNOWLEDGMENTS

The authors thank Editor Keisuke Hirano, an associate editor, and two referees for constructive comments that have led to substantial improvement in the presentation. Research is funded by Spanish Plan Nacional de I+D+I grant number SEJ2007-62908.

[Received June 2011. Revised June 2012.]

REFERENCES

Anderson, G. J. (1996), "Nonparametric Tests of Stochastic Dominance in Income Distributions," *Econometrica*, 64, 1183–1193. [16]

Andrews, D. W. K., and Shi, X. (2010), "Inference Based on Conditional Moment Inequalities," Cowless Comission Discussion Paper 1761, Yale University. [16]

Barlow, R. E., Bartholomew, D. J., Bremmer, J. M., and Brunk, H. D. (1972), *Statistical Inference Under Order Conditions*, New York: Wiley. [16,18]

Barrett, G., and Donald, S. (2003), "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, 71–104. [16]

Bril, G., Dykstra, R., Pillers, C., and Robertson, T. (1984), "Algorithm AS 206: Isotonic Regression in Two Independent Variables," *Journal of the Royal Statistical Society,* Series C, 33, 352–357. [18]

Bücher, A., and Dette, H. (2010), "A Note on Bootstrap Approximations for the Empirical Copula Process," *Statistics and Probability Letters*, 80, 1925–1932. [18]

Cabus, P. (1998), "Un Test de Type Kolmogorov–Smirnov Dans Le Cadre de Comparaison de Fonctions de Régression," *Comptes Rendus de l'Académie des Sciences,* Series I, 327, 939–942. [24]

Chen, X., Linton, O. B., and van Keilegom, I. (2003), "Estimation of Semiparametric Models When the Criterion Function is not Smooth," *Econometrica*, 71, 1591–1608. [24]

Cran, G. W. (1980), "Algorithm AS 149: Amalgamation of Means in the Case of Simple Ordering," *Journal of the Royal Statistical Society,* Series C, 29, 209–211. [18]

Davidson, R., and Duclos, J. Y. (2000), "Statistical Inference for Stochastic Dominance and Measurement for the Poverty and Inequality," *Econometrica*, 68, 1435–1464. [16]

Dehejia, R. H., and Wahba, S. (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062. [22]

de Leeuw, J., Hornik, K., and Mair, P. (2009), "Isotone Optimization in R: Pool-Adjacent-Violators (PAVA) and Active Set Methods," *Journal of Statistical Software*, 32, 1–24. [18]

Delgado, M. A. (1993), "Testing for the Equality of Nonparametric Regression Curves," *Statistics and Probability Letters*, 17, 199–204. [16]

Delgado, M. A., and Escanciano, J. C. (2012), "Distribution-Free Tests of Stochastic Monotonicity," *Journal of Econometrics*, 170, 68–75. [16]

Delgado, M. A., Fariñas, J. C., and Ruano, S. (2002), "Firm Productivity and Export Markets: A Non-Parametric Approach," *Journal of International Economics*, 57, 397–422. [19]

Delgado, M. A., and González-Manteiga, W. (2001), "Significance Testing in Nonparametric Regression Based on the Bootstrap," *The Annals of Statistics*, 29, 1469–1507. [18,24]

Donald, S., and Hsu, Y.-C. (2011), "Estimation and Inference of Distribution Functions and Quantile Functions in Treatment Effect Models," *Preprint*. [19]

Durot, C. (2003), "A Kolmogorov-Type Test for Monotonicity of Regression," *Statistics and Probability Letters*, 63, 425–433. [16]

Durot, C., and Tocquet, A.-S. (2003), "On the Distance Between the Empirical Process and its Concave Majorant in a Monotone Regression Framework," *Annals of the Institute Henri Poincare Probability and Statistics*, 39, 217–240. [25]

Dykstra, R. L., and Robertson, T. (1982), "Order Restricted Statistical Tests on Multinomial and Poisson Parameters: The Starshaped Restriction," *The Annals of Statistics*, 10, 1246–1252. [16]

——— (1983), "On Testing Monotone Tendencies," *Journal of the American Statistical Association*, 78, 342–350. [16]

Escanciano, J. C. (2006), "A Consistent Diagnostic Test for Regression Models Using Projections," *Econometric Theory*, 22, 1030–1051. [20]

Ferreira, E., and Stute, W. (2004), "Testing for Differences Between Conditional Means in a Time Series Context," *Journal of the American Statistical Association*, 99, 169–174. [16]

Friedman, J. H., and Tukey, J. W. (1974), "A Projection Pursuit Algorithm for Exploratory Data Analysis," *IEEE Transactions on Computers*, C-23, 881–890. [20]

Hall, P., and Hart, J. D. (1990), "Bootstrap Test for Difference Between Means in Nonparametric Regression," *Journal of the American Statistical Association*, 85, 1039–1049. [16]

Hall, P., Huber, C., and Speckman, P. L. (1997), "Covariate-Matched One-Sided Tests for the Difference Between Functional Means," *Journal of the American Statistical Association*, 92, 1074–1083. [24]

Hall, P., and Turlach, B. A. (1997), "Enhancing Convolution and Interpolation Methods for Nonparametric Regression," *Biometrika*, 84, 779–790. [24]

Hall, P., and Yatchew, A. (2005), "Unified Approach to Testing Functional Hypotheses in Semiparametric Contexts," *Journal of Econometrics*, 127, 225–252. [16]

Hansen, P. (2005), "A Test for Superior Predictive Ability," *Journal of Business and Economics Statistics*, 23, 365–380. [24]

Härdle, W., and Marron, S. (1990), "Semiparametric Comparison of Regression Curves," *The Annals of Statistics*, 18, 63–89. [16]

Hsu, Y.-C. (2011), Consistent Tests for Conditional Treatment Effects, Preprint, University of Missouri at Columbia. [19]

Khan, S., and Tamer, E. (2009), "Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities," *Journal of Econometrics*, 152, 104–119. [16]

Kim, K. (2008), "Set Estimation and Inference with Models Characterized by Conditional Moment Restrictions," Preprint, University of Minnesota. [16,20]

King, E. C., Hart, J. D., and Wehrly, T. E. (1991), "Testing the Equality of Two Regression Curves Using Linear Smoothers," *Statistics and Probability Letters*, 12, 239–247. [16]

Kodde, D. A., and Palm, F. C. (1986), "Wald Criteria for Jointly Testing Equality and Inequality Restrictions," *Econometrica*, 54, 1243–1248. [16]

Koul, H. L., and Schick, A. (1997), "Testing for the Equality of Two Nonparametric Regression Curves," *Journal of Statistical Planning and Inference*, 65, 293–314. [24]

——— (2003), "Testing for Superiority Among Two Regression Curves," *Journal of Statistical Planning and Inference*, 117, 15–33. [24]

LaLonde, R. J. (1986), "Evaluating the Econometric Evaluations of Training Programs With Experimental Data," *American Economic Review*, 76, 604–620. [17,22]

Lee, S., and Whang, Y.-J. (2009), "Nonparametric Tests of Conditional Treatment Effects," Preprint, Cowles Foundation Discussion Paper No. 1740. [16,19,23]

Lifshits, M. A. (1982), "On the Absolute Continuity of Distributions of Functionals of Random Processes," *Theory of Probability and Its Applications*, 27, 600–607. [26]

Linton, O., Maasoumi, E., and Whang, Y.-J. (2005), "Consistent Testing for Stochastic Dominance Under General Sampling Schemes," *Review of Economic Studies*, 72, 735–765. [16,24]

Linton, O., Song, K., and Whang, Y.-J. (2010), "An Improved Bootstrap Test of Stochastic Dominance," *Journal of Econometrics*, 154, 186–202. [16]

Mammen, E. (1993), "Bootstrap and Wild Bootstrap for High-Dimensional Linear Model," *The Annals of Statistics*, 21, 225–285. [21]

McFadden, D. (1989), "Testing for Stochastic Dominance," in *Studies in the Economics of Uncertainty: In Honor of Josef Hadar*, eds. T. B. Fomby and T. K. Seo, New York: Springer. [16,19]

Neumeyer, N., and Dette, H. (2003), "Nonparametric Comparison of Regression Curves: An Empirical Process Approach," *The Annals of Statistics*, 31, 880–920. [21,24]

Pardo-Fernández, J. C., van Keilegom, I., and González-Manteiga, W. (2007), "Testing for the Equality of k Regression Curves," *Statistica Sinica*, 17, 1115–1137. [24]

Rémillard, B., and Scaillet, O. (2009), "Testing for Equality Between Two Copulas," *Journal of Multivariate Analysis*, 100, 377–386. [18]

Robertson, T., Wright, F. T., and Dykstra, F. T. (1988), *Order Restricted Inference*, New York: Wiley. [16]

Scaillet, O. (2005), "A Kolmogorov–Smirnov Type Test for Positive Quadrant Dependence," *Canadian Journal of Statistics*, 33, 415–427. [18]

Scaillet, O., and Topaloglou, N. (2010), "Testing for Stochastic Dominance Efficiency," *Journal of Business and Economic Statistics*, 28, 169–180. [16,24]

Smirnov, N. V. (1939), "An Estimate of Divergence Between Empirical Curves of a Distribution in two Independent Samples," *Bulletin Mathématique del Úniversité de Moscou*, 2, 3–14 (in Russian). [16]

Srihera, R., and Stute, W. (2010), "Nonparametric Comparison of Regression Functions," *Journal of Multivariate Analysis*, 9, 2039–2059. [24]

Stute, W., and Zhu, L. (2005), "Nonparametric Tests for Single Index Models," *The Annals of Statistics*, 33, 1048–1083. [20]

van der Vaart, A. W. (1998), Asymptotic Statistics, Cambridge: Cambridge University Press. [25]

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer. [18,24,25]

Wolak, F. A. (1989), "Testing Inequality Constraints in Linear Econometric Models," *Journal of Econometrics*, 41, 205–235. [16]

Xia, Y., Tong, H., Li, W. K., and Zhu, L. (2002), "An Adaptive Estimation of Dimension Reduction Space," *Journal of the Royal Statistical Society,* Series B, 64, 363–410. [21,23]