Distribution-free tests of stochastic monotonicity ☆

Miguel A. Delgado [a,*], Juan Carlos Escanciano [b]

[a] *Departamento de Economía, Universidad Carlos III de Madrid, 28903 Getafe (Madrid), Spain*
[b] *Economics Department, Indiana University, Bloomington, IN, USA*

A B S T R A C T

This article proposes a nonparametric test of monotonicity for conditional distributions and its moments. Unlike previous proposals, our method does not require smooth estimation of the derivatives of nonparametric curves. Distinguishing features of our approach are that critical values are pivotal under the null in finite samples and that the test is invariant to any monotonic continuous transformation of the explanatory variable. The test statistic is the sup-norm of the difference between the empirical copula function and its least concave majorant with respect to the explanatory variable coordinate. The resulting test is able to detect local alternatives converging to the null at the parametric rate $n^{-1/2}$, with n the sample size. The finite sample performance of the test is examined by means of a Monte Carlo experiment and an application to testing intergenerational income mobility.

JEL classification:
C14
C15

Keywords:
Stochastic monotonicity
Conditional moments
Least concave majorant
Copula process
Distribution-free in finite samples
Tests invariant to monotone transforms

1. Introduction

Let (Y, X) be a bivariate random vector taking values in $\mathcal{Y}\times\mathcal{X}\subseteq\mathbb{R}^2$ and with induced joint distribution

$$F(y,x)=\int_{-\infty}^{x}F_{Y|X}(y|\bar{x})\,F_X(d\bar{x}),\quad (y,x)\in\mathcal{Y}\times\mathcal{X}, \tag{1}$$

where $F_{Y|X}$ is the conditional distribution function of Y given X and, henceforth, F_ξ denotes the marginal cumulative distribution function (cdf) of the generic random variable (r.v.) ξ. This article proposes a nonparametric test for the monotonicity of $F_{Y|X}$ with respect to the covariate X. That is, the null hypothesis is

$$H_0: F_{Y|X}(y|\cdot)\in\mathcal{M}\quad\text{for each } y\in\mathcal{Y}, \tag{2}$$

where $\mathcal{M}$ is the set of monotonically non-increasing functions with support $\mathcal{X}$, *i.e.*,

$$\mathcal{M}=\left\{m:\mathcal{X}\subseteq\mathbb{R}\to\mathbb{R}\text{ s.t. } m\left(x'\right)\geq m\left(x''\right)\text{ for } x'\leq x''\right\}.$$

We consider omnibus tests where the alternative hypothesis, H_1, is the negation of H_0. The discussion and results below obviously apply to the monotonically non-decreasing case *mutatis mutandis*. This testing problem has been recently addressed by Lee et al. (2009), LLW henceforth, which generalizes the test of monotonicity for regression functions proposed by Ghosal et al. (2000), GSV henceforth.

Testing monotonicity is interesting, first of all, because estimators of nonparametric monotonic curves can be obtained without imposing smoothness restrictions. See e.g. Brunk (1958) and the monograph by Barlow et al. (1972). The efficiency of these isotonic estimators can be improved when it is additionally known that the nonparametric curve is smooth. See e.g. Mammen (1991) and Mukerjee (1988).

The null hypothesis H_0 states a stochastic dominance assumption on subpopulations defined by means of the values taken by the covariate X. For instance, when $Y = Y(t+1)$ and $X = Y(t)$, for a Markov process $\{Y(t)\}_{t\in\mathbb{Z}}$, this generalizes the usual stochastic dominance concept for the transition probabilities of Markov chains to a continuous state space, see e.g. Kadi et al. (2009) for a discussion. Stochastic monotonicity plays a crucial role in stochastic dynamic programming in order to ensure the uniqueness of the equilibrium solution. See Chapters 9 and 12 of Lucas and Stokey (1989). This property is often assumed when modeling industrial economics dynamics. See e.g. Ericson and Pakes (1995), Pakes (1986) or Olley and Pakes (1996). Monotonicity is also an important identification assumption in many nonparametric and semiparametric settings, see Matzkin (1994) for a survey and Aguirregabiria (2010), Banerjee et al. (2009), Lewbel and Linton (2007), and Tanaka (2008) for some recent applications. The monotonicity of the intergenerational transition function is also worth testing in the analysis of intergenerational mobility; i.e. having a parent from a high social-economic status is never worse than having one with a lower status. This testing problem has been considered by LLW and Dardanoni et al. (2012) using different data sets. Many theories in finance also imply monotonic patterns in expected returns and other financial variables, see e.g. Boudoukh et al. (1999) and Richardson et al. (1992). Recently, Patton and Timmermann (2009) have proposed tests of monotonicity and have applied them to test whether expected returns are monotonically decreasing or monotonically increasing in securities' risk or liquidity characteristics. The tests presented in this article can be used to extend their results to a continuous covariate.

The null hypothesis H_0 implies that for any non-increasing function in the second argument $\gamma : \mathcal{Y} \times \mathcal{X} \to [0, \infty)$,

$$H_0^\gamma : \mathbb{E}\left(\gamma\left(Y, X\right) \middle| X = \cdot\right) = \int_{\mathcal{Y}} \gamma\left(y, \cdot\right) F_{Y|X}\left(dy|\cdot\right) \in \mathcal{M}. \quad (3)$$

The function γ could be non-parametric. In fact, H_0^γ with nonparametric γ is crucial in modeling under asymmetric information. For instance, in signaling models, the analysis is conducted by a monotonicity property; e.g. more talented workers buy higher education (Spence, 1973) or work faster (Akerlof, 1976) than their less talented competitors. Monotonicity also plays a crucial role in adverse selection; e.g. Akerlof (1970) "lemons" model, where higher prices in the used car market results in a higher average quality of the cars available. Additional examples of the role of monotonicity can be found in the literature on search, advertising and bidding. See Milgrom (1981) for discussion.

Testing H_0^γ with γ known or parametric is interesting on its own in many circumstances. Testing the monotonicity of regression curves is a natural hypothesis to test. In fact, the monotonicity of reduced form mean responses forms a basis for the identification of non-parametric structural relations. See Manski and Pepper (2000). Monotonicity of a regression function is also essential for the root-n consistent estimation of convolution density estimators in Escanciano and Jacho-Chávez (2012) and references therein. The test for H_0^γ with γ known of GSV, extended to testing H_0 by LLW, as well as the vast majority of existing monotonicity tests, rely on the assumption that the nonparametric curve is smooth enough, and the tests are based on some kind of smooth nonparametric estimator of the first derivatives. See also previous proposals by Bowman et al. (1998), Schlee (1982), or Hall and Heckman (2000). The performance of these tests depends on the satisfaction of several assumptions on the nonparametric curve whose monotonicity is tested, as well as other underlying nonparametric curves, despite the nuisance of suitably choosing some smoothing parameter.

In this article, rather than looking at the first derivative of the curve, we pay attention to its integral. To that end, we introduce the copula function

$$C\left(u, v\right) := F\left(F_Y^{-1}\left(u\right), F_X^{-1}\left(v\right)\right), \quad \left(u, v\right) \in [0, 1]^2,$$

where F_ξ^{-1} denotes the quantile function, i.e. the generalized inverse $F_\xi^{-1}(u) := \inf\{t \in \mathbb{R} : F_\xi(t) \geq u\}$, $u \in [0, 1]$, associated to the cdf F_ξ. We shall assume that F_X and F_Y are continuous. Hence, from (1) we can write

$$C\left(u, v\right) = \int_0^v F_{Y|X}\left(F_Y^{-1}\left(u\right) \middle| F_X^{-1}(\bar{v})\right) d\bar{v}, \quad \left(u, v\right) \in [0, 1]^2.$$

Let $\mathcal{C}$ be the set of concave functions on [0, 1]. The condition

$$C\left(u, \cdot\right) \in \mathcal{C} \quad \text{for each } u \in [0, 1] \quad (4)$$

is necessary and sufficient for H_0. Sufficiency is guaranteed because a concave function has non-increasing derivatives, and necessity is proved, for instance, in Apostol (1967, Theorem 2.9).

Therefore, the null hypothesis can be alternatively characterized using the least concave majorant (l.c.m) operator, $\mathcal{T}$ say, applied to the explanatory variable coordinate. That is, the l.c.m of $C\left(u, \cdot\right)$ for each $u \in [0, 1]$ fixed, $\mathcal{T}C\left(u, \cdot\right)$, is the function satisfying the following two properties: (i) $\mathcal{T}C\left(u, \cdot\right) \in \mathcal{C}$ and (ii) if there exists $h \in \mathcal{C}$ with $h \geq C\left(u, \cdot\right)$, then $h \geq \mathcal{T}C\left(u, \cdot\right)$. Henceforth, $\mathcal{T}C$ denotes the function obtained by applying the operator $\mathcal{T}$ to the function $C\left(u, \cdot\right)$ for each $u \in [0, 1]$. Thus, we can alternatively write H_0 as

$$H_0 : \mathcal{T}C = C. \quad (5)$$

Obviously, the greatest convex minorant must be used for characterizing H_0 in the monotonically non-decreasing case. The copula function C, and therefore $\mathcal{T}C$, can be estimated by its sample analog. Notice that the slope of $\mathcal{T}C$ with respect to the second coordinate is a restricted version of $F_{Y|X}$, i.e. concave with respect to X. Our approach is then related to the classical literature on inference under shape restrictions. Grenander (1956) first found that the slope of the l.c.m of the empirical distribution is the maximum likelihood estimator of a monotonic non-increasing probability density. Chernoff (1964) applied Grenander's ideas to the estimation of a mode and Prakasa Rao (1969) to the estimation of an unimodal probability density. Brunk (1958) extended this idea to estimating a monotonic (isotonic) regression function, see Barlow et al. (1972) for a monograph on isotonic regression. These ideas are behind the classical DIP test of unimodality proposed by Hartigan and Hartigan (1985). More recently, Durot (2003) has used the difference between the empirical integrated regression function and its l.c.m. for testing monotonicity of a regression curve in a fixed regressor setting with independent and identically distributed (iid) errors. The fixed regressor assumption is rather restrictive and rules out most applications of interest in economics. Moreover, a näive extension of Durot's (2003) method to stochastic regressors is not valid because the integrated regression function is not necessarily concave or convex when the regression function is monotone.

Estimates of the l.c.m. of the copula process are used in this article for testing monotonicity of the conditional cdf, only assuming continuity of the marginal distributions. Distinguishing features of our approach are that test's critical values are pivotal under the null and that the test is invariant to any monotonic continuous transformation of the explanatory variable in finite samples. The former feature is inherited from the use of the copula process, the latter should be a minimal requirement for any test of monotonicity. Our proposal permits us to relax different smoothness assumptions on the underlying nonparametric curves imposed by LLW and related tests. In particular, with our approach there is no need to estimate derivatives of nonparametric conditional curves, which requires a bandwidth choice.

The rest of the article is organized as follows. The test is discussed in Section 2. Section 3 presents an asymptotic test for H_0^γ with γ known. The results of a Monte Carlo study are reported in Section 4, together with an application of the new test to studying intergenerational income mobility. Mathematical proofs are gathered in a technical mathematical Appendix at the end of the article.

2. Testing stochastic monotonicity

Given independent copies $\mathcal{Z}_n := \{(Y_i, X_i)\}_{i=1}^n$ of (Y, X), the natural estimator of $C\left(u, v\right)$ is the empirical copula process

$$C_n\left(u, v\right) := \frac{1}{n} \sum_{i=1}^n 1_{\{F_{Yn}(Y_i) \leq u\}} 1_{\{F_{Xn}(X_i) \leq v\}}, \quad \left(u, v\right) \in [0, 1]^2, \quad (6)$$

where, for a sample $\{\xi_i\}_{i=1}^n$ of a generic r.v. ξ, $F_{\xi n}(\cdot) := n^{-1}\sum_{i=1}^n 1_{\{\xi_i \le \cdot\}}$ denotes the sample analog of F_ξ and $1_{\{A\}}$ is the indicator function of the event A, i.e. $1_{\{A\}} = 1$ if A occurs and $= 0$, otherwise. The empirical copula process

$$K_n := \sqrt{n}\,(C_n - C)$$

is a random element of the càdlàg space $D[0, 1]^2$ endowed with the J_1 Skorohod's topology constructed by Bickel and Wichura (1971) and Neuhaus (1971). Deheuvels (1981a,b) first obtained the exact law of K_n when Y and X are independent, i.e. $C(u, v) = u \cdot v$, see also Gänssler and Stute (1987).

Notice that $\mathcal{T}C_n(u, \cdot)$, taking u fixed, is the corresponding sample version of $\mathcal{T}C(u, \cdot)$. Omnibus tests of H_0 are based on the empirical process

$$T_n := \sqrt{n}\,(\mathcal{T}C_n - C_n).$$

The *least favorable case* (*l.f.c.*) under the null hypothesis, which is the case closest to the alternative, corresponds to the situation where X and Y are independent. In that case, $T_n \equiv \mathcal{T}K_n - K_n$, after taking advantage of the fact that $\mathcal{T}(C_n(u, v) - uv) = \mathcal{T}C_n(u, v) - uv$. See Appendix for a proof.

Test statistics can be some suitable functional of T_n, like other tests based on empirical processes. We propose using the uniform norm, i.e the Kolmogorov–Smirnov criteria. That is, the test statistic is

$$\tau_n := \|T_n\|_\infty, \tag{7}$$

where $\|z\|_\infty := \sup_{(u,v)\in[0,1]^2} |z(u, v)|$. The test statistic is simple to compute and does not require numerical optimization. By well-known results from the classical Kolmogorov–Smirnov tests, we compute τ_n as

$$\tau_n = \max_{1\le i\le n}\max_{1\le j\le n}\sqrt{n}\left(\mathcal{T}C_n\left(\frac{i}{n}, \frac{j}{n}\right) - C_n\left(\frac{i}{n}, \frac{j-1}{n}\right)\right), \tag{8}$$

where $C_n(i/n, 0) \equiv 0$. Hence, all that is needed in the computation of τ_n are the elements $C_n(i/n, j/n)$ and $\mathcal{T}C_n(i/n, j/n)$. Computation of the elements $C_n(i/n, j/n)$ is straightforward, and it can be done recursively once the covariates are ordered. To compute $\mathcal{T}C_n(i/n, \cdot)$ for each $i = 1, \ldots, n$, one can use the Pool–Adjacent–Violators (PAV) algorithm described in Barlow et al. (1972, p. 13). See e.g. Bril et al. (1984) and Cran (1980) for FORTRAN implementations and de Leeuw et al. (2009) for R routines.

The results in Deheuvels (1981a,b) directly imply that the finite sample distribution of T_n is pivotal under the *l.f.c.* and can be tabulated. Thus, a finite sample test at the α − *level* of significance rejects H_0 if $\tau_n > \tau_{n\alpha}$, where $\tau_{n\alpha} := \inf\{t \in [0, \infty) : \mathbb{P}(\tau_n \le t | l.f.c.) \ge 1 - \alpha\}$ is the $(1 - \alpha)$ − *quantile* of τ_n in the *l.f.c.* The next theorem establishes that the resulting test is consistent and its Type I error never exceeds the significance level under the following minimal assumption.

Assumption A1. $\{(Y_i, X_i)\}_{i=1}^n$ are iid distributed as (Y, X), with continuous marginal distributions F_X and F_Y.

Theorem 1. *Let Assumption A1 hold. Then,*

(a) $\mathbb{P}(\tau_n > \tau_{n\alpha}) \le \alpha$ *under* H_0.

If in addition, F is continuous, then

(b) $\lim_{n\to\infty} \mathbb{P}(\tau_n > \tau_{n\alpha}) = 1$ *under* H_1.

Part (a) follows applying properties of $\mathcal{T}$ detailed in Appendix. Part (b) follows proving that τ_n converges in distribution under the *l.f.c.* and diverges under H_1. Rüschendorf (1976), see also Deheuvels (1981a,b) and Gänssler and Stute (1987, Ch. V), establishes that, in the *l.f.c.* of independence, K_n converges in distribution to a "completely tucked" Brownian sheet, i.e. a continuous Gaussian process with mean zero and covariance function given in (13). Then, convergence in distribution of τ_n in the *l.f.c.* and divergence under H_1 follows from applying the continuous mapping theorem after showing that $\mathcal{T}$ is a continuous functional with respect to the uniform norm. See Appendix for details.

Table 1
Simulated critical values of τ_n based on 3,000,000 MC simulations.

$\alpha \setminus n$	10	20	50	100	200	500	1000
0.10	0.790	0.804	0.803	0.805	0.808	0.811	0.813
0.05	0.878	0.859	0.861	0.861	0.868	0.871	0.873
0.01	0.948	0.961	0.977	0.983	0.988	0.992	0.995

Since $\tau_{n\alpha}$ is difficult to calculate analytically, it is approximated by Monte Carlo as accurately as desired. Table 1 reports the approximated critical values of τ_n for different sample sizes based on 3 million Monte Carlo simulations.

We also study the power of the test under a sequence of local alternatives. Therefore, the copula function may depend on the sample size n and is denoted by $C^{(n)}$ rather than C. The sequence of local alternatives is expressed as

$$H_{1n}^\eta : \lim_{n\to\infty}\inf \sqrt{n}\left\|\mathcal{T}C^{(n)} - C^{(n)}\right\|_\infty \ge \eta,$$

where $\eta > 0$. In order to show that our test does not have trivial power in the direction of H_{1n}^η, we need to introduce assumptions on the underlying data generating process to guarantee the weak convergence of K_n. Rüschendorf (1976) has derived the limiting distribution of K_n in the non-iid case, extending the result of Neuhaus (1975) for the standard empirical process. We assume that the observations $\left\{\left(Y_{i,n}, X_{i,n}\right), 1 \le i \le n, n \ge 1\right\}$ form a triangular array with $\left\{\left(Y_{i,n}, X_{i,n}\right)\right\}_{i=1}^n$ iid for each $n \ge 1$ with continuous joint distribution $F^{(n)}$, and marginals $F_Y^{(n)}$ and $F_X^{(n)}$. The standard empirical process

$$\alpha_n(u, v) := \sqrt{n}\left(F_n\left(F_Y^{(n)-1}(u), F_X^{(n)-1}(u)\right) - C^{(n)}(u, v)\right)$$

has covariance function

$$\begin{aligned} L^{(n)}((u_1, v_1), (u_2, v_2)) &= \mathbb{E}(\alpha_n(u_1, v_1)\alpha_n(u_2, v_2)) \\ &= C^{(n)}(u_1 \wedge u_2, v_1 \wedge v_2) \\ &\quad - C^{(n)}(u_1, v_1)\,C^{(n)}(u_2, v_2) \end{aligned}$$

for each n, which is assumed to converge uniformly to a continuous limit. Assumption A2 below adapts Assumptions A and B in Rüschendorf (1976) to the current circumstances.

Assumption A2. Under the local alternatives, $\left\{\left(Y_{i,n}, X_{i,n}\right)\right\}_{i=1}^n$ is a sequence of iid arrays for each $n \ge 1$, with continuous joint distribution $F^{(n)}$ and continuously differentiable copula function $C^{(n)}$ with partial derivatives $c_Y^{(n)}(u, v) := \partial C^{(n)}(u, v)/\partial u$ and $c_X^{(n)}(u, v) := \partial C^{(n)}(u, v)/\partial v$ which converge uniformly to continuous functions c_Y and c_X, respectively. The covariance function $L^{(n)}$ converges uniformly to a continuous function L, which satisfies $L((u, v), (u, v)) \ne 0$ for all (u, v) in the interior of $[0, 1]^2$.

Theorem 2. *Let Assumption A2 hold. Then for any* $\beta \in (0, 1)$ *there is some* $\eta > 0$ *such that, under* H_{1n}^η

$$\lim_{n\to\infty}\inf \mathbb{P}(\tau_n > \tau_{n\alpha}) \ge \beta.$$

The theorem states that the test is able to detect local alternatives converging to the null at the rate $n^{-1/2}$.

In case we are interested in testing monotonicity in a subset $\mathcal{S} \subset \mathcal{Y} \times \mathcal{X}$, we should use as test statistic $\tau_n(\mathcal{W}_n)$, where, for $\mathcal{A} \subseteq [0, 1]^2$,

$$\tau_n(\mathcal{A}) := \sup_{(u,v)\in\mathcal{A}} T_n(u, v),$$

and $\mathcal{W}_n := \left\{(u, v) \in [0, 1]^2 : \left(F_{Yn}^{-1}(u), F_{Xn}^{-1}(v)\right) \in \mathcal{S}\right\}$ is the natural estimator of the set

$$\mathcal{W} := \left\{(u, v) \in [0, 1]^2 : \left(F_Y^{-1}(u), F_X^{-1}(v)\right) \in \mathcal{S}\right\}.$$

The next proposition shows that $\tau_n(\mathcal{W}_n)$ and $\tau_n(\mathcal{W})$ are asymptotically equivalent under the $l.f.c.$

Proposition 3. *Under the l.f.c. and Assumption A1,* $\tau_n(\mathcal{W}_n) = \tau_n(\mathcal{W}) + o_{\mathbb{P}}(1)$.

The proposition justifies using as critical values of the test

$$\tau_{n\alpha}(\mathcal{W}_n) := \inf\{t \in [0, \infty) : \mathbb{P}(\tau_n(\mathcal{W}_n) \le t) \ge 1 - \alpha \text{ under the } l.f.c.\}.$$

The critical values for different sets $\mathcal{W}_n$ can be obtained by Monte Carlo.

3. Testing monotonicity of conditional moments

Consider testing H_0^γ against its negation, say H_1^γ, with γ known in (3). The test statistic is in this case

$$\tau_{\gamma n} := \sup_{v \in [0,1]} T_{\gamma n}(v),$$

where

$$T_{\gamma n}(v) := \frac{\sqrt{n}\left(\mathcal{T} C_{\gamma n} - C_{\gamma n}\right)(v)}{\sigma_{\gamma n}(1)}$$

is based on the weighted empirical process

$$C_{\gamma n}(v) := \frac{1}{n}\sum_{i=1}^{n}\{\gamma(Y_i, X_i) - \overline{\gamma}\}\left\{1_{\left\{X_i \le F_{Xn}^{-1}(v)\right\}} - v\right\},$$

$\overline{\gamma} := n^{-1}\sum_{i=1}^{n}\gamma(Y_i, X_i)$, and

$$\sigma_{\gamma n}^2(v) := \frac{1}{n}\sum_{i=1}^{n}(\gamma(Y_i, X_i) - \overline{\gamma})^2\, 1_{\left\{X_i \le F_{Xn}^{-1}(v)\right\}}.$$

Let B denote the standard Brownian Motion in $[0, 1]$; i.e. B is a Gaussian process with zero mean and covariance function $\mathbb{E}(B(v_1)B(v_2)) = v_1 \wedge v_2$. Also define the function $\eta_\gamma(v) := \sigma_\gamma^2(v)/\sigma_\gamma^2(1)$ with $\sigma_\gamma^2(v) := \mathbb{E}\left(\mathbb{V}ar(\gamma(Y, X)|X)\, 1_{\{F_X(X) \le v\}}\right)$, and note that η_γ is an increasing function with $\eta_\gamma(0) = 0$ and $\eta_\gamma(1) = 1$. Theorem 4 below is proved using the fact that in the $l.f.c.$,

$$T_{\gamma n} \to_d \left\{\mathcal{T}\left(B \circ \eta_\gamma\right) - \left(B \circ \eta_\gamma\right)\right\} \quad \text{on the space } D[0, 1].$$

An asymptotic test is available after noticing that

$$\sup_{v \in [0,1]}\left|\left(\mathcal{T}\left(B \circ \eta_\gamma\right) - \left(B \circ \eta_\gamma\right)\right)(v)\right| \overset{d}{=} Z \equiv \sup_{v \in [0,1]}|(\mathcal{T}B - B)(v)|,$$

where "$\overset{d}{=}$" denotes equality in distribution. The distribution of Z has been already tabulated by Durot (2003). The next theorem justifies the application of the test based on rejecting H_0^γ for large values of $\tau_{\gamma n}$. Let Z_α be the $(1 - \alpha) - th$ quantile of Z.

Theorem 4. *Assume F_X is continuous and $\mathbb{E}\left(\gamma(Y, X)^2\right) < \infty$. Then, under H_0^γ,*

$$\lim_{n\to\infty}\mathbb{P}\left(\tau_{\gamma n} > Z_\alpha\right) \le \alpha.$$

If instead, H_1^γ holds, then

$$\lim_{n\to\infty}\mathbb{P}\left(\tau_{\gamma n} > Z_\alpha\right) = 1.$$

The test $1_{\{\tau_{\gamma n} > Z_\alpha\}}$ extends the proposal in Durot (2003) to stochastic regressors and regression errors possibly depending on X. In particular, we allow in Theorem 4 for conditional heteroscedasticity of unknown form. Simulation results in Durot (2003) suggest that the asymptotic test has large size distortions for moderate sample sizes. To improve the asymptotic approximation we propose approximating Z_α by means of a multiplier-type bootstrap. See Chapter 2.9 in van der Vaart and Wellner (1996). Specifically, once random variables $\mathcal{V}_n := \{V_i\}_{i=1}^n$ are independently generated from the sample $\mathcal{Z}_n$ according to a random variable V with bounded support, mean zero and variance one, the asymptotic critical value Z_α is estimated by the $[N(1 - \alpha)]$-th order statistic computed from N replicates $\left\{\tau_{\gamma nj}^*\right\}_{j=1}^N$ of $\tau_{\gamma n}^*$, where

$$\tau_{\gamma n}^* := \frac{\sqrt{n}}{\sigma_{\gamma n}(1)}\sup_{v \in [0,1]}\left(\mathcal{T} C_{\gamma n}^* - C_{\gamma n}^*\right)(v)$$

and

$$C_{\gamma n}^*(v) := \frac{1}{n}\sum_{i=1}^{n}\{\gamma(Y_i, X_i) - \overline{\gamma}\}\left\{1_{\left\{X_i \le F_{Xn}^{-1}(v)\right\}} - v\right\}V_i.$$

The bootstrap consistency follows from our Theorem 4 and the results in Stute et al. (1998). We can proceed as in Theorem 2 to prove that the test can also detect local alternatives. Details are omitted.

4. Finite sample performance

4.1. Monte Carlo

We carried out a simulation study to demonstrate the finite-sample performance of the proposed test, in comparison with LLW's approach. For the sake of completeness we briefly describe their test statistic. LLW's approach is an extension of that by Ghosal et al. (2000) to test for monotonicity in the whole conditional distribution rather than just in the regression function. Their test is based on the U-process

$$\hat{U}_n(x, y) = \binom{n}{2}^{-1}\sum_{1 \le i < j \le n}\left\{1_{\{Y_i \le y\}} - 1_{\{Y_j \le y\}}\right\} \times sgn(X_i - X_j)k_{hi}(x)\,k_{hj}(x), \quad (y, x) \in \mathcal{Y} \times \mathcal{X},$$

where sgn denotes the sign function, $k_{h\ell}(\cdot) = h^{-1}k(X_\ell - \cdot/h)$, k is a kernel function and h is a bandwidth such that $h \to 0$ as $n \to \infty$. Notice that $\hat{U}_n(x, y)$ estimates $\partial F_{Y|X}(y|x)/\partial x$ times a positive function, see LLW. They consider the Kolmogorov–Smirnov criterion

$$\widehat{\mathcal{U}}_n = \sup_{(y,x) \in \mathcal{Y} \times \mathcal{X}}\frac{\hat{U}_n(x, y)}{c_n(x)},$$

for a suitable standardized factor $c_n(x) = n^{-1/2}\hat{\sigma}_n(x)$. Their test rejects for large values of $\widehat{\mathcal{U}}_n$. Notice that the values of the test statistic $\widehat{\mathcal{U}}_n$ may change under monotonic continuous transformations of the explanatory variable X, while τ_n is always invariant to such transformations for each n. Under H_0, $\widehat{\mathcal{U}}_n$ is asymptotically distributed as an extreme value random variable and the level accuracy is poor in finite samples. To overcome this problem, LLW suggests computing critical values by an approximation to the asymptotic distribution, as in Ghosal et al. (2000). We refer the reader to LLW's article for an explicit expression of the test's rejection region. We report results using their choice for the kernel function, the Epanechnikov kernel $k(u) = 0.75(1 - u^2)$, and bandwidth values $h = 0.4, 0.5, 0.6$ and 0.7. We denote their test by $LLW_{n,h}$ in our simulations.

We consider the following data generating processes (DGPs). Let $\{\varepsilon_i\}_{i=1}^n$ be a sequence of iid $N(0, 0.1^2)$ random variables, and let $\{X_i\}_{i=1}^n$ be a sequence of iid $U[0, 1]$ variables, independent of the sequence $\{\varepsilon_i\}_{i=1}^n$. Then, the sample $\{Y_i\}_{i=1}^n$ is generated according to:

Table 2
Rejection frequencies at 5% nominal level based on 5000 MC simulations.

Model	n	τ_n	$LLW_{n,0.4}$	$LLW_{n,0.5}$	$LLW_{n,0.6}$	$LLW_{n,0.7}$
	50	0.044	0.022	0.023	0.028	0.029
N1	200	0.049	0.035	0.035	0.038	0.038
	500	0.046	0.031	0.034	0.036	0.036
	50	0.001	0.005	0.005	0.005	0.010
N2	200	0.000	0.002	0.002	0.006	0.023
	500	0.000	0.000	0.002	0.015	0.053
	50	0.463	0.680	0.759	0.774	0.758
ALT1	200	0.998	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000
	50	0.385	0.120	0.207	0.284	0.343
ALT2	200	0.901	0.537	0.743	0.854	0.906
	500	0.999	0.955	0.994	0.999	1.000
	50	0.079	0.056	0.065	0.065	0.058
ALT3	200	0.279	0.249	0.243	0.231	0.202
	500	0.732	0.641	0.595	0.557	0.496
	50	0.011	0.009	0.012	0.019	0.031
ALT4	200	0.157	0.018	0.012	0.011	0.017
	500	0.812	0.057	0.018	0.013	0.007
	50	0.268	0.077	0.116	0.141	0.167
ALT5	200	0.759	0.465	0.523	0.547	0.561
	500	0.991	0.923	0.935	0.935	0.927

N1: $Y_i = \varepsilon_i$.
N2: $Y_i = 0.1X_i + \varepsilon_i$.
ALT1: $Y_i = X_i(1 - X_i) + \varepsilon_i$.
ALT2: $Y_i = -0.1X_i + \varepsilon_i$.
ALT3: $Y_i = -0.1 \exp\left(-250(X_i - 0.5)^2\right) + \varepsilon_i$.
ALT4: $Y_i = 0.2X_i - 0.2 \exp\left(-250(X_i - 0.5)^2\right) + \varepsilon_i$.
ALT5: $Y_i = 0.2 \cdot 1_{\{X_i \le 0.5\}} X_i(1 - X_i) + \varepsilon_i$.

Models N1 and ALT1 were considered in LLW, models ALT2–ALT4 have been used in the isotonic regression literature, see Durot (2003) and references therein, and model ALT5 is a non-smooth variant of ALT1 in LLW to see the robustness of the procedures to lack of smoothness. In these simulations we compare LLW's test with ours. Table 2 reports the proportion of rejections in 5000 Monte Carlo replications of the two tests at the 5% significance level under the seven designs and with sample sizes $n = 50$, 200 and 500. The results with other nominal levels were similar, and hence, they are not reported.

The reported empirical sizes for τ_n are accurate for N1. In agreement with the results in LLW, their test shows some underrejection for the *l.f.c.* in N1. The design N2 corresponds to a data generating process in the interior of the null hypothesis. Hence, as expected, the proportion of rejections in N2 is small and converging to zero with the sample size. As for the alternatives, none of the tests is uniformly better than the others. LLW's test performs best for the alternative ALT1, but our test outperforms theirs for ALT2–ALT5. These alternatives suggest that our test based on τ_n can be complementary to LLW's test. In Fig. 1(a) we plot the regression function corresponding to ALT4. We observe that this alternative is relatively close to the null hypothesis, so it requires large sample sizes ($n = 500$) to be detected by our procedure. LLW has no power against this alternative for the sample sizes considered. Our test outperforms LLW under ALT5, the non-smooth variant of ALT1.

To better understand the power properties of our test, we consider the following DGP:

$$\text{ALT6:} \quad Y_i = a 1_{\{X_i \le 0.5\}} (X_i - 0.5)^3 - \exp\left(-250(X_i - 0.5)^2\right) + \varepsilon_i,$$

where $\{\varepsilon_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$ are as in the previous simulations. ALT6 represents a model on the alternative hypothesis which becomes farther away from the *l.f.c.* as $a \to \infty$. In Fig. 1(b) we plot the regression function corresponding to $a = 15$. From this plot we observe that this represents another alternative close to the null hypothesis.

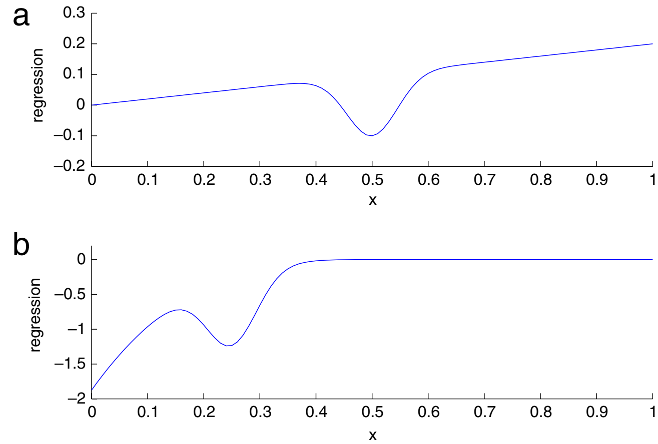

Fig. 1. Regression functions for alternatives ALT4 (top panel) and ALT6 (bottom panel) with $a = 15$.

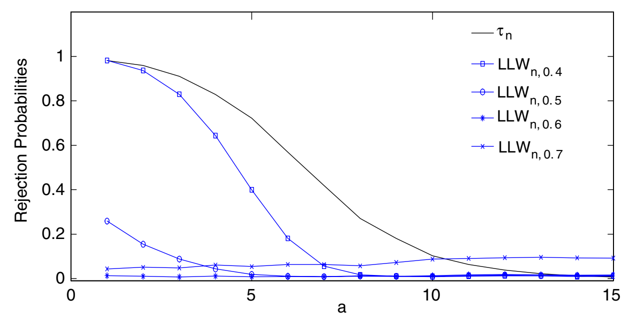

Fig. 2. Rejection probabilities for ALT6 as a function of a. 5000 Monte Carlo simulations. Sample size $n = 200$.

In Fig. 2, we plot the empirical rejection probabilities for ALT6, based on 5000 Monte Carlo replications at the 5% nominal level and using the sample size $n = 200$. Several remarks are in order. On one hand, LLW's test only has power against this alternative for low values of a and low values of the bandwidth parameter. The proportions of rejections are very sensitive to the bandwidth choice. On the other hand, τ_n performs best, particularly for moderate values of a. For $a = 15$ none of the tests have power. In unreported simulations, we have observed that, for $n = 500$ and $a = 15$, τ_n is able to detect this alternative, whereas LLW's test shows a flat power at the nominal level.

To summarize, these simulations suggest that the performance of our supremum statistic is satisfactory, and compares favorably to the only competing alternative in LLW. Our test does not require bandwidth choices and, hence, it should be appealing to practitioners.

4.2. Application to intergenerational income mobility

In this section, we investigate the monotonicity of the intergenerational transition function using a data set from the Panel Study of Income Dynamics (PSID). We use the same data set as in LLW. The dependent variable Y is the logarithm of sons' averaged full-time real labor income at ages 28 and 29, and the X variable is the logarithm of parental predicted permanent income. The number of observations is $n = 616$.[1] We aim to test for

[1] The data set can be obtained from the *Econometrica* web site.

stochastic monotonicity of the distribution of Y given X. Recently, Dardanoni et al. (2012) has considered the same testing problem using a different data set.

We have applied the proposed test to this data set and obtained a test statistic of $\tau_n = 0.4025$. The 10% critical values with $n = 500$ and 1000 are, respectively, 0.811 and 0.813, and hence, we fail to reject the null hypothesis of stochastic monotonicity. LLW obtained a similar conclusion with their test. Interestingly enough, if a monotonic transformation of X is performed, such as $\log(X)$, the LLW's test may change its conclusion. For instance, an application of LLW proposal to this data set after the transformation $\log(X)$ results in rejection of the null at 10% with p-values 0.093, 0.077 and 0.084, corresponding to bandwidths $h = 0.4, 0.5$ and 0.6, respectively. In contrast, our test statistic takes the same value $\tau_n = 0.4025$, and hence, it leads to the same conclusion. We conclude this application emphasizing that invariance should be a minimum requirement for any test of stochastic monotonicity. The proposed procedure possesses this property, in addition to other attractive features such as its distribution-free property, consistency, and simple and general applicability.

Appendix. Proofs of the main results

Proof of Theorem 1. (a) Define $G_n := n^{-1/2}K_n = C_n - C$. Then, by definition of the l.c.m, the function $\mathcal{T}G_n(u,\cdot) + C(u,\cdot)$ is above $C_n(u,\cdot)$ and is concave in v under H_0, for each $u \in [0,1]$, since both $\mathcal{T}G_n(u,\cdot)$ and $C(u,\cdot)$ are concave for each $u \in [0,1]$. Hence, $\mathcal{T}G_n + C$ is uniformly above $\mathcal{T}C_n$. Thus, under H_0, uniformly in $[0,1]^2$,

$$\begin{aligned} T_n &= \sqrt{n}\,(\mathcal{T}C_n - C_n) \\ &\le \sqrt{n}\,(\mathcal{T}G_n - G_n) \\ &=: \tilde{T}_n. \end{aligned} \quad (9)$$

When $C(u,v) = uv$,

$$\mathcal{T}G_n(u,v) = \mathcal{T}C_n(u,v) - uv,\ (u,v) \in [0,1]^2. \quad (10)$$

To see this, note that $\mathcal{T}G_n(u,v) + uv \ge C_n(u,v)$ by definition, and since $\mathcal{T}G_n(u,v) + uv$ is concave in v for each $u \in [0,1]$, then $\mathcal{T}G_n(u,v) + uv \ge \mathcal{T}C_n(u,v)$. Similarly, $\mathcal{T}C_n(u,v) - uv \ge C_n(u,v) - uv$, and since $\mathcal{T}C_n(u,v) - uv$ is concave in v for each $u \in [0,1]$, then $\mathcal{T}G_n(u,v) \le \mathcal{T}C_n(u,v) - uv$. Thus, (9) becomes equality when $C(u,v) = uv$. This result has been applied by Prakasa Rao (1969) among others. Hence,

$$\mathbb{P}(\tau_n > \tau_{n\alpha} | H_0) \le \mathbb{P}(\tilde{\tau}_n > \tau_{n\alpha} \mid l.f.c) \le \alpha,$$

where $\tilde{\tau}_n := \left\|\tilde{T}_n\right\|_\infty$. Note that τ_n is a random variable, i.e. measurable, since it is the maximum of a finite number of random variables, see (8).

(b) First, we show that $\mathcal{T} : D[0,1]^2 \longmapsto C[0,1]^2$ is a continuous functional with respect to the uniform norm $\|\cdot\|_\infty$. Lemma 2.2 in Durot and Tocquet (2003) implies that for each $f, g \in D[0,1]^2$,

$$\sup_{v\in[0,1]} |(\mathcal{T}f - \mathcal{T}g)(u,v)| \le \sup_{v\in[0,1]} |(f-g)(u,v)|$$
$$\text{for each } u \in [0,1] \text{ fixed.} \quad (11)$$

Since the last inequality holds for all $u \in [0,1]$, for any $f, g \in D[0,1]^2$, $\|\mathcal{T}f - \mathcal{T}g\|_\infty \le \|f-g\|_\infty$, which proves that $\mathcal{T}$ is continuous.

In order to prove that τ_n diverges under H_1, using the fact that the marginal distributions are continuous, we can assume without loss of generality that the marginals are uniform, see Lemma 1 in Fermanian et al. (2004). Then, write

$$C_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} 1_{\left\{X_i \le F_{Xn}^{-1}(u)\right\}} 1_{\left\{Y_i \le F_{Yn}^{-1}(v)\right\}}.$$

Therefore,

$$\|C_n - C\|_\infty \le \sup_{(u,v)\in[0,1]^2} \left|C_n(u,v) - C\left(F_{Yn}^{-1}(u), F_{Xn}^{-1}(v)\right)\right| + \sup_{(u,v)\in[0,1]^2} \left|F\left(F_{Yn}^{-1}(u), F_{Xn}^{-1}(v)\right) - F(u,v)\right|, \quad (12)$$

by triangle inequality and noticing that $C((F_Y \circ f)(u), (F_X \circ g)(v)) = F(f(u), g(v))$. Therefore, $\|C_n - C\|_\infty \to 0$ a.s. $- \mathbb{P}$ since the first term on the r.h.s of (12) converges applying the Glivenko–Cantelli Theorem, as the second term does using the fact that F is continuous and, for $\xi = Y$ or $\xi = X$ with uniform marginals,

$$\sup_{u\in[0,1]} \left|F_{\xi n}^{-1}(u) - u\right| \to_{a.s.} 0.$$

Therefore, by (11), $\|\mathcal{T}C_n - \mathcal{T}C\|_\infty \to_{a.s.} 0$ as $n \to \infty$. Hence, under fixed alternatives, $\|\mathcal{T}C_n - C_n\|_\infty$ converges to $\|\mathcal{T}C - C\|_\infty > 0$, and τ_n diverges to $+\infty$. Then, part (b) of the theorem is proved by showing that $\tau_{n\alpha} = O(1)$, which is done in the following arguments.

Rüschendorf (1976), see also Deheuvels (1981a,b) and Gänssler and Stute (1987), have shown that under the $l.f.c.$, assuming $A1$, K_n converges in distribution to K_∞ on the space $D[0,1]^2$ endowed with the Skorohod's norm $\|\cdot\|_S$, $(D[0,1]^2, \|\cdot\|_S)$, where K_∞ is a "completely tucked" Brownian sheet, i.e. a continuous Gaussian process with mean zero and covariance function

$$\mathbb{E}(K_\infty(u_1,v_1)K_\infty(u_2,v_2)) = ((u_1 \wedge u_2) - u_1u_2)((v_1 \wedge v_2) - v_1v_2). \quad (13)$$

The limiting distribution of $T_n = \sqrt{n}\,(\mathcal{T}C_n - C_n)$ under the $l.f.c.$ is obtained as a consequence of the continuous sample path property of K_∞ together with the fact that $\mathcal{T}$ is $\|\cdot\|_\infty$-continuous, and that every $\|\cdot\|_S$-convergent sequence in $D[0,1]^2$ is $\|\cdot\|_\infty$-convergent, if the limit belongs to $C[0,1]^2$. A proof of it may use a general continuous mapping theorem (c.f. Billingsley, 1999, Theorem 2.7) or may involve almost surely convergent constructions using a Skorohod embedding Theorem. The second strategy provides directly the asymptotic distribution of τ_n under the $l.f.c.$ According to Skorohod embedding Theorem, there are versions $\hat{K}_n$ and $\hat{K}_\infty$ of K_n and K_∞, respectively (defined on some probability space $(\widehat{\Omega}, \widehat{\mathcal{A}}, \widehat{\mathbb{P}})$), such that $\left\|\hat{K}_n - \hat{K}_\infty\right\|_S \to 0$ $a.s. - \widehat{\mathbb{P}}$ and therefore $\left\|\hat{K}_n - \hat{K}_\infty\right\|_\infty \to 0$ $a.s. - \widehat{\mathbb{P}}$. Hence, by continuity of $\mathcal{T}$, $\left\|\left(\mathcal{T}\hat{K}_n - \hat{K}_n\right) - \left(\mathcal{T}\hat{K}_\infty - \hat{K}_\infty\right)\right\|_\infty \to 0$ $a.s. - \widehat{\mathbb{P}}$, and $\tau_n = \|\mathcal{T}K_n - K_n\|_\infty$ converges in distribution to $\tau_\infty \overset{d}{=} \|\mathcal{T}K_\infty - K_\infty\|_\infty$. By Lifshits (1982) the distribution of τ_∞ is continuous, and then the convergence of τ_n to τ_∞ implies $\tau_{n\alpha} = O(1)$. □

Proof of Theorem 2. Recall $K_n = \sqrt{n}(C_n - C^{(n)})$ and write

$$T_n = \sqrt{n}\left(\mathcal{T}C^{(n)} - C^{(n)}\right) + \left(\sqrt{n}\left(\mathcal{T}C_n - \mathcal{T}C^{(n)}\right) - K_n\right). \quad (14)$$

By (11)

$$\left\|\sqrt{n}\left(\mathcal{T}C_n - \mathcal{T}C^{(n)}\right) - K_n\right\|_\infty \le 2\|K_n\|_\infty. \quad (15)$$

Therefore, the theorem follows after showing that K_n converges in distribution. Under Assumption A2, α_n converges in distribution on $\left(D[0,1]^2, \|\cdot\|_S\right)$ to α_∞, a continuous Gaussian process, centered at zero and with covariance function L. See Neuhaus (1975) for convergence of α_n under general non-iid conditions. Therefore, Assumptions A and B in Theorem 3.3 of Rüschendorf (1976) are satisfied and K_n converges in distribution on $\left(D[0,1]^2, \|\cdot\|_S\right)$ to K'_∞, where

$$K'_\infty(u,v) \overset{d}{=} \alpha_\infty(u,v) - c_Y(u,v)\,\alpha_\infty(1,v) - c_X(u,v)\,\alpha_\infty(u,1).$$

Notice that $K'_\infty \overset{d}{=} K_\infty$ in the $l.f.c$. Therefore, $T_n = \sqrt{n}\left(\mathcal{T}C^{(n)} - C^{(n)}\right) + O_{\mathbb{P}}(1)$ and the Theorem follows using the fact that the critical value $\tau_{n\alpha} = O(1)$. □

Proof of Proposition 3. From triangle inequality, the stochastic equicontinuity of T_n, and Glivenko–Cantelli's theorem,

$$\begin{aligned}
&\left| \sup_{(u,v)\in\mathcal{W}_n} T_n(u,v) - \sup_{(u,v)\in\mathcal{W}} T_n(u,v) \right| \\
&\quad \le \sup_{(y,x)\in\mathcal{S}} \left| T_n(F_{Yn}(y), F_{Xn}(x)) - T_n(F_Y(y), F_X(x)) \right| \\
&\quad \le \sup_{|u-u'|\le\delta_{yn},\, |v-v'|\le\delta_{xn}} \left| T_n(u,v) - T_n(u',v') \right| \\
&\quad = o_{\mathbb{P}}(1),
\end{aligned}$$

where $\delta_{yn} := \sup_{y\in\mathcal{Y}} |F_{Yn}(y) - F_Y(y)|$ and $\delta_{xn} := \sup_{x\in\mathcal{X}} |F_{Xn}(x) - F_X(x)|$. □

Proof of Theorem 4. Define $\gamma_0 := \mathbb{E}(\gamma(Y_i, X_i))$ and write

$$C_{\gamma n}(v) = \tilde{C}_{\gamma n}(v) + R_{\gamma n}(v),$$

where

$$\tilde{C}_{\gamma n}(v) := \frac{1}{n}\sum_{i=1}^{n} \{\gamma(Y_i, X_i) - \gamma_0\} \left\{ 1_{\left\{X_i \le F_X^{-1}(v)\right\}} - v \right\}$$

$$\begin{aligned}
R_{\gamma n}(v) &:= \frac{1}{n}\sum_{i=1}^{n} \{\gamma(Y_i, X_i) - \gamma_0\} \left\{ 1_{\left\{X_i \le F_{Xn}^{-1}(v)\right\}} - 1_{\left\{X_i \le F_X^{-1}(v)\right\}} \right\} \\
&\quad + \{\gamma_0 - \overline{\gamma}\} \frac{1}{n}\sum_{i=1}^{n} \left\{ 1_{\left\{X_i \le F_{Xn}^{-1}(v)\right\}} - v \right\}.
\end{aligned}$$

Applying the same arguments as for (12),

$$\sup_{v\in[0,1]} \left| F_{Xn}(F_{Xn}^{-1}(v)) - F_X(F_X^{-1}(v)) \right| = o_{\mathbb{P}}(1). \qquad (16)$$

Hence, the last display and the central limit theorem yield $\sup_{v\in[0,1]} \left|R_{\gamma n}(v)\right| = o_{\mathbb{P}}\left(n^{-1/2}\right)$. Stute (1997) has shown that under continuity of F_X and $\mathbb{E}\left(\gamma(Y,X)^2\right) < \infty$,

$$\tilde{C}_{\gamma n}(\cdot) \overset{d}{\to} C_\gamma \quad \text{on } (D[0,1], \|\cdot\|_S) \text{ in the } l.f.c.,$$

where $C_\gamma(v) \overset{d}{=} B \circ \eta_\gamma(v) - v\sigma_\gamma(1)B(1)$. Thus, by the equality

$$\mathcal{T}\left(B \circ \eta_\gamma(v) - B(1)\sigma_\gamma(1)v\right) = \mathcal{T}\left(B \circ \eta_\gamma(v)\right) - B(1)\sigma_\gamma(1)v,$$

which is proved as (10), we obtain that in the $l.f.c.$,

$$T_{\gamma n} \overset{d}{\to} \left\{\mathcal{T}\left(B \circ \eta_\gamma\right) - \left(B \circ \eta_\gamma\right)\right\} \quad \text{on the space } D[0,1].$$

From this point the proof follows exactly the same arguments as those of Theorem 1. □

References

Aguirregabiria, V., 2010. Another look at the identification of dynamic discrete decision processes: an application to retirement behavior. Journal of Business and Economic Statistics 28, 201–218.
Akerlof, G., 1970. The market for lemons: qualitative uncertainty and the market mechanism. Quarterly Journal of Economics 84, 488–500.
Akerlof, G., 1976. The economics of caste, the rat race, and other woeful tales. Quarterly Journal of Economics 90, 599–617.
Apostol, T., 1967. Calculus, Vol. I, second ed. Wiley, New York.
Banerjee, M., Mukherjee, D., Mishra, S., 2009. Semiparametric binary regression models under shape constraints with an application to Indian schooling data. Journal of Econometrics 149, 101–117.
Barlow, R.E., Bartholomew, D.J., Bremmer, J.M., Brunk, H.D., 1972. Statistical Inference Under Order Conditions. Wiley, New York.
Bickel, P.J., Wichura, M.J., 1971. Convergence criteria for multiparameter stochastic processes and some applications. Annals of Mathematical Statistics 42, 1656–1670.
Billingsley, P., 1999. Convergence of Probability Measures, second ed. Wiley, New York.
Boudoukh, J., Richardson, M., Smith, T., Whitelaw, R.F., 1999. Ex-ante bond returns and the liquidity preference hypothesis. Journal of Finance 54, 1153–1167.
Bowman, A.W., Jones, M.C., Gijbels, I., 1998. Testing monotonicity of regression. Journal of Computations and Graphical Statistics 7, 489–500.
Bril, G., Dykstra, R., Pillers, C., Robertson, T., 1984. Algorithm AS 206: isotonic regression in two independent variables. Journal of the Royal Statistical Society, Series C (Applied Statistics) 352–357.
Brunk, H.D., 1958. On the estimation of parameters restricted by inequalities. Annals of Mathematical Statistics 29, 437–454.
Chernoff, H., 1964. Estimation of the mode. Annals of the Institute of Statistical Mathematics 16, 31–41.
Cran, G.W., 1980. Algorithm AS 149: amalgamation of means in the case of simple ordering. Journal of the Royal Statistical Society, Series C (Applied Statistics) 29, 209–211.
Dardanoni, V., Fiorini, M., Forcina, A., 2012. Stochastic monotonicity in intergenerational mobility tables. Journal of Applied Econometrics 27, 85–107.
de Leeuw, J., Hornik, K., Mair, P., 2009. Isotone Optimization in R: Pool–Adjacent–Violators (PAVA) and active set methods. Journal of Statistical Software 32, 1–24.
Deheuvels, P., 1981a. A Kolmogorov–Smirnov type test for independence and multivariate samples. Revue Roumaine de Mathematiques Pures et Appliquees 26, 213–226.
Deheuvels, P., 1981b. A nonparametric test for independence. Publications de l'Institut de Statistique de l'Université Paris 26, 29–50.
Durot, C., 2003. A Kolmogorov-type test for monotonicity of regression. Statistics & Probability Letters 63, 425–433.
Durot, C., Tocquet, A.S., 2003. On the distance between the empirical process and its least concave majorant in a monotone regression framework. Annales de l'Institut Henri Poincaré B 29, 217–240.
Ericson, R., Pakes, A., 1995. Markov-perfect industry dynamics: a framework for empirical work. Review of Economic Studies 62, 53–82.
Escanciano, J.C., Jacho-Chávez, D.T., 2012. $\sqrt{n}$-uniformly consistent density estimation in nonparametric regression models. Journal of Econometrics 167, 305–316.
Fermanian, J.D., Radulovic, D., Wegkamp, M., 2004. Weak convergence of empirical copula processes. Bernoulli 10, 847–860.
Gänssler, P., Stute, W., 1987. Seminar on Empirical Processes. DMV Seminar. Birkhäuser, Basel.
Ghosal, S., Sen, A., van der Vaart, A.W., 2000. Testing monotonicity of regression. The Annals of Statistics 28, 1054–1082.
Grenander, U., 1956. On the theory of mortality measurement. part II. Skandinavisk Aktuarietidskrift 39, 125–153.
Hall, P., Heckman, E., 2000. Testing for monotonicity of a regression mean by calibrating for linear functions. The Annals of Statistics 28, 20–39.
Hartigan, J.A., Hartigan, P.M., 1985. The dip test of unimodality. The Annals of Statistics 13, 70–84.
Kadi, I., Pekergin, N., Vincent, J-M., 2009. Different monotonicity definitions in stochastic modelling, ASMTA'09 Proceedings of the 16th International Conference on Analytical and Stochastic Modeling Techniques and Applications.
Lee, S., Linton, O.B., Whang, Y-J., 2009. Testing for stochastic monotonicity. Econometrica 77, 585–602.
Lewbel, A., Linton, O.B., 2007. Nonparametric matching and efficient estimators of homothetically separable functions. Econometrica 75, 1209–1227.
Lifshits, M.A., 1982. On the absolute continuity of distributions of functionals of random processes. Theory of Probability and Its Applications 27, 600–607.
Lucas, R.E., Stokey, N.L., 1989. Recursive Methods in Dynamics Economics. Harvard University Press, Cambridge, MA.
Mammen, E., 1991. Estimating a smooth nonparametric regression function. The Annals of Statistics 19, 724–740.
Manski, C., Pepper, J., 2000. Monotone instrumental variables: with an application to the returns to schooling. Econometrica 68, 997–1010.
Matzkin, R.L., 1994. Restrictions of economic theory in nonparametric methods. In: Engel, R.F., McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. pp. 2523–2558.
Milgrom, P.R., 1981. Good and bad news: representation theorems and applications. The Bell Journal of Economics 12, 380–391.
Mukerjee, H., 1988. Monotone nonparametric regression. The Annals of Statistics 16, 741–750.
Neuhaus, G., 1971. On weak convergence of stochastic processes with multidimensional time parameter. Annals of Mathematical Statistics 42, 1285–1295.
Neuhaus, G., 1975. Convergence of the reduced empirical process. The Annals of Statistics 3, 528–531.
Olley, G.S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64, 1263–1297.
Pakes, A., 1986. Patents as options: some estimates of the value of holding European patent stocks. Econometrica 54, 755–784.
Patton, A., Timmermann, A., 2009. Monotonicity in asset returns: new tests with applications to the term structure, the CAPM and portfolio sorts. Journal of Financial Economics 98, 605–625.
Prakasa Rao, B.L.S., 1969. Estimation of a unimodal density. Sankhyá (A) 31, 23–36.
Richardson, M., Richardson, P., Smith, T., 1992. The monotonicity of the term premium: another look. Journal of Financial Economics 31, 97–106.

Rüschendorf, L., 1976. Asymptotic distributions of multivariate rank order statistic. The Annals of Statistics 4, 912–923.

Schlee, W., 1982. Nonparametric tests of the monotony and convexity of regression. In: Gnedenko, B.V., Puri, M.L., Vincze, I. (Eds.), Nonparametric Statistical Inference, Vol. 2. North Holland, Amsterdam, pp. 823–836.

Spence, A.M., 1973. Job market signaling. Quarterly Journal of Economics 87, 355–374.

Stute, W., 1997. Nonparametric model checks for regression. The Annals of Statistics 25, 613–641.

Stute, W., González-Manteiga, W., Presedo-Quindimil, M., 1998. Bootstrap approximations in model checks for regression. Journal of the American Statistical Association 93, 141–149.

Tanaka, H., 2008. Semiparametric least squares estimation of monotone single index models and its application to the iterative least squares estimation of binary choice models, Mimeo. Department of Economics. University of Wisconsin-Madison.

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