Distribution-free Tests of Stochastic Monotonicity[1]

Miguel A. Delgado[2] Juan Carlos Escanciano [3]
Universidad Carlos III de Madrid Indiana University

September 9th, 2010.

Abstract: This article proposes an omnibus test for monotonicity of nonparametric conditional distributions and its moments. Unlike previous proposals, our method does not require smooth estimation of the derivatives of nonparametric curves and it can be implemented even when probability densities do not exist. In fact, we only require continuity of the marginal distributions under the null and fixed alternatives. Distinguishing features of our approach are that critical values are pivotal under the null in finite samples and 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 article also discusses several applications and extensions of the proposal. These include testing monotonicity of general conditional moments and the extension to multivariate explanatory variables. The finite sample performance of the test is examined by means of a Monte Carlo experiment.

Key words and phrases: Stochastic monotonicity; conditional moments; least concave majorant; copula process; distribution-free in finite samples; tests invariant to monotone transforms.

JEL Classification Number: C14, C15.

[1] This research is funded by Spanish Plan Nacional de I+D+I grant number SEJ2007-62908.

[2] Departamento de Economía, Universidad Carlos III de Madrid, 28903 Getafe (Madrid), Spain. E-mail: miguelangel.delgado@uc3m.es

[3] Economics Department, Indiana University, Bloomington, Indiana, USA; E-mail: jescanci@indiana.edu

1 Introduction

Let (Y, X) be a bivariate random vector taking values in $\mathcal{Y} \times \mathcal{X} \subseteq \mathbb{R}^2$ with 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 is primarily concerned with nonparametric testing of the monotonicity of $F_{Y|X}$ with respect to the explanatory variable X. That is, the null hypothesis is

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

where

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

is the set of monotonically non-increasing functions with support $\mathcal{X}$. 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 mutandi.

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. Mukerjee (1988) and Mammen (1991). A test for H_0 has been recently proposed by Lee, Linton and Whang (2009), LLW henceforth, which generalizes the test of monotonicity for regression functions proposed by Ghosal, Seen and van der Vaart (2001). LLW offers a fairly comprehensive account of motivations for testing H_0 in economics research. See also Matzkin (1994) for a survey on how the monotonicity restriction, amongst other shape restrictions, can be derived from an economic model and how these restrictions can be used for identification and estimation of structural nonparametric curves.

The LLW and Ghosal, Seen and van der Vaart (2001) tests, 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 Schlee (1982), Bowman, Jones and Gijbels (1998) 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 a suitable choice of 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

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

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

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

Therefore, since F_X^{-1} is a non-decreasing function, we can characterize H_0 as

$$H_0 : C(u,\cdot) \in \mathcal{C} \text{ for each } u \in [0,1],$$

where $\mathcal{C}$ is the set of concave functions.

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 $C(u,\cdot)$ for each $u \in [0,1]$ fixed, $\mathcal{T}C(u,\cdot)$, is the function satisfying the following two properties: (i) $\mathcal{T}C(u,\cdot) \in \mathcal{C}$ and (ii) if there exists $h \in \mathcal{C}$ with $h \geq C(u,\cdot)$, then $h \geq \mathcal{T}C(u,\cdot)$. Henceforth, $\mathcal{T}C$ denotes the function resulting of applying the operator $\mathcal{T}$ to the function $C(u,\cdot)$ for each $u \in [0,1]$. Thus, we can alternatively write H_0 as

$$H_0 : \mathcal{T}C \equiv C. \tag{3}$$

Obviously, the greatest convex majorant must be used for characterizing H_0 in the monotonically non-decreasing case. Grenander (1956) 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 set up 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 naive application 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 cdf´s, only assuming continuity of the marginal distributions. Distinguishing features of our approach are that the test's critical values are pivotal under the null and the test is invariant to any monotonic continuous transformation of the explanatory variable in finite samples. The latter is a minimal requirement for any test of monotonicity.[4] Our proposal permits to relax different smoothness assumptions on the underlying nonparametric curves imposed by LLW and related tests. In particular, continuity of the c.d.f. with respect to the conditioning variable or the existence of conditional densities are not needed under the null and fixed alternatives. The minimal continuity assumption on the marginal distributions is satisfied in many relevant situations where the conditional distribution is discontinuous with respect the explanatory variable. For instance, bimodality in marginal income distributions is often explained because of different income distributions in two subpopulations, which can be defined in terms of an explanatory variable (e.g. education) and some threshold. The conditional distribution consists of a continuous distribution for each sub-population and the marginal, obtained by integrating out the explanatory variable, is naturally continuous. A mixture of two continuous distributions with mixing parameter a possibly discontinuous function could model this situation. For examples in economics where densities may not exists or are discontinuous see e.g. Chernozhukov and Hong (2004) and Zinde-Walsh (2008). Finally, unlike with competing methods, the exact computation of our test is straightforward, its performance does not depend on the choice of a smoothing number and the test is able to detect local alternatives that approach the null hypothesis at the rate $n^{-1/2}$, with n the sample size.

The rest of the article is organized as follows. Next section introduces the new test, discussing its asymptotic behavior under H_0 and local alternatives. The results of a Monte Carlo study are summarized in Section 3. Last Section is devoted to final remarks, which include extensions of the basic framework to testing the monotonicity of general conditional moments and extensions with a vector of explanatory variables. For the multivariate case, we consider monotonicity with respect to only one coordinate and the hypothesis of stochastic semimonoticity, in the sense of Manski (1997). Proofs are placed in a technical mathematical appendix at the end of the article.

[4]For instance, if X is total expenditure and Y is expenditure on food, our method delivers a test for monotonicity of Engle curves that is invariant to whether X is measured in dollars or euros.

2 Testing monotonicity of a conditional distribution

Given an idependent and identically distributed (iid) sample $\{(Y_i, X_i),\ i = 1, ..., n\}$, distributed as (Y, X), the natural estimator of $C(u, v)$ is

$$C_n(u, v) := \frac{1}{n} \sum_{i=1}^{n} 1_{\{F_{Yn}(Y_i) \le u\}} 1_{\{F_{Xn}(X_i) \le v\}},\ (u, v) \in [0, 1]^2, \tag{4}$$

where, given a sample $\{\xi_i\}_{i=1}^{n}$ of a generic r.v. ξ, $F_{\xi n}(\cdot) := n^{-1} \sum_{i=1}^{n} 1_{\{\xi_i \le \cdot\}}$ is the sample analog of F_ξ. The process

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

is the standard empirical copula process. Deheuvels (1981a, 1981b) first obtained the exact law and the limiting distribution of K_n when Y and X are independent, see also Gänssler and Stute (1987). In particular, Deheuvels (1981a, 1981b) proved that,

$$K_n \to_d K_\infty \text{ on the extended Skorohod's space in } D[0, 1]^2,$$

where K_∞ is a "completely tucked" Brownian sheet, a continuous Gaussian process with mean zero and covariance function

$$\mathbb{E}(K_\infty(u_1, v_1) K_\infty(u_2, v_2)) = (u_1 \wedge u_2 - u_1 u_2)(v_1 \wedge v_2 - v_1 v_2),$$

for $(u_i, v_i) \in [0, 1]^2$, $i = 1, 2$. That is, K_∞ is distributed as the product of two independent standard Brownian Bridges in $[0, 1]$.

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 $\mathcal{T}(C_n(u, v) - uv) = \mathcal{T} C_n(u, v) - uv$, by well-known properties of the l.c.m operator. Hence, applying the continuous mapping theorem, under the l.f.c.,

$$T_n \to_d T_\infty \text{ on the extended Skorohod's space in } D[0, 1]^2,$$

where $T_\infty := \mathcal{T} K_\infty - K_\infty$. The stochastic process T_∞ seems to be new in the literature.

The properties of $T_\infty(u,\cdot)$, with $u \in [0,1]$ fixed, have been studied by Groeneboom (1983), amongst others.

Test statistics can be some suitable functional of T_n, like other tests based on empirical processes. We propose to use the $sup - norm$, i.e the Kolmogorov-Smirnov criteria. That is, the test statistic is

$$\tau_n = \|T_n\|_\infty , \tag{5}$$

where, henceforth, with some abuse of notation we denote by $\|\cdot\|_\infty$ the $sup - norm$ in the corresponding space of functions. For instance, for any generic function $f : [0,1]^2 \to \mathbb{R}$, $\|f\|_\infty = \sup_{(u,v)\in[0,1]^2} |f(u,v)|$. Notice that T_n is a positive function.

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

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

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 $i = 1,...,n$, one can use the Pool-Adjacent-Violators (PAV) algorithm described in Barlow et al. (1972, p.13), which is already implemented in many statistical software packages such as R.

The results in Deheuvels (1981a, 1981b) and continuity of $\mathcal{T}$ 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 $\alpha - level$ of significance rejects H_0 if $\tau_n > \tau_{n\alpha}$, where $\tau_{n\alpha} := \inf\{t \in \mathbb{R} : \mathbb{P}(\tau_n \le t |\, l.f.c.) \ge 1-\alpha\}$ is the $(1-\alpha) - quantile$ of τ_n in the l.f.c. Since $\tau_{n\alpha}$ is difficult to calculate analytically, it is approximated by Monte Carlo as accurately as desired. Table I reports the approximated critical values of τ_n for different sample sizes based on 50,000 Monte Carlo simulations.

TABLE I ABOUT HERE

The asymptotic test rejects H_0 at the $\alpha - level$ of significance if $\tau_n > \tau_{\infty\alpha}$, where $\lim_{n\to\infty} \Pr[\tau_n > \tau_{\infty\alpha}|\, l.f.c.] = \alpha$. Next theorem justifies that the tests has the appropriate level under the following mild condition.

Assumption A1: The sequence $\{(Y_i, X_i),\ i = 1,...,n\}$ is an iid sample, distributed as (Y,X). The cdfs F_X and F_Y are continuous.

Theorem 1 *Under H_0 and Assumption A1,*

$$\Pr(\tau_n > \tau_{n\alpha}) \leq \alpha.$$

Moreover,

$$\lim_{n\to\infty} \Pr(\tau_n > \tau_{\infty\alpha}) \leq \alpha.$$

Next Theorem states that the proposed test is able to detect a large class of alternatives, including local alternatives converging to the null at the parametric rate $n^{-1/2}$. The following assumption is needed to ensure the weak convergence of the empirical copula processes K_n under general local alternative hypotheses; see Gänssler and Stute (1987).

ASSUMPTION A2: Under the local alternatives $\{(Y_{i,n}, X_{i,n}),\ i = 1, ..., n\}$ is a sequence of iid arrays for each $n \geq 1$, with continuous marginal cdfs $F_X^{(n)}$ and $F_Y^{(n)}$ and a continuously differentiable copula function.

Notice that in order to justify the behaviour of the test under general local alternatives we do need more smoothness than assumed in Theorem 1. As discussed in Fermanian, Radulovic and Wegkamp (2004, Theorem 4) Assumption A2 is minimal for weak convergence of the copula process.

Theorem 2 *Under the alternative hypothesis and Assumption A1,*

$$\lim_{n\to\infty} \Pr(\tau_n > \tau_{n\alpha}) = 1.$$

If in addition, Assumption A2 holds, then for any $\beta \in (0,1)$ there is some $\gamma > 0$ such that

$$\lim_{n\to\infty} \inf \Pr(\tau_n > \tau_{\infty\alpha}) \geq \beta,$$

provided $\lim_{n\to\infty} \inf \sqrt{n}\, \|\mathcal{T} D_n - D_n\|_\infty > \gamma$, *where* $D_n(u,v) = \mathbb{E}[C_n(u,v)]$, *with the expectation taken under A2.*

Theorem 2 shows that our test is consistent against fixed alternatives and is able to detect local alternatives of the form

$$H_{1n} : \mathcal{T} D_n(u,v) = D_n(u,v) + \frac{a(u,v)}{\sqrt{n}},\ \ (u,v) \in [0,1]^2,$$

with $a : [0,1]^2 \to \mathbb{R}^+$ such that $\|a\|_\infty > \gamma$. Note that these local alternatives are not necessarily local to the l.f.c. but could be local to hypotheses where $F_{Y|X}$ is strictly monotonic with respect to X. This consistency property against $\sqrt{n}-$local alternatives is not shared by LLW's test. Next section investigates the finite-sample properties of the proposed test.

3 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, Seen and van der Vaart (2001) to test for monotonicity in the whole conditional distribution rather than just in the regression function. Their test is based on the U-process

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

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 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 suggest to compute critical values by an approximation to the asymptotic distribution, as in Ghosal, Seen and van der Vaart (2001). 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 their 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 (DGP). 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:

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: $Y_i = 0.2X_i - 0.2 \exp\left(-250(X_i - 0.5)^2\right) + \varepsilon_i$.

Models N1 and ALT1 were considered in LLW, whereas the rest of models have been used in the isotonic regression literature, see Durot (2003) and references therein. We compare LLW's test with ours. Table 2 reports the proportion of rejections in 1,500 Monte Carlo replications of the two tests at 5% of significance under the six designs and with sample sizes $n = 50$, 200 and 500. The results with other nominal levels were similar, and hence, they are not reported.

TABLE II ABOUT HERE

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 null hypothesis but different from the l.f.c. Hence, as expected, the proportion of rejection 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-ALT4. These alternatives suggest that our test based on τ_n can be complementary to LLW's test. In Figure 1(a) we plot the regression function corresponding to ALT4. We observe that this alternative is relatively close to the null hypothesis.

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

ALT5: $Y_i = a1_{\{X_i \leq 0.5\}}(X_i - 0.5)^3 - \exp\left(-250(X_i - 0.5)^2\right) + \varepsilon_i$,

where $\{\varepsilon_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$ are as in the previous simulations. ALT5 represents a model on the alternative hypothesis which becomes farther away from the l.f.c. as $a \to \infty$. In Figure 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.

Figure 1 ABOUT HERE

In Figure 2, we plot the empirical rejection probabilities for ALT5, based on 1500 Monte Carlo replications at 5% nominal level and sample size $n = 300$. 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 the LLW's test shows a flat power at the nominal level.

Figure 2 ABOUT HERE

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, should be appealing to practitioners.

4 Final remarks and extensions

We have proposed a test for the monotonicity of a conditional distribution function, which is pivotal under fairly primitive assumptions, without resorting to smooth estimators of first derivatives. With slightly more efforts, our basic framework can be extended to other interesting situations presented below.

Our procedure can be extended to the case of nonparametric tests of the hypothesis

$$H_0^\gamma : \mathbb{E}\left(\gamma\left(Y,X\right)\middle|\, X=\cdot\right) \in \mathcal{M}$$

for some given function $\gamma : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}$. This includes monotonicity tests for the regression, conditional variances and other conditional moments. In this situation, tests are based on continuous functionals of the empirical process

$$T_n^\gamma := \sqrt{n}\left(\mathcal{T}C_n^\gamma - C_n^\gamma\right),$$

where

$$C_n^\gamma\left(v\right) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\gamma\left(Y_i,X_i\right) - \bar{\gamma}_n\right)1_{\{F_{Xn}(X_i)\le v\}},\ v \in [0,1]\,,$$

with $\bar{\gamma}_n := n^{-1}\sum_{i=1}^{n}\gamma\left(Y_i,X_i\right)$. The l.f.c corresponds now to mean independence, i.e. $\mathbb{E}\left(\gamma\left(Y,X\right)\middle|\, X=\cdot\right) = \mathbb{E}\left(\gamma\left(Y,X\right)\right)$ a.s. Similarly to our Theorem 1 and using standard results in e.g. Stute (1997), it can be shown that if $\mathbb{E}\left(\gamma^2\left(Y,X\right)\right) < \infty$ and F_X is continuous, under the l.f.c,

$$C_n^{\gamma} \to_d W^{\gamma} \text{ on the extended Skorohod's space in } D[0,1],$$

where $W^{\gamma}(v) \stackrel{d}{=} B\left(\tau_{\gamma}^2(v)\right) - vB\left(\tau_{\gamma}^2(1)\right)$, $\tau_{\gamma}^2(v) := \mathbb{E}\left((\gamma(Y,X) - \mathbb{E}(\gamma(Y,X)))^2 1_{\{F_X(X)\leq v\}}\right)$, $v \in [0,1]$ and B is the standard Brownian Motion on $[0,1]$. The test statistic is $\tau_n^{\gamma} := \|T_n^{\gamma}\|_{\infty}$.

Also, note that, unlike τ_n, τ_n^{γ} is in general no longer distribution-free under the l.f.c, even asymptotically.[5] However, the critical values of the test based on τ_n^{γ} can be generally approximated with the assistance of bootstrap using resamples $\{(Y_i^*, X_i)\}_1^n$ with $Y_i^* = \bar{\gamma}_n + V_i(Y_i - \bar{\gamma}_n)$ for a sequence $\{V_i\}_1^n$ of iid variables with zero mean and unit variance, draw independently of $\{(Y_i, X_i)\}_1^n$.

In some applications, we may be interested in testing monotonicity of $F_{Y|X}$ on a strict subset $\mathcal{K} \subset \mathcal{Y} \times \mathcal{X}$. Assume for simplicity that $\mathcal{K} = [l_y, u_y] \times [l_x, u_x]$, and define $\mathcal{S} := [F_Y(l_y), F_Y(u_y)] \times [F_X(l_x), F_X(u_x)]$. To handle this case, we could use as test statistic $\tau_n^{\mathcal{S}} = \sup_{(u,v)\in\mathcal{S}} |T_n(u,v)|$. Since the l.f.c in this case does not entail full independence of Y and X, the test is not anymore distribution-free, even asymptotically, and some approximation of the asymptotic critical values is needed. A convenient resampling process in this case is the subsampling approximation, see Politis, Romano and Wolf (1999). In subsampling the test statistic is computed over the $\binom{n}{b}$ different possible subsamples of size b (taken without replacement from the original data), and the empirical distribution of the resulting sample of test statistics is used to approximate the original test statistic's distribution. Theorem 4 in Fermanian, Radulovic and Wegkamp (2004) proves that, under Assumption A2, K_n converges weakly to a Gaussian process with zero mean in $D[\mathcal{S}]$. The limit distribution of $\tau_n^{\mathcal{S}}$ under the l.f.c is absolutely continuous because it is a functional of a Gaussian process whose covariance function is nonsingular, see Lifshits (1982). Hence, Theorem 2.2.1 of Politis, Romano and Wolf (1999) justifies the validity of the subsampling approximation.

Another important extension is to allow for multivariate explanatory variables. Consider a $1{+}d{-}valued$ vector of r.v.'s $(Y, \mathbf{X})$ taking values in $\mathcal{Y}\times\underline{\mathcal{X}} \subseteq \mathbb{R}^{1+d}$, with $\mathbf{X} = \left(X^{(1)}, ..., X^{(d)}\right)$ and $\underline{\mathcal{X}} \equiv \mathcal{X}^{(1)} \times ... \times \mathcal{X}^{(d)} \subseteq \mathbb{R}^d$. We may be interested in testing monotonicity with respect to a particular coordinate, the $j-th$ say, i.e. testing that a partial effect for $X^{(j)}$ is always negative, or positive. This hypothesis can be written, for a given $j \in \{1, .., d\}$, as

$$H_0^{(j)} : F_{Y|\mathbf{X}}\left(y|\,\mathbf{x}^{(-j)}, \cdot\right) \in \mathcal{M} \text{ for each } \left(y, \mathbf{x}^{(-j)}\right) \in \mathcal{Y}\times\underline{\mathcal{X}}^{(-j)}$$

[5] An important example for which the asymptotic distribution-free property still holds is when $\gamma(Y,X)$ is binary. Under the l.f.c, the model is conditionally homoskedastic, and a suitable standardization of τ_n^{γ} becomes asymptotically distribution-free. As an application of this situation, consider Aguirregabiria's (2010) structural model of dynamic discrete choice. His identification strategy requires agents' choice probabilities that are strictly increasing in a covariate.

where we use the notation $\mathbf{x}^{(-j)}$ to denote the subvector of $\mathbf{x} = \left(x^{(1)}, ..., x^{(d)}\right)$ that excludes $x^{(j)}$ and $\underline{\mathcal{X}}^{(-j)} = \prod_{\ell \neq j, \ell=1}^{d} \mathcal{X}^{(\ell)}$ its corresponding support. Hence, $H_0^{(j)}$ can also be expressed as (3), in terms of the multivariate copula function

$$C\left(u, \mathbf{v}\right) := F\left(F_Y^{-1}\left(u\right), F_{X^{(1)}}^{-1}\left(v^{(1)}\right), ..., F_{X^{(d)}}^{-1}\left(v^{(d)}\right)\right), \ \left(u, \mathbf{v}\right) \in [0,1]^{1+d},$$

where F is the joint distribution of $(Y, \mathbf{X})$ and $\mathbf{v} = \left(v^{(1)}, ..., v^{(d)}\right)$. In this situation, $\mathcal{T}^{(j)}C$ denotes the function obtained by applying the l.c.m. operator $\mathcal{T}^{(j)}$ to the function C, for each $\left(u, \mathbf{v}^{(-j)}\right) \in [0,1]^d$ fixed. Given a random sample $\{Y_i, \mathbf{X}_i\}_{i=1}^n$, $\mathbf{X}_i = \left(X_i^{(1)}, ..., X_i^{(d)}\right)$, C is estimated by its sample analog, as in (4),

$$C_n\left(u, \mathbf{v}\right) := \frac{1}{n} \sum_{i=1}^{n} 1_{\{F_{Yn}(Y_i) \leq u\}} \prod_{\ell=1}^{d} 1_{\left\{F_{X^{(\ell)}n}\left(X_i^{(\ell)}\right) \leq v^{(\ell)}\right\}},$$

resulting in the extension to the multiple explanatory variable case of the test statistic in (5)

$$\tau_n^{(j)} := \left\|T_n^{(j)}\right\|_\infty,$$

where $T_n^{(j)} := \sqrt{n}\left(\mathcal{T}^{(j)}C_n - C_n\right)$. The computational burden increases with the number of explanatory variables considered. The test statistic is not distribution free when $d > 1$ under the l.f.c., which consists now of the conditional independence between Y_i and $X_i^{(j)}$, given $\mathbf{X}_i^{(-j)}$. However, the test can be implemented with the assistance of the subsampling method described above.

The extension to testing stochastic semimonoticity in the sense of Manski (1997) is also straightforward. The stochastic semimonotonicity hypothesis with d explanatory variables is stated as

$$\bar{H}_0^{(d)} : F_{Y|\mathbf{X}}\left(y|\cdot\right) \in \bar{\mathcal{M}}^{(d)} \text{ for each } y \in \mathcal{Y},$$

were

$$\bar{\mathcal{M}}^{(d)} = \left\{ \begin{array}{l} m : \underline{\mathcal{X}} \subset \mathbb{R}^d \to \mathbb{R} \text{ s.t. } m\left(\mathbf{x}'\right) \geq m\left(\mathbf{x}''\right) \text{ if } x^{(j)\prime} \leq x^{(j)\prime\prime} \text{ for all} \\ \qquad j = 1, ..., d \text{ and } \mathbf{x}' = \left(x^{(1)\prime}, ..., x^{(d)\prime}\right), \ \mathbf{x}'' = \left(x^{(1)\prime\prime}, ..., x^{(d)\prime\prime}\right) \in \underline{\mathcal{X}} \end{array} \right\}.$$

It is straightforward to prove that $\bar{H}_0^{(d)}$ can be alternatively written as

$$\bar{H}_0^{(d)} : \mathcal{T}^{(j)}C \equiv C \text{ for each } j = 1, ..., d,$$

which suggests that one can use the following test statistic,

$$\boldsymbol{\tau}_n = \max_{1 \le j \le d} \tau_n^{(j)}.$$

The asymptotic critical values of $\boldsymbol{\tau}_n$ can be approximated using the subsampling procedure discussed above. These extensions to multivariate explanatory variables naturally apply to stochastic semimonotonicity of general conditional moments.

5 Appendix: Proofs of the main results

Proof of Theorem 1: Define $G_n = C_n - C$. Then, by definition of 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, for each $u \in [0,1]$, under H_0, 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,

$$\begin{aligned} T_n &= \sqrt{n}\left(\mathcal{T}C_n - C_n\right) \\ &\le \sqrt{n}\left(\mathcal{T}G_n - G_n\right) \\ &: = \tilde{T}_n \end{aligned} \tag{6}$$

When $C(u,v) = uv$, it holds that $\mathcal{T}G_n(u,v) = \mathcal{T}C_n(u,v) - uv$, $(u,v) \in [0,1]^2$, by well-known properties of the l.c.m operator. So (6) becomes equality. Hence,

$$\Pr\left(\tau_n > \tau_{n\alpha}\right) \le \Pr\left(\tilde{\tau}_n > \tau_{n\alpha} \mid l.f.c\right) \le \alpha,$$

where $\tilde{\tau}_n := \left\|\tilde{T}_n\right\|_\infty$, and

$$\lim_{n\to\infty} \Pr\left(\tau_n > \tau_{\infty\alpha}\right) \le \lim_{n\to\infty} \Pr\left(\tilde{\tau}_n > \tau_{\infty\alpha} \mid l.f.c\right) = \alpha,$$

where the last equality follows from the continuous mapping theorem.

Proof of Theorem 2: Assumption A1, Glivenko-Cantelli's theorem and the continuous mapping theorem imply $\|C_n - C\|_\infty \to_{a.s.} 0$ as $n \to \infty$. Likewise, $\|\mathcal{T}\left(C_n - C\right)\|_\infty \to_{a.s.} 0$ as $n \to \infty$, since by well-known properties of the l.c.m operator, there exists a constant A such that $\|\mathcal{T}\left(C_n - C\right)\|_\infty \le A\|C_n - C\|_\infty$. Hence, under fixed alternatives, $\|\mathcal{T}C_n - C_n\|_\infty$ converges to $\|\mathcal{T}C - C\|_\infty > 0$. Hence, τ_n diverges to $+\infty$, and the test is consistent.

To prove the second part of the theorem, we note that, uniformly,

$$\begin{aligned} T_n &= \sqrt{n}\left(\mathcal{T}D_n - D_n\right) + \sqrt{n}\left(\mathcal{T}C_n - \mathcal{T}D_n - C_n + D_n\right) \\ &= \sqrt{n}\left(\mathcal{T}D_n - D_n\right) + O_P(1). \end{aligned}$$

The $O_P(1)$ term follows from the weak uniform convergence of $\sqrt{n}\left(C_n - D_n\right)$. To see this convergence, notice that by Example 2.11.8 in van der Vaart and Wellner (1996, p. 210) the standard bivariate empirical process

$$\alpha_n\left(y,x\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[1_{\{Y_{i,n}\leq y\}}1_{\{X_{i,n}\leq x\}} - \mathbb{E}\left(1_{\{Y_{i,n}\leq y\}}1_{\{X_{i,n}\leq x\}}\right)\right],$$

converges weakly in $D\left[-\infty,\infty\right]^2$. Now, the weak convergence of $\sqrt{n}\left(C_n - D_n\right)$ follows from the functional delta-method as in Fermanian, Radulovic and Wegkamp (2004, Theorem 3).

References

[1] 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.

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

[3] Bowman, A.W., M.C. Jones and I. Gijbels (1988): "Testing monotonicity of regression", *Journal of Computations and Graphical Statistics,* **7,** 489-500.

[4] Brunk, H.D. (1958): "On the estimation of parameters restricted by inequalities", *Annals of Mathematical Statistics,* **29,** 437-454.

[5] Chernoff, H. (1964): "Estimation of the mode", *Annals of the Institute of Statistical Mathematics*, **16**, 31-41.

[6] Chernozhukov, V. and Hong, H. (2004): "Likelihood estimation and inference in a class of nonregular econometric models", *Econometrica*, **72**, 1445–1480.

[7] Deheuvels, P. (1981a): "A Kolmogorov-Smirnov type test for independence and multivariate samples", *Revue Roumaine de Mathematiques Pures et Appliquees,* **26**, 213-226.

[8] Deheuvels, P. (1981b): "A nonparametric test for independence", *Publications de l'Institut de Statistique de l'Université Paris,* **26**, 29-50.

[9] Durot, C. (2003): "A Kolmogorov-type test for monotonicity of regression", *Statistics & Probability Letters,* **63**, 425-433.

[10] Fermanian, J.D., Radulovic, D. and Wegkamp, M. (2004): "Weak convergence of empirical copula processes", *Bernoulli,* **10,** 847-860.

[11] Gänssler, P. and Stute, W. (1987): *Seminar on Empirical Processes*. DMV Seminar. Birkhäuser, Basel.

[12] Ghosal, S., A. Sen and W. van der Vaart (2000): "Testing monotonicity of regression", *The Annals of Statistics,* **28**, 1054-1082.

[13] Grenander, U. (1956): "On the theory of mortality measurement. Part II", *Skandinavisk Aktuarietidskrift*, **39,** 125-153.

[14] Groeneboom, P. (1983): "The concave majorant of Brownian Motion", *The Annals of Probability*, **11,** 1016-1027.

[15] Hall, P. and E. Heckman (2000): "Testing for monotonicity of a regression mean by calibrating for linear functions", *The Annals of Statistics,* **28**, 20-39.

[16] Hartigan, J.A. and P.M. Hartigan (1985): "The dip test of unimodality", *The Annals of Statistics,* **13**, 70-84.

[17] Lee, S., O. Linton and Y-J Whang (2009): "Testing for stochastic monotonicity", *Econometrica,* **77,** 585-602.

[18] Lifshits, M. A. (1982) "On the absolute continuity of distributions of functionals of random processes," *Theory of Probability and Its Applications,* **27**, 600-607.

[19] Manski, C. F. (1997): "Monotone Treatment Response," *Econometrica*, **65**, 1311–1334.

[20] Matzkin, R.L. (1994): "Restrictions of economic theory in nonparametric methods", *Handbook of Econometrics*, Vol. 4, pp. 2523-2558, edited by R.F. Engel and D. McFadden.

[21] Mammen, E. (1991): "Estimating a smooth nonparametric regression function", *The Annals of Statistics,* **19**, 724-740.

[22] Murkerjee, H. (1988): "Monotone nonparametric regression", *The Annals of Statistics,* **16**, 741-750.

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

[24] Prakasa Rao, B.L.S. (1969): "Estimation of a unimodal density", *Sankhyá (A),* **31,** 23-36.

[25] Schlee, W. (1982): "Nonparametric tests of the monotony and convexity of regression", In *Nonparametric Statistical Inference, Vol. 2, 823-836,* B.V. Gnedenko, M.L. Puri and I. Vincze eds., North Holland, Amnsterdam.

[26] Stute, W. (1997): "Nonparametric model checks for regression", *The Annals of Statistics*, **25**, 613-641.

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

[28] Zinde-Walsh, V. (2008): "Kernel estimation when density may not exist," *Econometric Theory*, **24**, 696-725.

Table I

Simulated Critical Values of τ_n based on 50000 MC simulations.

α/n	**10**	**25**	**50**	**100**	**200**	**500**	**1000**
0.10	0.759	0.783	0.792	0.800	0.806	0.811	0.811
0.05	0.791	0.840	0.848	0.861	0.864	0.870	0.872
0.01	0.885	0.947	0.970	0.980	0.980	0.988	0.993

Table II

Rejection Frequencies at 5%. 1500 MC simulations.

Model	n	τ_n	$LLW_{n,0.4}$	$LLW_{n,0.5}$	$LLW_{n,0.6}$	$LLW_{n,0.7}$
N1	50	0.045	0.020	0.024	0.032	0.034
	200	0.056	0.027	0.028	0.031	0.033
	500	0.048	0.036	0.043	0.045	0.044
N2	50	0.004	0.004	0.003	0.003	0.006
	200	0.000	0.000	0.004	0.012	0.023
	500	0.000	0.000	0.002	0.012	0.044
ALT1	50	0.511	0.672	0.742	0.764	0.749
	200	0.997	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000
ALT2	50	0.436	0.121	0.190	0.264	0.325
	200	0.911	0.550	0.760	0.862	0.920
	500	0.999	0.949	0.994	0.999	1.000
ALT3	50	0.090	0.048	0.062	0.061	0.054
	200	0.281	0.259	0.238	0.227	0.201
	500	0.744	0.648	0.609	0.570	0.512
ALT4	50	0.012	0.014	0.016	0.019	0.032
	200	0.170	0.022	0.016	0.014	0.010
	500	0.806	0.052	0.021	0.008	0.008

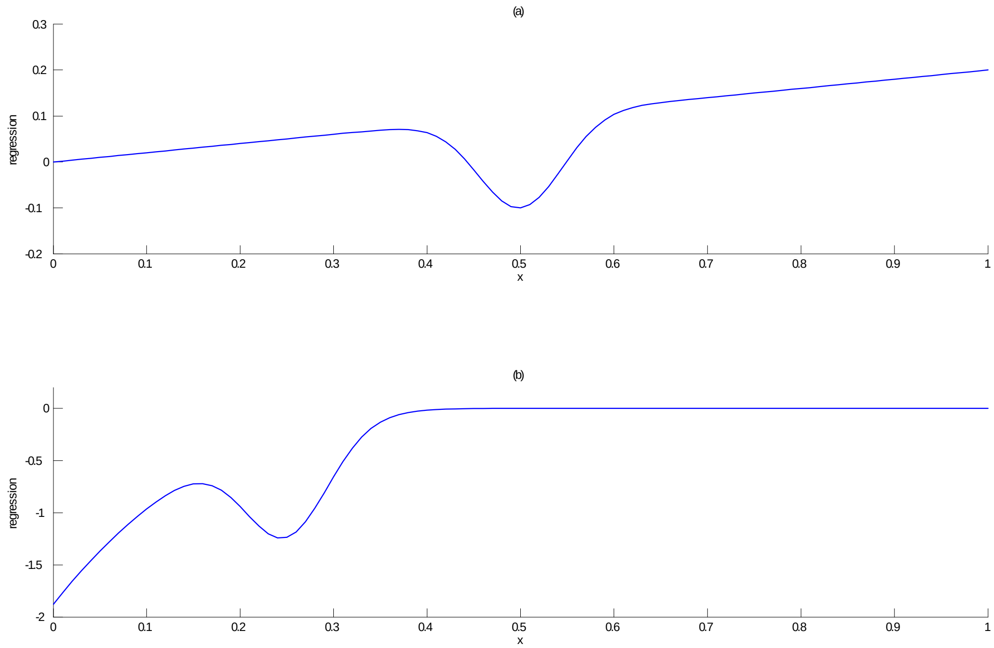

Figure 1. Regression functions for alternatives ALT4 (top panel) and ALT5 (bottom panel) with $a = 15$.

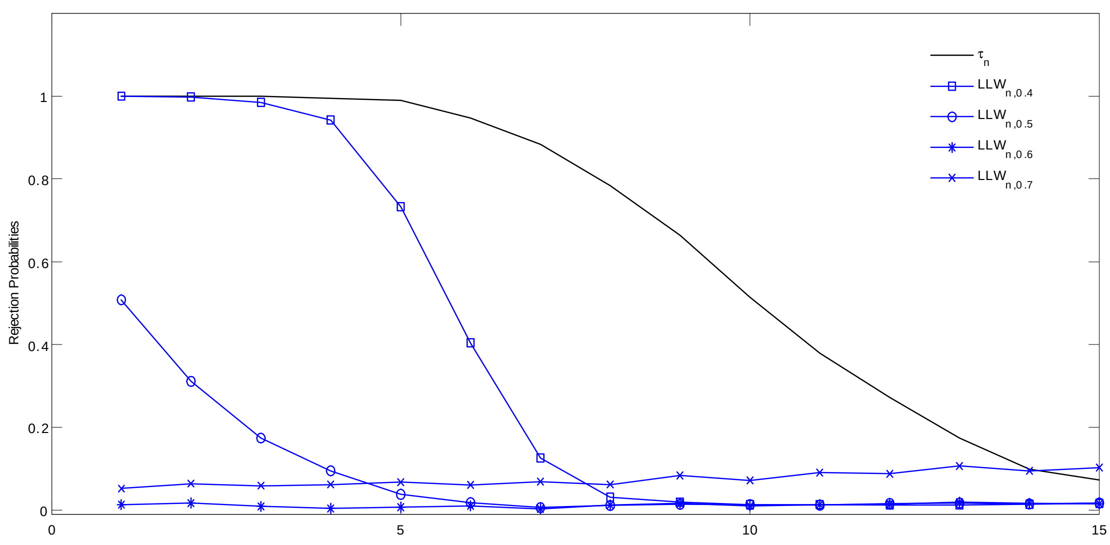

Figure 2. Rejection probabilities for ALT5 as a function of a. 1500 Monte Carlo simulations. Sample size $n = 300$.