



Working Paper  
Economic Series 10-17  
March 2010

Departamento de Economía  
Universidad Carlos III de Madrid  
Calle Madrid, 126  
28903 Getafe (Spain)  
Fax (34) 916249875

# Testing Conditional Monotonicity in the Absence of Smoothness<sup>1</sup>

Miguel A. Delgado<sup>2</sup>      Juan Carlos Escanciano<sup>3</sup>  
Universidad Carlos III de Madrid    Indiana University

March 2nd, 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 the probability densities do not exist. In fact, we only require continuity of the marginal distributions. Distinguishing features of our approach are that the test statistic is pivotal under the null and invariant to any monotonic continuous transformation of the explanatory variable in finite samples. 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  $n^{-1/2}$ ; like the classical goodness-of-fit tests. The article also discusses restricted estimation procedures under monotonicity and extensions of the basic framework to general conditional moments, estimated parameters and multivariate explanatory variables. The finite sample performance of the test is examined by means of a Monte Carlo experiment.

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

**JEL classification:** C14, C15

<sup>1</sup> This research is funded by Spanish Plan Nacional de I+D+I grant number SEJ2007-62908.

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

<sup>3</sup> Economics Department, Indiana University, Bloomington, Indiana, USA; E-mail: jescanci@iu.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

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

where  $F_{Y|X}$  is the conditional distribution function of  $Y$  given  $X$  and, henceforth,  $F_\xi$  denotes the marginal cumulative distribution function (cdf) of the generic random variable (r.v.)  $\xi$ . 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

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

where

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

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 procedure can also be applied to testing monotonicity in a subset of  $\mathcal{Y} \times \mathcal{X}$ . The discussion and results below obviously apply to the monotonically non-decreasing case *mutatis mutandi*.

Notice that, when  $X$  is a dichotomous random variable, the null hypothesis is the classical stochastic dominance hypothesis using two samples, e.g. when comparing distributions between treatment and control groups. Thus,  $H_0$  can be interpreted as the generalization of the Smirnov's two sample test, or stochastic dominance test, to an arbitrary (infinite) number of samples.

Testing monotonicity is interesting, first of all, because estimators of nonparametric monotonic curves can be obtained without imposing smoothness restrictions, which may be hard to test in practice. 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 Wang (2009), LLW henceforth, generalizing 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 others, can be derived from an economic model and how these restrictions can be used for identification and estimation of nonparametric curves.

The LLW and Ghosal et al. (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 derivative. 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. These tests are not valid when the underlying nonparametric curve is not smooth enough. Also, testing the required smoothness, prior to implementing the monotonicity test, is often too involved and may lead into pretest problems. Therefore, testing monotonicity in the absence of smoothness is well motivated. The tests proposed in this article can be implemented only assuming that the marginal distributions of  $Y$  and  $X$  are continuous.

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

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

where  $F_\xi^{-1}$  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  $F_\xi$ . 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

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

Therefore, since  $F_X^{-1}(\bar{v})$  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 also used the difference between the empirical integrated regression function and its l.c.m. for testing monotonicity of a regression curve in a fixed regressors set up with independent and identically distributed (iid) errors.

Estimates of the l.c.m. of the copula process are used in this article for testing monotonicity in the context of general conditional models, only assuming continuity of the marginal distributions. Distinguishing features of our approach are that the test statistic is pivotal under the null and invariant to any monotonic continuous transformation of the explanatory variable in finite samples. Our proposal permits to relax different smoothness assumptions on the underlying nonparametric curves imposed by the LLW and related tests. Also, the performance of our test does not depend on the choice of a smoothing number and we are able to study its power in the direction of local alternatives converging to the null at the parametric rate  $n^{-1/2}$ .

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, a discussion on restricted estimation procedures under monotonicity, indications on how to implement the test in the presence of estimated parameters and the extension to a vector of explanatory variables, were we consider monotonicity with respect to only one coordinate and the hypothesis of stochastic semimonotonicity, in the sense of Manski (1997). A technical mathematical appendix at the end of the article contains the proofs of the results presented in the article.

## 2 Testing monotonicity of a conditional distribution

Given a random sample  $\{(Y_i, X_i), i = 1, \dots, n\}$  of  $(Y, X)$ , the natural estimator of  $C(u, v)$  is

$$C_n(u, v) := \frac{1}{n} \sum_{i=1}^n 1_{\{F_{Y_n}(Y_i) \leq u\}} 1_{\{F_{X_n}(X_i) \leq v\}}, \quad (u, v) \in [0, 1]^2, \quad (4)$$

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

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

where  $K_\infty$  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,  $K_\infty$  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

$$\hat{K}_n := \sqrt{n}(\mathcal{T}C_n - C_n).$$

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,  $\hat{K}_n \equiv \mathcal{T}K_n - K_n$ , 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 l.c.m. Hence, applying the continuous mapping theorem, under the l.f.c.

$$\hat{K}_n \rightarrow_d \hat{K}_\infty \text{ on the extended Skorohod's space in } D[0, 1]^2,$$

where  $\hat{K}_\infty := \mathcal{T}K_\infty - K_\infty$ . The l.c.m. of a Brownian Motion has been studied by Groeneboom (1983) amongst others.

Test statistics can be some suitable functional of  $\hat{K}_n$ , like other tests based on empirical

processes. We propose to use the *sup-norm*, i.e the Kolmorov-Smirnov criteria. That is, the test statistic is

$$\tau_n = \left\| \hat{K}_n \right\|_{\infty}, \quad (5)$$

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  $f : [0, 1]^2 \rightarrow \mathbb{R}$ ,  $\|f\|_{\infty} = \sup_{(u,v) \in [0,1]^2} |f(u, v)|$ . Notice that  $\hat{K}_n$  is a positive function.

The results in Deheuvels (1981a, 1981b) and continuity of  $\mathcal{T}$  imply that the finite sample distribution of  $\hat{K}_n$  is pivotal 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  $\tau_{n\alpha} := \inf\{t \in \mathbb{R} : \mathbb{P}(\tau_n \leq t | l.f.c.) \geq 1 - \alpha\}$  is the  $(1 - \alpha)$ -quantile of  $\tau_n$  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  $\tau_n$  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  $\lim_{n \rightarrow \infty} \Pr[\tau_n > \tau_{\infty\alpha} | l.f.c.] = \alpha$ . Next theorem justifies that the tests have the appropriate level under the following mild condition.

**ASSUMPTION A1:** The sequence  $\{(Y_i, X_i), i = 1, \dots, n\}$  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,

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

If we are interested in testing monotonicity of  $F_{Y|X}$  on a subset of  $\mathcal{S} \subseteq \mathcal{Y} \times \mathcal{X}$  we should suitably modify the sup-norm on the desired subset. The test statistic would be  $\tau_n^{\mathcal{S}} = \sup_{(u,v) \in \mathcal{S}} \left| \hat{K}_n(u, v) \right|$ , which critical values can be approximated by Monte Carlo.

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 both the null and (local) alternative hypotheses; see Gänssler and Stute (1987).

ASSUMPTION A2: Under the local alternatives  $\{(Y_{i,n}, X_{i,n}), i = 1, \dots, n\}$  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.

**Theorem 2** *Under Assumption A2, for any  $\beta \in (0, 1)$  there is some  $\gamma > 0$  such that*

$$\liminf_{n \rightarrow \infty} \Pr(\tau_n > \tau_{\infty\alpha}) \geq \beta,$$

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

Theorem 2 applies to both, fixed and local, alternatives. We first show that our Theorem 2 implies the consistency of our test for fixed alternatives. Under the alternative hypothesis and Assumption A2,  $\|D_n - C_n\|_\infty \rightarrow_{a.s.} 0$  as  $n \rightarrow \infty$ , by Glivenko-Cantelli's theorem and the continuous mapping theorem. Likewise,  $\|\mathcal{T}(D_n - C_n)\|_\infty \rightarrow_{a.s.} 0$  as  $n \rightarrow \infty$ , since by well-known properties of the l.c.m, there exists a constant  $A$  such that  $\|\mathcal{T}(D_n - C_n)\|_\infty \leq A \|D_n - C_n\|_\infty$ . Hence, under fixed alternatives  $\|\mathcal{T}D_n - D_n\|_\infty$  is close to  $\|\mathcal{T}C_n - C_n\|_\infty$ , which in turn converges to a positive constant. Thus, we can apply Theorem 2 to any  $\beta \in (0, 1)$ , which proves that the test is consistent against any fixed alternative.

Theorem 2 also shows that our test is able to detect local alternatives of the form

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

with  $a : [0, 1]^2 \rightarrow \mathbb{R}^+$  such that  $\|a\|_\infty > \gamma$ . Note that these local alternatives are not necessarily local to the l.f.c. but in the interior. This consistency against local alternatives in the “interior” of the null hypothesis is confirmed in our simulations below.

### 3 Monte Carlo

We carried out a simulation study to demonstrate the finite-sample performance of the proposed test, in comparison with the 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

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

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  $h \rightarrow 0$  as  $n \rightarrow \infty$ . 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

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

for a suitable standardized factor  $c_n(x) = n^{-1/2}\hat{\sigma}_n(x)$ . Their test rejects for large values of  $\hat{\mathcal{U}}_n$ . Notice that the values of the test statistic  $\hat{\mathcal{U}}_n$  may change under monotonic continuous transformations of the explanatory variable  $X$ , while  $\tau_n$  is always invariant under such transformations for each  $n$ . Under  $H_0$ ,  $\hat{\mathcal{U}}_n$  is asymptotically distributed as an extreme value random variable and the level accuracy is poor in finite samples. This is why LLW suggest to compute critical values by an approximation to the asymptotic distribution, as in Ghosal et al. (2001), rather than the asymptotic distribution itself. We refer the reader to LLW's article for an explicit expression of the test rejection region. We report results using their choice for the kernel function and consider the Epanechnikov kernel  $k(u) = 0.75(1 - u^2)$ , and the 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  $\{\varepsilon_i\}_1^n$  be a sequence of iid  $N(0, 0.1^2)$  random variables, and let  $\{X_i\}_1^n$  be a sequence of iid  $U[0, 1]$  variables, independent of the sequence  $\{\varepsilon_i\}_1^n$ . Then, the sample  $\{Y_i\}_1^n$  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:**  $Y_i = -0.1 \exp(-250(X_i - 0.5)^2) + \varepsilon_i$ .

**ALT4:**  $Y_i = 0.2X_i - 0.2 \exp(-250(X_i - 0.5)^2) + \varepsilon_i$ .

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 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  $\tau_n$  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 the interior of the null hypothesis and, as expected, the proportion of rejection 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  $\tau_n$  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 in the interior of the null hypothesis.

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

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

where  $\{\varepsilon_i\}_1^n$  and  $\{X_i\}_1^n$  are as in the previous simulations. ALT5 represents a model on the alternative hypothesis which becomes more far away from the l.f.c. as  $a \rightarrow \infty$ . In Figure 1(b) we plot the regression function corresponding to  $a = 15$ . From this plot we observe that this represents another local alternative close to the interior of 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 tests only have 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,  $\tau_n$  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  $a = 15$ ,  $\tau_n$  is able to detect this alternative, whereas the LLW's tests show 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 competing alternatives 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 distribution free under fairly primitive assumptions, without resorting to smooth estimators of first derivatives.

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

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

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

$$\hat{K}_n^\gamma := \sqrt{n}(\mathcal{T}C_n^\gamma - C_n^\gamma),$$

where

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

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

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

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

Also, note that, unlike  $\tau_n$ ,  $\tau_n^\gamma$  is no longer distribution-free under the l.f.c, even asymptotically. However, the critical values of the test based on  $\tau_n^\gamma$  can be 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  $\{(Y_i, X_i)\}_1^n$ . Details are omitted.

Once  $H_0$  in (2) is not rejected, nonparametric estimators of the conditional moments can be obtained without imposing further smoothness assumptions on the underlying nonparametric curves. That is, we can estimate nonparametrically  $F_{Y|X}$ , extending the work of Prakasa Rao (1969) and Brunk (1970) among others, by considering the estimator,

$$F_{nY|X}(y|x) := \arg \min_{m \in \mathcal{M}} \sum_{i=1}^n [1_{\{Y_i \leq y\}} - m(y, X_i)]^2 1_{\{X_i \leq x\}}.$$

This is in fact the slope of  $\mathcal{TC}_n(F_X(x), F_Y(y))$  with respect to  $y$ , which can be readily computed from

$$F_{nY|X}(y|X_{R_i}) = \max_{s \leq i} \min_{t \geq i} \sum_{j=s}^t 1_{\{Y_{R_i} \leq y\}},$$

where  $\{R_i\}_{i=1}^n$  is the sequence of  $X$ -ranks, i.e.  $X_{R_1} \leq X_{R_2} \leq \dots \leq X_{R_n}$ . Alternative monotone estimators can be constructed by monotone rearranging an smoothed estimator, as recently suggested by Chernozhukov, Fernandez-Val and Galichon (2009). Our estimator  $F_{nY|X}$  complements existing methods, as it does not require smoothness of the underlying conditional distribution  $F_{Y|X}$ . Reasoning as Brunk (1970), it can be proved that  $F_{nY|X}(y|x)$ , with  $(y, x)$  fixed, is  $n^{1/3}$ -consistent. The convergence rate can be improved, when it is known that  $F_{Y|X}$  is smooth enough, by smoothing  $F_{nY|X}$ , as proposed by Murkherjee (1988) for isotonic non-parametric regression. See also Mammen (1991) for an study of the efficiency gains. A thorough study of the properties of these estimators is beyond the scope of this article and it is left for future work.

In some circumstances, it may be interesting to apply the test to fitted values or residuals depending on estimated parameters, rather than to raw data. In these cases the test statistics are no longer distribution-free, even asymptotically. This is the case in most tests using empirical processes depending on estimated parameters, see e.g. Durbin (1973). However, the critical values of the tests can be approximated with the assistance of bootstrap using resamples  $\{Y_i^*, X_i\}_{i=1}^n$ , where  $\{Y_i^*\}_{i=1}^n$  is either, a naïve resample or a random permutation of  $\{Y_i\}_{i=1}^n$ . The bootstrap can be justified in the lines of other tests using empirical processes depending on estimated parameters, e.g. Andrews (1997).

Another important extension is to 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  $\mathbf{X} = (X^{(1)}, \dots, X^{(d)})$  and  $\underline{\mathcal{X}} \equiv \mathcal{X}^{(1)} \times \dots \times \mathcal{X}^{(d)} \subseteq \mathbb{R}^d$ . 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  $j \in \{1, \dots, d\}$ , as

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

where we use the notation  $\mathbf{x}^{(-j)}$  to denote the subvector of  $\mathbf{x} = (x^{(1)}, \dots, x^{(d)})$  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

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

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

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

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

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

where  $\hat{K}_n^{(j)} := \sqrt{n} (\mathcal{T}^{(j)}C_n - C_n)$ . 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 independence between  $Y$  and the vector  $\mathbf{X}$ , except in the unlikely case where all the explanatory variables in  $\mathbf{X}$  are independent. However, the test can be implemented with the assistance of the bootstrap using resamples  $\{Y_i^*, \mathbf{X}_i\}_{i=1}^n$ , where  $\{Y_i^*\}_{i=1}^n$  is either a naïve resample or a random permutation of  $\{Y_i\}_{i=1}^n$ .

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

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

where

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

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

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

which suggests to use as test statistic

$$\tau_n = \max_{1 \leq j \leq d} \tau_n^{(j)},$$

which asymptotic critical values can be approximated using the bootstrap procedure dis-

cussed above. These extensions to multivariate explanatory variables naturally apply to stochastic semimonotonicity of conditional moments.

## 5 Appendices

### 5.1 Appendix A: Computation of the test statistic

This appendix contains formulae for computing the test statistic. Following well-known algorithms for computing the classical Kolmogorov-Smirnov tests, we compute  $\tau_n$  as

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

where  $C_n(i/n, 0) \equiv 0$ . Hence, all that is needed 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. A Matlab algorithm to compute  $\mathcal{T}C_n(i/n, \cdot)$  for each  $i = 1, \dots, n$ , and our test statistic is available from the authors upon request.

### 5.2 Appendix B: 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  $H_0$ ,

$$\begin{aligned} \hat{K}_n &= \sqrt{n}(\mathcal{T}C_n - C_n) \\ &\leq \sqrt{n}(\mathcal{T}G_n - G_n) \\ &: = \tilde{K}_n \end{aligned} \tag{6}$$

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. So (6) becomes equality. Hence,

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

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

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

where the last equality follows from the continuous mapping theorem.

**Proof of Theorem 2:** It follows from the proof of Theorem 1 that, uniformly,

$$\begin{aligned}\hat{K}_n &= \sqrt{n}(\mathcal{T}D_n - D_n) + \sqrt{n}(\mathcal{T}C_n - \mathcal{T}D_n - C_n + D_n) \\ &= \sqrt{n}(\mathcal{T}D_n - D_n) + O_P(1).\end{aligned}$$

The  $O_P(1)$  term follows from the weak uniform convergence of  $\sqrt{n}(C_n - D_n)$ . 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

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

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

## References

- [1] ANDREWS, D.W.K. (1997): "A conditional Kolmogorov test", *Econometrica*, **65**, 1097-1128.
- [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] BRUNK, H.D. (1970): "Estimation of isotonic regression", In *Nonparametric Techniques in Statistical Inference*, 177-195, M.L. Puri eds., Cambridge University Press, New York.
- [6] CHERNOZHUKOV, V., FERNANDEZ-VAL, I. AND GALICHON, A. (2009): "Quantile and Probability Curves without Crossing", forthcoming in *Econometrica*.
- [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] DURBIN, J. (1973): "Weak convergence of the sample distribution function when parameters are estimated", *The Annals of Statistics*, **1**, 279-290.
- [10] DUROT, C. (2003): "A Kolmogorov-type test for monotonicity of regression", *Statistics & Probability Letters*, **63**, 425-433.
- [11] FERMANIAN, J.D., RADULOVIC, D. AND WEGKAMP, M. (2004): "Weak convergence of empirical copula processes", *Bernoulli*, **10**, 847-860.
- [12] GÄNSSLER, P. AND STUTE, W. (1987): *Seminar on Empirical Processes*. DMV Seminar. Birkhäuser, Basel.
- [13] GHOSAL, S., A. SEN AND W. VAN DER VAART (2000): "Testing monotonicity of regression", *The Annals of Statistics*, **28**, 1054-1082.
- [14] GRENANDER, U. (1956): "On the theory of mortality measurement. Part II", *Skandinavisk Aktuarietidskrift*, **39**, 125-153.
- [15] GROENEBOOM, P. (1983): "The concave majorant of Brownian Motion", *The Annals of Probability*, **11**, 1016-1027.
- [16] 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.
- [17] HARTIGAN, J.A. AND P.M. HARTIGAN (1985): "The dip test of unimodality", *The Annals of Statistics*, **13**, 70-84.
- [18] LEE, S., O. LINTON AND Y-J WHANG (2009): "Testing for stochastic monotonicity", *Econometrica*, **77**, 585-602.
- [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] PRAKASA RAO, B.L.S. (1969): "Estimation of a unimodal density", *Sankhyá (A)*, **31**, 23-36.
- [24] 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, Amsterdam.
- [25] VAN DER VAART, A. W. AND WELLNER, J. A (1996): *Weak Convergence and Empirical Processes*. Springer, New York.



**Table I**Simulated Critical Values of  $\tau_n$  based on 50000 MC simulations.

$\alpha/n$	<b>10</b>	<b>25</b>	<b>50</b>	<b>100</b>	<b>200</b>	<b>500</b>	<b>1000</b>
<b>0.10</b>	0.759	0.783	0.792	0.800	0.806	0.811	0.811
<b>0.05</b>	0.791	0.840	0.848	0.861	0.864	0.870	0.872
<b>0.01</b>	0.885	0.947	0.970	0.980	0.980	0.988	0.993

**Table II**

Rejection Frequencies at 5%. 1500 MC simulations.

Model	$n$	$\tau_n$	$LLW_{n,0.4}$	$LLW_{n,0.5}$	$LLW_{n,0.6}$	$LLW_{n,0.7}$
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

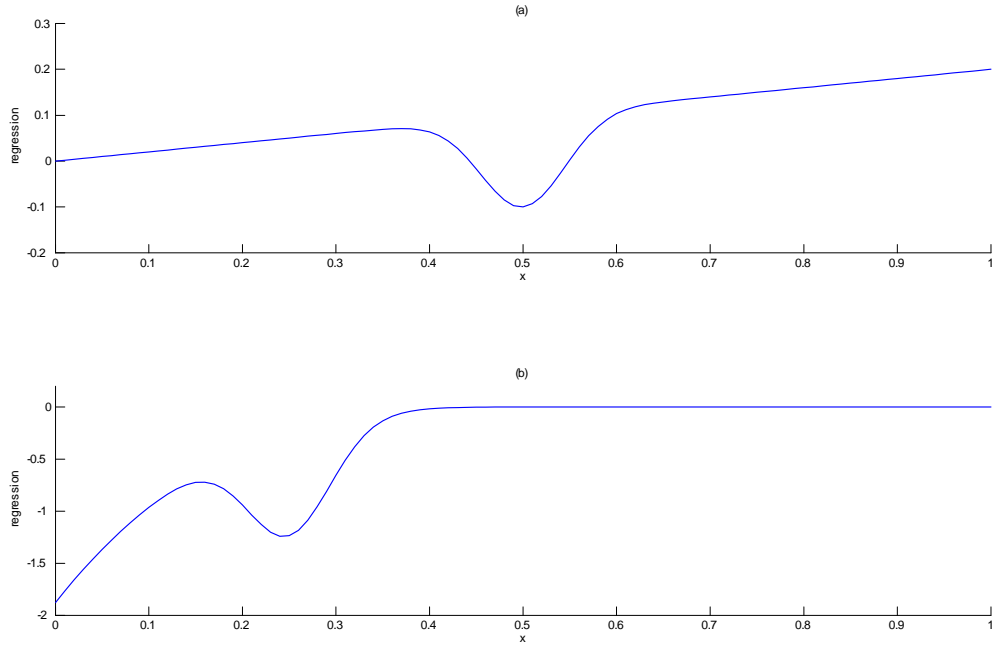


Figure 1. Regression functions for alternatives ALT4 (top panel) and ALT5 (bottom panel) with  $a = 15$ .

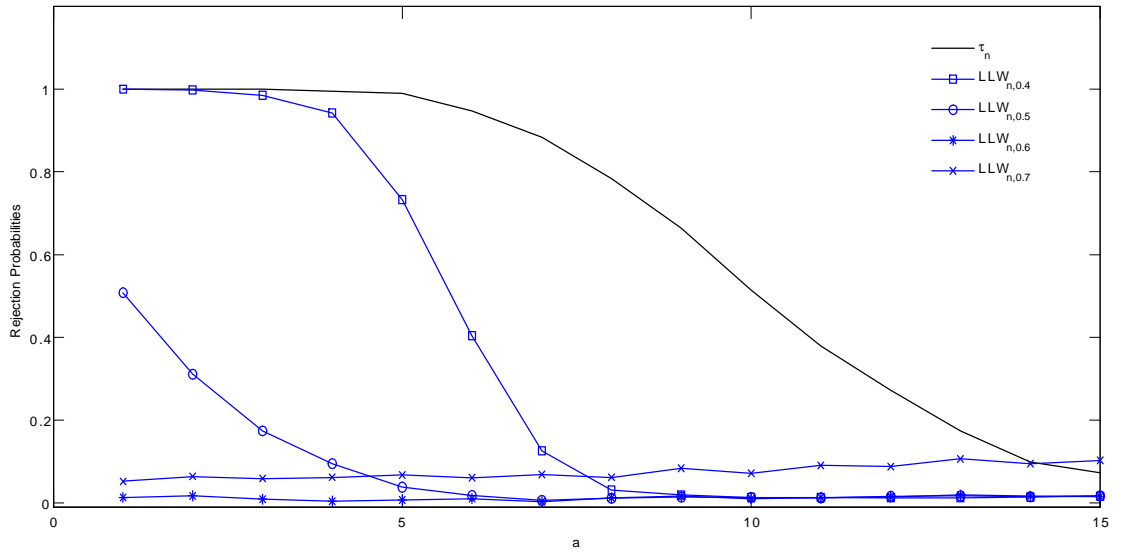


Figure 2. Rejection probabilities for ALT5 as a function of  $a$ . 1500 Monte Carlo simulations. Sample size  $n = 300$ .