

# ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

*An International Society for the Advancement of Economic  
Theory in its Relation to Statistics and Mathematics*

<http://www.econometricsociety.org/>

*Econometrica*, Vol. 77, No. 2 (March, 2009), 585–602

TESTING FOR STOCHASTIC MONOTONICITY

SOKBAE LEE

*University College London, London WC1E 6BT, U.K.*

OLIVER LINTON

*London School of Economics, London WC2A 2AE, U.K.*

YOON-JAE WHANG

*Seoul National University, Seoul 151-742, Korea*

---

The copyright to this Article is held by the Econometric Society. It may be downloaded, printed and reproduced only for educational or research purposes, including use in course packs. No downloading or copying may be done for any commercial purpose without the explicit permission of the Econometric Society. For such commercial purposes contact the Office of the Econometric Society (contact information may be found at the website <http://www.econometricsociety.org> or in the back cover of *Econometrica*). This statement must be included on all copies of this Article that are made available electronically or in any other format.

---

## TESTING FOR STOCHASTIC MONOTONICITY

BY SOKBAE LEE, OLIVER LINTON, AND YOON-JAE WHANG<sup>1</sup>

We propose a test of the hypothesis of stochastic monotonicity. This hypothesis is of interest in many applications in economics. Our test is based on the supremum of a rescaled  $U$ -statistic. We show that its asymptotic distribution is Gumbel. The proof is difficult because the approximating Gaussian stochastic process contains both a stationary and a nonstationary part, and so we have to extend existing results that only apply to either one or the other case. We also propose a refinement to the asymptotic approximation that we show works much better in finite samples. We apply our test to the study of intergenerational income mobility.

KEYWORDS: Distribution function, extreme value theory, Gaussian process, monotonicity.

### 1. INTRODUCTION

LET  $Y$  AND  $X$  DENOTE two random variables whose joint distribution is absolutely continuous with respect to Lebesgue measure on  $\mathbb{R}^2$ . Let  $F_{Y|X}(\cdot|x)$  denote the distribution of  $Y$  conditional on  $X = x$ . This paper is concerned with testing the stochastic monotonicity of  $F_{Y|X}$ . Specifically, we consider the hypothesis

$$(1) \quad H_0: \text{For each } y \in \mathcal{Y}, F_{Y|X}(y|x) \leq F_{Y|X}(y|x') \text{ whenever } x \geq x' \text{ for } x, x' \in \mathcal{X},$$

where  $\mathcal{Y}$  and  $\mathcal{X}$ , respectively, are the supports of  $Y$  and  $X$ . We propose a test statistic and obtain asymptotically valid critical values. We are not aware of any existing test for (1) in the literature.

This hypothesis can be of interest in a number of applied settings. If  $X$  is some policy, dosage, or other input variable, one might be interested in testing whether its effect on the distribution of  $Y$  is increasing in this sense. Also, one can test whether stochastic monotonicity exists in well-known economic relationships such as expenditures ( $Y$ ) versus incomes ( $X$ ) at household levels, wages ( $Y$ ) versus cognitive skills ( $X$ ) using individual data, outputs ( $Y$ ) versus the stock of capital ( $X$ ) at the country level, sons' incomes ( $Y$ ) versus fathers' incomes ( $X$ ) using family data, and so on.

<sup>1</sup>We would like to thank a co-editor, three anonymous referees, John Einmahl, Jerry Hausman, Hidehiko Ichimura, Steve Machin, Charles Manski, and participants at several conferences and seminars for useful suggestions and helpful comments. Lee extends thanks to the ESRC and the Leverhulme Trust through funding of the Centre for Microdata Methods and Practice and of the research programme Evidence, Inference and Inquiry. Linton thanks the ESRC for financial support. Whang thanks the Korea Research Foundation Grant funded by the Korean government (MOEHRD) (KRF-2005-041-B00074).

The notion of stochastic monotonicity is important in instrumental variables estimation. Manski and Pepper (2000) have introduced monotone instrumental variables assumptions that hold when the average outcome varies monotonically across the levels of instrumental variables. Small and Tan (2007) have used the stochastic monotonicity condition that does not require that a monotonic increasing relationship hold within individuals, thus allowing for “defiers” in treatments.

Blundell, Gosling, Ichimura, and Meghir (2007) have recently adopted this hypothesis and obtained tight bounds on an unobservable cross-sectional wage distribution, thus allowing them to characterize the evolution of its inequality over time. Specifically, they assumed that the distribution of wages  $W$  for employed given observed characteristics  $X$  and an instrument  $Z$  is increasing in  $Z$ . Their instrument was the out of work income. They derived a bound on the implied distribution of wages given characteristics under this assumption of stochastic monotonicity. They also suggested a test of this hypothesis based on the implied bounds, using the bootstrap to calculate critical values. They found that the hypothesis was not rejected on their data at standard significance levels; indeed the  $p$ -values were very high. They did not provide any theory to justify their critical values and, moreover, did not test the monotonicity hypothesis itself, but an implication of it.

This concept arises often in dynamic economic models. Thus suppose that  $Y = Y_{t+1}$ ,  $X = Y_t$ , and  $Y_t$  is a Markov process so that  $F_{Y|X} = F_{t+1|t}$  is the transition measure of the process  $Y_t$ . In that case the property, along with mild technical conditions, implies that the process has a stationary distribution. The influential monograph of Lucas and Stokey (1989) used the stochastic monotonicity property frequently in solving dynamic optimization problems of the Markov type and characterizing the properties of the solution. It is particularly important in problems where nonconvexities give rise to discontinuous stochastic behavior and it provides a route to proving the existence of stationary equilibria without requiring smoothness. Hopenhayn and Prescott (1992) argued that it arises “in economic models from the monotonicity of decision rules or equilibrium mappings that results from the optimizing behaviour of agents.” Pakes (1986) assumed that the distribution of the return to holding a patent conditional on current returns was nonincreasing in current returns. Consequently he showed that the optimal renewal policy took a very simple form based on the realization of current returns compared with the cost of renewing. Ericson and Pakes (1995), Olley and Pakes (1996), and Buettner (2003) have all used a similar property in various dynamic models of market structures. It is possible to test these restrictions with our methods given suitable data.

Testing stochastic monotonicity can be relevant for testing the existence of firms’ strategic behaviors in industrial organization. Recently, Ellison and Ellison (2007) have shown that under some suitable conditions, investment levels are monotone in market size if firms are not influenced by strategic entry deterrence and are nonmonotone if influenced by a desire to deter entry. Ellison

and Ellison (2007) have also developed a couple of monotonicity tests, based on Hall and Heckman (2000), and have implemented them using pharmaceutical data. In addition to the tests used in Ellison and Ellison (2007), our test can be adopted to test the existence of strategic entry deterrence.

We propose a simple test of hypothesis (1) for observed or (partially) estimated independent and identically distributed (i.i.d.) data. Our statistic is based on the supremum of a rescaled second-order  $U$ -process indexed by two parameters  $x$  and  $y$  (Nolan and Pollard (1987)). It generalizes the corresponding statistic introduced by Ghosal, Sen, and van der Vaart (2000) for testing the related hypothesis of monotonicity of a regression function. Our first contribution is to prove that the asymptotic distribution of our test statistic is Gumbel with certain nonstandard norming constants, thereby facilitating inference using critical values obtained from the limiting distribution. We also show that the test is consistent against all alternatives. The proof technique is quite complicated and novel because the approximating Gaussian stochastic process contains both a stationary part (corresponding to  $x$ ) and a nonstationary part (corresponding to  $y$ ), and so we have to extend existing results that apply only to one or the other case. For example, Stute (1986) established the weak convergence to a Brownian bridge of a conditional empirical process (effectively holding  $x$  constant in our problem). In the other direction, using the local strong invariance principle of Rio (1994), Ghosal, Sen, and van der Vaart (2000) established local (in  $x$  in our notation) weak convergence of an empirical process to a stationary limit, generalizing the seminal work of Bickel and Rosenblatt (1973). The most closely related work to ours is Beirlant and Einmahl (1996), who considered the asymptotics of some functional of a conditional empirical process except that they considered a maximum over a discrete partition of the support of the covariate. See also Einmahl and Van Keilegom (2008). We use some results of Piterbarg (1996) to establish the approximation. These results can be of use elsewhere. See Appendix A.1 of the electronic supplement to this article (Lee, Linton, and Whang (2009)) for some informal discussion on the proof technique.

One known issue with the extreme value limiting distributions is the poor quality of the asymptotic approximation in the sense that the error declines only at a logarithmic (in sample size) rate. The usual approach to this has been to use the bootstrap, which provides an asymptotic refinement by removing the logarithmic error term and giving an error of polynomial order (Hall (1993)). In a special case of ours (of a stationary Gaussian process), Piterbarg (1996) provided a higher order analytic approximation to the limiting distribution that involves including the (known) logarithmic factor in the first-order error. His Theorem G1 shows that this corrected distribution is closer to the actual distribution and indeed has an error of polynomial (in sample size) magnitude. We apply this analysis to our more complicated setting and compute the corresponding "correction" term. Our simulation study shows that this approach gives a noticeable improvement in size. An alternative approach is to use a

standard bootstrap resample applied to the (recentered) statistic (or a bootstrap resample imposing independence between  $Y$  and  $X$ ) to improve the size of the test, motivated by the reasoning of Hall (1993). This method should also yield an asymptotic refinement (Horowitz (2001)), but is much more time consuming than using the asymptotic critical values.

The hypothesis (1) implies that the regression function  $E(Y|X = x)$ , when it exists, is monotonic increasing. It also implies that all conditional quantile functions are increasing. It is a strong hypothesis, but can be reduced in strength by limiting the set of  $X$  and  $Y$  for which this property holds. See, for example, Bowman, Jones, and Gijbels (1998), Hall and Heckman (2000), Ghosal, Sen, and van der Vaart (2000), and Gijbels, Hall, Jones, and Koch (2000) for existing tests of the hypothesis that  $E(Y|X = x)$  is increasing in  $x$ . Note that the transformation regression model structure considered in Ghosal, Sen, and van der Vaart (2000) that is,  $\phi(Y) = m(X) + \varepsilon$ , where  $\varepsilon$  is independent of  $X$ , and both  $\phi$  and  $m$  are monotonic functions, actually implies stochastic monotonicity. See also Ekeland, Heckman, and Nesheim (2004). Also, a test of the hypothesis (1) can be viewed as a continuum version of the stochastic dominance test (see Linton, Maasoumi, and Whang (2005) and references therein for details on the stochastic dominance test).

The remainder of the paper is organized as follows. Section 2 defines our test statistic and Section 3 states the asymptotic results and describes how to carry out the test. Section 4 contains results of some Monte Carlo experiments. Section 5 illustrates the usefulness of our test by applying it to the study of intergenerational income mobility. Section 6 considers a multivariate extension and Section 7 concludes. All the proofs are given in the electronic supplement to this article (Lee, Linton, and Whang (2009)).

## 2. THE TEST STATISTIC

This section describes our test statistic. Let  $\{(Y_i, X_i) : i = 1, \dots, n\}$  denote a random sample from  $(Y, X)$ . We suppose throughout that the data are i.i.d., but the main result also holds for the Markov time series case where  $Y_i = Y_{i+1}$  and  $X_i = Y_i$ . We actually suppose that  $X_i$  is not observed, but an estimate  $\widehat{X}_i = \psi(W_i, \widehat{\theta})$  is available, where  $X_i = \psi(W_i, \theta_0)$  is a known function of observable  $W_i$  for some true parameter value  $\theta_0$  and  $\widehat{\theta}$  is a root- $n$  consistent estimator thereof. The vector  $W_i$  can contain discrete and continuous variables. Let  $1(\cdot)$  denote the usual indicator function and let  $K(\cdot)$  denote a one-dimensional kernel function with a bandwidth  $h_n$ . Consider the  $U$ -process

$$\begin{aligned} \widehat{U}_n(y, x) &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} [1(Y_i \leq y) - 1(Y_j \leq y)] \operatorname{sgn}(\widehat{X}_i - \widehat{X}_j) \\ &\quad \times K_{h_n}(\widehat{X}_i - x) K_{h_n}(\widehat{X}_j - x), \end{aligned}$$

where  $K_{h_n}(\cdot) = h_n^{-1}K(\cdot/h_n)$  and  $\text{sgn}(x) = 1(x > 0) - 1(x < 0)$ . Note that the  $U$ -process  $\widehat{U}_n(y, x)$  can be viewed as a locally weighted version of Kendall's tau statistic applied to  $1(Y \leq y)$  and that  $\widehat{U}_n(y, x)$  is related to the  $U$ -process considered in Ghosal, Sen, and van der Vaart (2000, equation (2.1)).

Let  $U_n(y, x)$  denote  $\widehat{U}_n(y, x)$  computed using  $X_i$  instead of  $\widehat{X}_i$ . First, notice that under regularity conditions including smoothness of  $F_{Y|X}(y|x)$ , as  $n \rightarrow \infty$ ,

$$h_n^{-1}EU_n(y, x) \rightarrow F_x(y|x) \left( \iint |u_1 - u_2|K(u_1)K(u_2) du_1 du_2 \right) [f_X(x)]^2,$$

where  $F_x(y|x)$  is a partial derivative of  $F_{Y|X}(y|x)$  with respect to  $x$ . Therefore, since  $\widehat{\theta}$  is a consistent estimator, under the null hypothesis such that  $F_x(y|x) \leq 0$  for all  $(y, x) \in \mathcal{Y} \times \mathcal{X}$ ,  $\widehat{U}_n(y, x)$  is less than or equal to zero on average for large  $n$ . Under the alternative hypothesis such that  $F_x(y|x) > 0$  for some  $(y, x) \in \mathcal{Y} \times \mathcal{X}$ , a suitably normalized version of  $\widehat{U}_n(y, x)$  can be very large. In view of this, we define our test statistic as a supremum statistic

$$(2) \quad S_n = \sup_{(y,x) \in \mathcal{Y} \times \mathcal{X}} \frac{\widehat{U}_n(y, x)}{c_n(x)}$$

with some suitably defined  $c_n(x)$ , which may depend on  $(X_1, \dots, X_n)$  but not on  $(Y_1, \dots, Y_n)$ . The  $U$ -statistic structure suggests that we use the scaling factor  $c_n(x) = \widehat{\sigma}_n(x)/\sqrt{n}$ , where

$$\begin{aligned} \widehat{\sigma}_n^2(x) &= \frac{4}{n(n-1)(n-2)} \sum_{1 \leq i \neq j \neq k \leq n} \text{sgn}(\widehat{X}_i - \widehat{X}_j) \text{sgn}(\widehat{X}_i - \widehat{X}_k) \\ &\quad \times K_{h_n}(\widehat{X}_j - x) K_{h_n}(\widehat{X}_k - x) [K_{h_n}(\widehat{X}_i - x)]^2. \end{aligned}$$

REMARK 2.1: (i) An alternative class of test statistics is based on explicit estimation of conditional cumulative distribution functions (c.d.f.'s). Thus, consider  $T_n = \sup_{y \in \mathcal{Y}, x, x' \in \mathcal{X}: x \geq x'} [\widehat{F}_{Y|X}(y|x) - \widehat{F}_{Y|X}(y|x')]$ , where, for example,  $\widehat{F}_{Y|X}(y|x)$  is some kernel estimate of the conditional c.d.f.; see Hall, Wolff, and Yao (1999). The advantage that  $T_n$  has is that it does not require smoothness of  $F_{Y|X}(y|x)$ . The disadvantage is that its limiting distribution is not pivotal and it is not known how to make it so. (ii) One might also be interested in testing second or higher order dominance (Levy (2006)) of the conditional distribution functions, which can be achieved by straightforward modification of either  $S_n$  or  $T_n$ .

REMARK 2.2: In applications one may also be interested in the following extension where there are multiple covariates. Specifically, suppose that  $X$  is replaced by  $X, Z$ , where  $Z$  is a vector, and the hypothesis is that

$$H_0: \text{For each } y \in \mathcal{Y}, F_{Y|X,Z}(y|x, z) \leq F_{Y|X,Z}(y|x', z) \text{ whenever } x \geq x' \text{ for } x, x' \in \mathcal{X} \text{ and } z \in \mathcal{Z}.$$

This hypothesis allows the variable  $Z$  to affect the response in a general way. The hypothesis is nonnested with (1) for the same reason that a conditional independence hypothesis is non-nested with an independence hypothesis; see Dawid (1979) and Phillips (1988). In the case that  $Z$  are discrete random variables, our test statistic can be trivially adapted to test this hypothesis. If  $Z$  included some continuous random variables, then a modified version of our test statistic might work, but its asymptotic distribution would be different.

REMARK 2.3: As an alternative norming constant, one can use  $c_n(x) = \tilde{\sigma}_n(x)/\sqrt{n}$ , where

$$\tilde{\sigma}_n^2(x) = 4h_n^{-1} \left[ \int q^2(u)K^2(u) du \right] \times \hat{f}_X^3(x),$$

$q(u) = \int \text{sgn}(u - w)K(w) dw$ , and  $\hat{f}_X(x)$  is the kernel density estimator of  $f_X(x)$ . It can be shown easily that  $\tilde{\sigma}_n(x)$  is asymptotically equivalent to  $\hat{\sigma}_n(x)$ . In finite samples,  $\hat{\sigma}_n(x)$  may work better than  $\tilde{\sigma}_n(x)$  since the former is based on a more direct sample analog, but the latter is easier to compute.

REMARK 2.4: In some applications, it might be more desirable to assume that  $X_i = \psi(W_i, \theta_0) + \varepsilon_i$  ( $i = 1, \dots, n$ ), where  $\varepsilon_i$  is an unobserved random variable. In this case, our test does not apply with  $\hat{X}_i = \psi(W_i, \hat{\theta})$ . To resolve this case, one could assume certain regularity conditions that ensure that the stochastic monotonicity between  $Y$  and  $\psi(W, \theta_0)$  holds if the stochastic monotonicity between  $Y$  and  $X$  holds (e.g., the monotone likelihood ratio property as in Proposition 2 of Ellison and Ellison (2007)).

### 3. ASYMPTOTIC THEORY

This section provides the asymptotic behavior of the test statistic when the null hypothesis is true and when it is false. In particular, we determine the asymptotic critical region of the test and show that the test is consistent against general fixed alternatives at any level.

3.1. *Distribution of the Test Statistic*

Since the hypothesis (1) is a composite hypothesis, it is necessary to find a case when the type I error probability is maximized asymptotically. First of all, under regularity conditions assumed below, it can be shown that

$$(3) \quad \widehat{U}_n(y, x) - U_n(y, x) = O_p(n^{-1/2}) \quad \text{and} \quad h_n^{1/2} \widehat{\sigma}_n(x) = O_p(1)$$

uniformly over  $(y, x)$  (see Lemmas A.6, A.7, and A.9 of Lee, Linton, and Whang (2009)). Then if  $h_n \rightarrow 0$ ,

$$(4) \quad \widehat{U}_n(y, x) = U_n(y, x)[1 + o_p(1)]$$

uniformly over  $(y, x)$ . Thus, the asymptotic distribution of the test statistic  $S_n$  is the same as if  $X_i$  were observed directly.

Now define

$$\begin{aligned} \tilde{U}_n(y, x) &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} [F_{Y|X}(y|X_i) - F_{Y|X}(y|X_j)] \text{sgn}(X_i - X_j) \\ &\quad \times K_{h_n}(X_i - x) K_{h_n}(X_j - x). \end{aligned}$$

Since  $E[U_n(y, x) - \tilde{U}_n(y, x)|X_1, \dots, X_n] = 0$ , under regularity conditions assumed below, using the empirical process method (see, e.g., Ghosal, Sen, and van der Vaart (2000, Appendix) and van der Vaart and Wellner (1996)), it can be shown that

$$U_n(y, x) - \tilde{U}_n(y, x) = O_p\left[\left(\frac{\log n}{nh_n}\right)^{1/2}\right] \quad \text{and} \quad \tilde{U}_n(y, x) = O_p(h_n)$$

uniformly over  $(y, x)$ . Then if  $\log n / (nh_n^3) \rightarrow 0$ ,

$$(5) \quad U_n(y, x) = \tilde{U}_n(y, x)[1 + o_p(1)]$$

uniformly over  $(y, x)$ .

Under the null hypothesis (1), note that

$$(6) \quad [F_{Y|X}(y|X_i) - F_{Y|X}(y|X_j)] \text{sgn}(X_i - X_j) \leq 0.$$

Hence, by (4), (5), and (6), the type I error probability is maximized asymptotically when  $F_x(y|x) \equiv 0$ , equivalently  $F_{Y|X}(y|x) = F_Y(y)$  for any  $(y, x) \in \mathcal{Y} \times \mathcal{X}$ . Therefore, to derive the limiting distribution under the null hypothesis, we consider the case that  $F_x(y|x) \equiv 0$ , equivalently  $F_{Y|X}(y|x) = F_Y(y)$  for any  $(y, x)$ . That is,  $Y$  and  $X$  are independent. Further assume that without loss of generality, the support of  $X$  is  $\mathcal{X} = [0, 1]$ .

To establish the asymptotic null distribution of the test statistic, we make the following assumptions, which are standard in the literature on nonparametric estimation and testing.

ASSUMPTION 3.1: Assume that (a)  $Y$  and  $X$  are independent; (b)  $\mathcal{X} = [0, 1]$ ; (c) the distribution of  $X$  is absolutely continuous with respect to Lebesgue measure and the probability density function of  $X$  is continuously differentiable and strictly positive in  $\mathcal{X}$ ; (d) the distribution of  $Y$  is absolutely continuous with respect to Lebesgue measure; (e)  $K$  is a second-order kernel function with support  $[-1, 1]$  and is twice continuously differentiable; (f)  $\theta_0$  is a finite-dimensional parameter and  $\|\hat{\theta} - \theta_0\| = O_p(n^{-1/2})$ ; (g) for each  $w$ ,  $\psi(w, \theta)$  is continuously differentiable with respect to  $\theta$ ; (h) for any  $(W_i, x, h_n)$ , there exists a positive constant  $C_L < \infty$  such that  $|\xi(W_i, x, \theta_1, h_n) - \xi(W_i, x, \theta_2, h_n)| \leq C_L \|\theta_1 - \theta_2\|$  for all  $\theta_1$  and  $\theta_2$  in a neighborhood of  $\theta_0$ , where

$$\begin{aligned} &\xi(W_i, x, \theta, h_n) \\ &= \int \operatorname{sgn}[\psi(W_i, \theta) - \psi(\tilde{w}, \theta)] K_{h_n}[\psi(\tilde{w}, \theta) - x] dF_W(\tilde{w}). \end{aligned}$$

To describe the limiting distribution of  $S_n$ , recall that  $q(u) = \int \operatorname{sgn}(u - w)K(w) dw$ . Let  $\beta_n$  be the largest solution to the equation

$$(7) \quad h_n^{-1} \left( \frac{8\lambda}{\pi} \right)^{1/2} \beta_n \exp(-2\beta_n^2) = 1,$$

where

$$(8) \quad \lambda = - \frac{\int q(x)K^2(x)K'(x) dx + \int q^2(x)K(x)K''(x) dx}{\int q^2(x)K^2(x) dx}.$$

The following theorem gives the asymptotic distribution of the test statistic when the null hypothesis is true.

THEOREM 3.1: Let Assumption 3.1 hold. Let  $h_n$  satisfy  $h_n \log n \rightarrow 0$ ,  $nh_n^3/(\log n) \rightarrow \infty$ , and  $nh_n^2/(\log n)^4 \rightarrow \infty$ . Then for any  $x$ ,

$$(9) \quad \Pr(4\beta_n(S_n - \beta_n) < x) = \exp\left\{-\exp\left(-x - \frac{x^2}{8\beta_n^2}\right)\left[1 + \frac{x}{4\beta_n^2}\right]\right\} + o(1).$$

In particular,

$$\lim_{n \rightarrow \infty} \Pr(4\beta_n(S_n - \beta_n) < x) = \exp(-e^{-x}) \equiv F_\infty(x).$$

REMARK 3.1: It is necessary to compute  $\beta_n$  in (7) to construct a test based on Theorem 3.1. The constant  $\lambda$  in (8) can be computed easily for commonly used kernels that are twice differentiable and have compact support. For example,  $\lambda = 1177/118 \approx 9.975$  for the Epanechnikov kernel  $K(u) =$

$0.75(1 - u^2)1(|u| \leq 1)$  and  $\lambda = 131,689/11,063 \approx 11.904$  for the biweight kernel  $K(u) = (15/16)(1 - u^2)^21(|u| \leq 1)$ . It is straightforward to show that

$$(10) \quad \beta_n = \left(\frac{1}{2} \log[h_n^{-1}c^*]\right)^{1/2} + \frac{\log[\frac{1}{2} \log[h_n^{-1}c^*]]}{8(\frac{1}{2} \log[h_n^{-1}c^*])^{1/2}} + o\left[\frac{1}{(\log[h_n^{-1}c^*])^{1/2}}\right],$$

where  $c^* = (8\lambda/\pi)^{1/2}$ . Then one can use an approximation to  $\beta_n$  by the first two terms on the right side of (10) or solve the nonlinear equation (7) numerically.

REMARK 3.2: We note that the regularity conditions on  $h_n$  are not very restrictive. Bandwidth sequences  $h_n$  that converge to zero at a rate of  $n^{-\eta}$ ,  $\eta < 1/3$ , or  $(\log n)^{-\nu}$ ,  $\nu > 1$ , satisfy the conditions imposed in Theorem 3.1. We might also consider data-dependent bandwidths such as provided by cross-validation, for example. That is, let  $\hat{h}_n$  be a data-dependent sequence such that  $\hat{h}_n/h_n \xrightarrow{P} 1$ , where  $h_n$  is a deterministic sequence that satisfies the assumptions of Theorem 3.1. In view of the results in Einmahl and Mason (2005), one expects that under some suitable regularity conditions, the asymptotic distribution of the test  $S_n$  with data-dependent bandwidths  $\hat{h}_n$  is the same as that given in Theorem 3.1. However, it is beyond the scope of this paper to provide such regularity conditions and corresponding proofs.

As in Theorem 4.2 of Ghosal, Sen, and van der Vaart (2000), the theorem suggests that one can construct a test with an asymptotic level  $\alpha$ :

$$(11) \quad \text{Reject } H_0 \text{ if } F_\infty(4\beta_n(S_n - \beta_n)) \geq 1 - \alpha$$

for any  $0 < \alpha < 1$ . Alternatively, one can construct an  $\alpha$ -level test with (9):

$$(12) \quad \text{Reject } H_0 \text{ if } F_n(4\beta_n(S_n - \beta_n)) \geq 1 - \alpha,$$

where for each  $n$ ,  $F_n(x)$  is the “distribution function” of the form<sup>2</sup>

$$F_n(x) = \exp\left\{-\exp\left(-x - \frac{x^2}{8\beta_n^2}\right)\left[1 + \frac{x}{4\beta_n}\right]\right\}.$$

Although (11) yields the correct size asymptotically, the results of Hall (1979, 1991) suggest that  $\Pr[(11) \text{ is true} | H_0] = \alpha + O(1/\log n)$ , which is rather slow for practical purposes. However, the results of Piterbarg (1996, Theorem G1) suggest that  $\Pr[(12) \text{ is true} | H_0] = \alpha + O(n^{-q})$  for some  $q > 0$ , which is potentially much better. In the next section, we carry out Monte Carlo experiments using both critical regions (11) and (12). In our experiments, a test based on

<sup>2</sup>The approach in (12) to defining the critical region is motivated partly by the normalizing transformation approach (Phillips (1979)).

(12) performs much better in finite samples and yields size quite close to the nominal value.

An alternative approach to constructing critical values would be to use a bootstrap resampling method (that imposes independence between  $X$  and  $Y$ ) and then to reject if  $F_\infty(4\beta_n(S_n - \beta_n))$  exceeds the  $1 - \alpha$  critical value of the bootstrap distribution of  $F_\infty(4\beta_n(S_n^* - \beta_n))$ , where  $S_n^*$  is the bootstrapped test statistic (Horowitz (2001)). Hall (1993) showed in the related context of density estimation that a bootstrapped test yields error of order  $n^{-q}$  for some  $q > 0$ . We expect a similar result can be established here. The bootstrap approach is much more computationally demanding than the asymptotic approach outlined above.

We now turn to the consistency of the test. It is straightforward to show that the test specified by (11) or (12) is consistent against general alternatives.

**THEOREM 3.2:** *Assume that  $nh_n^3/\log h_n^{-1} \rightarrow \infty$ . If  $F_x(y|x) > 0$  for some  $(y, x) \in \mathcal{Y} \times \mathcal{X}$ , then the test specified by (11) or (12) is consistent at any level.*

We end this subsection by mentioning that the test and its asymptotic properties obtained in this section can be extended easily to the case when the null hypothesis in (1) holds only for  $\mathcal{Y}$  and  $\mathcal{X}_1$ , where  $\mathcal{X}_1$  is a compact interval and a strict subset of  $\mathcal{X}$ . In this case,  $F_x(y|x) \equiv 0$  does not imply that  $Y$  and  $X$  are independent; however, this would not matter since our test statistic depends only on observations inside an open interval containing  $\mathcal{X}_1$ . Thus, the asymptotic properties of the supremum test statistic would be the same with  $\mathcal{X}_1$  except that  $h_n^{-1}$  in (7) and (10) is replaced with  $\text{measure}(\mathcal{X}_1)/h_n$ .

#### 4. MONTE CARLO EXPERIMENTS

This section presents the results of some Monte Carlo experiments that illustrate the finite-sample performance of the test. For each Monte Carlo experiment,  $X$  was independently drawn from a uniform distribution on  $[0, 1]$ . To evaluate the performance of the test under the correct null hypothesis,  $Y \equiv U$  was generated independently from  $X$ , where  $U \sim \mathbf{N}(0, 0.1^2)$ . In addition, to see the power of the test,  $Y$  was also generated from  $Y = m(X) + U$ , where  $m(x) = x(1 - x)$ . The simulation design considered here is similar to that of Ghosal, Sen, and van der Vaart (2000). To save computing time the test statistic was computed by the maximum of  $\sqrt{n}U_n(y, x)/\tilde{\sigma}_n(x)$  over  $\mathcal{Y} \times \mathcal{X}$ , where  $\mathcal{Y} = \{Y_1, Y_2, \dots, Y_n\}$ ,  $\mathcal{X} = \{0.05, 0.10, \dots, 0.90, 0.95\}$ , and  $\tilde{\sigma}_n(x)$  was defined in Remark 2.3. The kernel function was  $K(u) = 0.75(1 - u^2)$  for  $-1 \leq u \leq 1$ . The simulations used sample sizes of  $n = 50, 100, 200$ , and  $500$ , and all the simulations were carried out in GAUSS using GAUSS pseudo-random number generators. For each simulation, the number of replications was 1500.

Table I reports results of Monte Carlo experiments using critical values obtained from the asymptotic expansion  $F_n$  of the limiting distribution (see (12))

TABLE I  
SIMULATION RESULTS

Sample Size $n$	Bandwidth			
	$h_n = 0.4$	$h = 0.5$	$h = 0.6$	$h = 0.7$
Using critical values obtained from the asymptotic expansion $F_n$ of the limiting distribution (1500 replications in each experiment)				
Rejection proportions when the null hypothesis is true				
50	0.014	0.021	0.025	0.030
100	0.028	0.033	0.034	0.034
200	0.025	0.031	0.036	0.033
500	0.032	0.039	0.033	0.037
Rejection proportions when the null hypothesis is false				
50	0.687	0.762	0.771	0.760
100	0.976	0.988	0.989	0.977
200	1.000	1.000	1.000	1.000
Using critical values obtained from the type I extreme value distribution (1500 replications in each experiment)				
Rejection proportions when the null hypothesis is true				
50	0.009	0.017	0.013	0.017
100	0.022	0.024	0.022	0.021
200	0.015	0.021	0.022	0.021
500	0.021	0.021	0.022	0.023
Rejection proportions when the null hypothesis is false				
50	0.618	0.693	0.697	0.694
100	0.966	0.976	0.983	0.965
200	1.000	1.000	1.000	1.000
Using bootstrap critical values with 500 bootstrap samples (500 replications in each experiment)				
Rejection proportions when the null hypothesis is true				
50	0.064	0.062	0.046	0.058
Rejection proportions when the null hypothesis is false				
50	0.814	0.872	0.880	0.856

and also using those from the type I extreme value distribution (see (11)). The nominal level was 5%. First, consider the first panel of the table that shows results with the critical values from  $F_n$ . When the null hypothesis is true, each rejection proportion is below the nominal level for all the bandwidths and is maximized at  $n = 500$  and  $h_n = 0.5$ . It can be seen that the best  $h_n$  is decreasing with the sample size and the performance of the test is less sensitive to  $h_n$  as  $n$  gets large. When the null hypothesis is false, for all values of  $h_n$ , the powers of the test are high for  $n = 50$ , almost 1 for  $n = 100$ , and 1 for  $n = 200$ . The performance of the test with critical values from the type I extreme value distribution is uniformly worse, as seen from the second panel of the table.

Hence, our simulation study shows that the approximation based on (12) gives a substantial improvement in size.

In addition, Table I gives results with bootstrap critical values. Each bootstrap resample is generated by random sampling of  $Y$  and  $X$  separately with replacement (i.e., imposing independence between  $Y$  and  $X$ ). Because of very lengthy computation times, the Monte Carlo experiments are carried out with only  $n = 50$  and only 500 replications in each experiment. There were 500 bootstrap resamplings for each replication in the Monte Carlo experiment. Not surprisingly, it can be seen that when the null hypothesis is true, the difference between actual and nominal rejection proportions is smaller than either of the asymptotic critical values. As a result, the test with bootstrap critical values has better power. In view of these experiment results, we recommend using bootstrap critical values when the sample size is small or moderate.

##### 5. APPLICATION TO INTERGENERATIONAL INCOME MOBILITY

This section presents an empirical example in which the test statistic  $S_n$  is used to test a hypothesis about the stochastic monotonicity between sons' incomes and parental incomes. See Solon (1999, 2002) for a detailed survey on intergenerational income mobility in the United States and other countries. A large body of this literature focuses on the extent to which sons' incomes are correlated with fathers' or parental incomes. Testing the hypothesis of stochastic monotonicity in (1) with  $Y$  being sons' incomes and  $X$  being parental incomes can give further insights into understanding intergenerational income mobility. For example, if one fails to reject the hypothesis, then that would imply that sons' incomes with high parental incomes are higher not only on average, but also in the stochastic dominance sense, than those with low parental incomes.<sup>3</sup>

We use data from the Panel Study of Income Dynamics (PSID), which has been used frequently to study mobility after it was used in a highly influential paper by Solon (1992). In particular, we use Minicozzi's (2003) data extract that is available on the *Journal of Applied Econometrics* website. The  $Y$  variable 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.<sup>4</sup>

<sup>3</sup>The related paper Dearden, Machin, and Reed (1997) investigated the intergenerational income mobility in Britain using the quantile transition matrix approach.

<sup>4</sup>Minicozzi (2003) computed sons' average incomes only when incomes at both ages 28 and 29 are available and regarded those with only one or no income record as censored observations. In our empirical work, we define sons' average incomes as the average of observed incomes. Hence, sons' average incomes are defined for those with only one income record at age 28 or at age 29. Using our definition, only 12 cases have missing sons' incomes. This is only 2% of the 628 original observations in Minicozzi's (2003) data extract. Hence, censoring is not a serious issue with our definition. Parental permanent incomes are predicted values. The asymptotic distribution of  $S_n$  is the same as long as a parametric model used by Minicozzi (2003, equations (8) and (9)) gives consistent estimates of parental permanent incomes.

Local Linear Quantile Regression Estimates

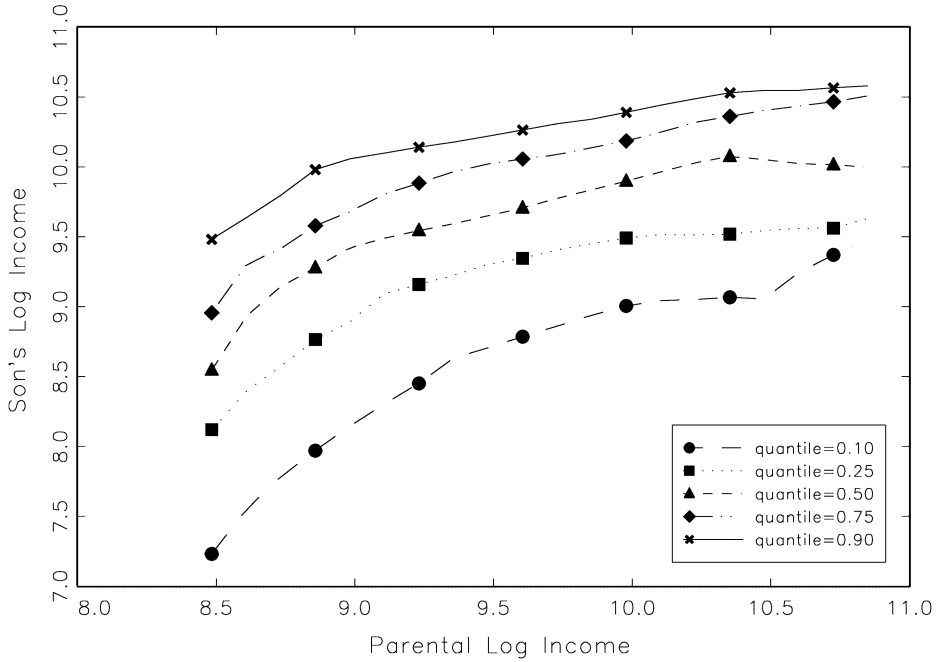


FIGURE 1.—Intergenerational income mobility: local linear quantile regression estimates of sons' log incomes on parental log incomes.

Figure 1 shows local linear quantile regression estimates of sons' log incomes on parental log incomes. The kernel function used in this estimation is the same as that used in the Monte Carlo experiment. For each quantile, the bandwidth is chosen by the simple rule of thumb suggested by Fan and Gijbels (1996, p. 202): 0.59 (quantile = 10%), 0.55 (25%), 0.55 (50%), 0.56 (75%), and 0.69 (90%). It can be seen that all the conditional quantiles of sons' incomes are increasing functions of parental incomes for most of the range of the support of  $X$ . This suggests that there may be stochastic monotonicity between sons' and parental incomes.

To test this formally, the test statistic  $S_n$  is computed by the maximum of  $\sqrt{n}U_n(y, x)/\hat{\sigma}_n(x)$  over  $\mathcal{Y} \times \mathcal{X}$ , where  $\mathcal{Y} = \{Y_1, Y_2, \dots, Y_n\}$  and  $\mathcal{X} = [8.48, 10.85]$ , where the two endpoints of  $\mathcal{X}$  correspond to 1st and 99th percentiles of parental log permanent incomes. The same kernel is used with a bandwidth of  $h_n = 0.55$ , which is used to estimate the local linear median estimator above. The test gives  $S_n = 0.5227$ . The normalizing constant  $\beta_n$  in (7) and (10) is obtained with the assumption that  $\mathcal{X} = [0, 1]$ , but it is trivial to extend this to a more general case,  $\mathcal{X} = [a, b]$ . One just needs to replace  $h_n^{-1}$  in (7) and (10) with  $(b - a)/h_n$ . After this simple modification, the critical values

at 10% nominal level are 1.71 using (11) and 1.72 using (12). Changing the value of the bandwidth to  $0.75h_n$  or to  $1.25h_n$  did not change this conclusion. Thus, we fail to reject the null hypothesis of stochastic monotonicity at any conventional level and this confirms findings from Figure 1.

6. TESTING FOR STOCHASTIC MONOTONICITY IN A VECTOR

In this section, we extend our analysis to the case where monotonicity in a vector is of interest. Let  $\mathbf{X}$  be a  $d$ -dimensional vector of random variables whose distribution is absolutely continuous with respect to Lebesgue measure on  $\mathbb{R}^d$ . We consider the following hypothesis, which is a multivariate generalization of (1):

$$(13) \quad H_0: \text{For each } y \in \mathcal{Y}, F_{Y|\mathbf{X}}(y|\mathbf{x}) \leq F_{Y|\mathbf{X}}(y|\mathbf{x}')$$

whenever  $x_j \geq x'_j$  for all  $j = 1, \dots, d$  and for  $\mathbf{x} \equiv (x_1, \dots, x_d), \mathbf{x}' \equiv (x'_1, \dots, x'_d) \in \mathcal{X}$ .

The hypothesis (13) restricts the stochastic ordering  $F_{Y|\mathbf{X}}(y|\mathbf{x})$  only when all components of  $\mathbf{x}$  are ordered componentwise. In other words, using the terminology of Manski (1997), testing (13) amounts to testing the stochastic semimonotonicity of  $F_{Y|\mathbf{X}}$ . The hypothesis (13) can be of interest in a number of empirical applications. For example,  $Y$  is the output and  $\mathbf{X}$  is a vector of inputs used for production (Manski (1997)).

We now describe a test statistic for (13). Let  $\{(Y_i, \mathbf{X}_i) : i = 1, \dots, n\}$  denote a random sample from  $(Y, \mathbf{X})$ . As in Section 2, we assume that  $\mathbf{X}_i$  is not observed, but  $\widehat{\mathbf{X}}_i$  is estimated with a root- $n$  consistent estimator of  $\widehat{\theta}$ . For  $\mathbf{u} \equiv (u_1, \dots, u_d)$ , let  $\mathbf{K}(\cdot)$  denote a  $d$ -dimensional product of univariate kernel functions  $\mathbf{K}(\mathbf{u}) = \prod_{j=1}^d K(u_j)$  and let  $\mathbf{I}(\mathbf{u} > 0) = \prod_{j=1}^d 1(u_j > 0)$ . Consider the  $U$ -process

$$\begin{aligned} \widehat{\mathbf{U}}_n(y, \mathbf{x}) &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} [1(Y_i \leq y) - 1(Y_j \leq y)] \text{sgn}(\widehat{\mathbf{X}}_i - \widehat{\mathbf{X}}_j) \\ &\quad \times \mathbf{K}_{h_n}(\widehat{\mathbf{X}}_i - \mathbf{x}) \mathbf{K}_{h_n}(\widehat{\mathbf{X}}_j - \mathbf{x}), \end{aligned}$$

where  $\mathbf{K}_{h_n}(\cdot) = h_n^{-d} \mathbf{K}(\cdot/h_n)$  and  $\text{sgn}(\mathbf{x}) = \mathbf{I}(\mathbf{x} > 0) - \mathbf{I}(\mathbf{x} < 0)$ . Note that  $\text{sgn}(\widehat{\mathbf{X}}_i - \widehat{\mathbf{X}}_j)$  has a nonzero value only when semimonotonicity between  $\widehat{\mathbf{X}}_i$  and  $\widehat{\mathbf{X}}_j$  holds.

Again, we define our test statistic as a supremum statistic

$$(14) \quad \mathbf{S}_n = \sup_{(y, \mathbf{x}) \in \mathcal{Y} \times \mathcal{X}} \frac{\widehat{\mathbf{U}}_n(y, \mathbf{x})}{\widehat{\mathbf{c}}_n(\mathbf{x})},$$

with  $\widehat{\mathbf{c}}_n(\mathbf{x}) = \widehat{\mathbf{s}}_n(\mathbf{x})/\sqrt{n}$ , where

$$\widehat{\mathbf{s}}_n^2(\mathbf{x}) = \frac{4}{n(n-1)(n-2)} \sum_{1 \leq i \neq j \neq k \leq n} \text{sgn}(\widehat{\mathbf{X}}_i - \widehat{\mathbf{X}}_j) \text{sgn}(\widehat{\mathbf{X}}_i - \widehat{\mathbf{X}}_k) \times \mathbf{K}_{h_n}(\widehat{\mathbf{X}}_j - \mathbf{x}) \mathbf{K}_{h_n}(\widehat{\mathbf{X}}_k - \mathbf{x}) [\mathbf{K}_{h_n}(\widehat{\mathbf{X}}_i - \mathbf{x})]^2.$$

Under the null hypothesis (13), note that

$$(15) \quad [F_{Y|\mathbf{X}}(y|\mathbf{X}_i) - F_{Y|\mathbf{X}}(y|\mathbf{X}_j)] \text{sgn}(\mathbf{X}_i - \mathbf{X}_j) \leq 0.$$

Thus, using arguments similar to those used in Section 3.1, it can be shown that the type I error probability is maximized asymptotically when the inequality (15) is equality for all  $i$  and  $j$ . This equality occurs when  $Y$  is independent of  $\mathbf{X}$ . Therefore, to derive the limiting distribution under the null hypothesis, we consider the case that  $Y$  and  $\mathbf{X}$  are independent.

ASSUMPTION 6.1: Assume that  $Y$  and  $\mathbf{X}$  are independent and the support of  $\mathbf{X}$  is  $\mathcal{X} = [0, 1]^d$ . Let conditions (c)–(h) of Assumption 3.1 hold.

Let  $\mathbf{b}_n$  be the largest solution to the equation

$$(16) \quad h_n^{-d} 2^{-(d-1)} \left(\frac{8\lambda}{\pi}\right)^{d/2} \mathbf{b}_n^d \exp(-2\mathbf{b}_n^2) = 1,$$

where  $\lambda$  is defined in (8). The following theorem is a generalization of Theorem 6.1.

THEOREM 6.1: Let Assumption 6.1 hold. Let  $h_n$  satisfy  $h_n \log n \rightarrow 0$ ,  $nh_n^{3d}/(\log n) \rightarrow \infty$ , and  $nh_n^{d+1}/(\log n)^{2(d+1)} \rightarrow \infty$ . Then for any  $x$ ,

$$(17) \quad \Pr(4\mathbf{b}_n(\mathbf{S}_n - \mathbf{b}_n) < x) = \exp\left\{-\exp\left(-x - \frac{x^2}{8\mathbf{b}_n^2}\right) \left[1 + \frac{x}{4\mathbf{b}_n^2}\right]^d\right\} + o(1).$$

In particular,

$$\lim_{n \rightarrow \infty} \Pr(4\mathbf{b}_n(\mathbf{S}_n - \mathbf{b}_n) < x) = \exp(-e^{-x}) \equiv F_\infty(x).$$

Then a test with asymptotically valid critical values can be constructed as in Section 3.1. Furthermore, it can be shown that the corresponding test is consistent at any level against fixed alternatives, provided that  $nh_n^{3d}/\log h_n^{-1} \rightarrow \infty$ .

## 7. CONCLUSIONS

We have proposed a test for stochastic monotonicity and have developed the asymptotic null distribution of our test statistic. There remain several research topics we have not addressed in this paper. First, we have only established the consistency of the test against fixed general alternatives. It would be useful to establish asymptotic results regarding local powers of the test. Second, we have not considered an “optimal” choice of the bandwidth used in the test statistic. Our theoretical results for the asymptotic null distribution and the consistency of the test do not distinguish different bandwidths, provided that a sequence of bandwidths satisfies some weak regularity conditions on rates of convergence. Thus, it would be necessary to develop a finer asymptotic result to discuss an optimal bandwidth choice. Doing this and developing a corresponding data-dependent bandwidth choice are topics for future research. Using a method similar to that used in this paper, we can extend the supremum test of Ghosal, Sen, and van der Vaart (2000), who considered the monotonicity of the regression function with a scalar explanatory variable, to the multivariate setup. This is another topic for future research.

## REFERENCES

- BEIRLANT, J., AND J. H. J. EINMAHL (1996): “Maximal Type Test Statistics Based on Conditional Processes,” *Journal of Statistical Planning and Inference*, 53, 1–19. [587]
- BICKEL, P. J., AND M. ROSENBLATT (1973): “On Some Global Measures of the Deviations of Density Function Estimates,” *Annals of Statistics*, 1, 1071–1095. [587]
- BLUNDELL, R., A. GOSLING, H. ICHIMURA, AND C. MEGHIR (2007): “Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds,” *Econometrica*, 75, 323–363. [586]
- BOWMAN, A. W., M. C. JONES, AND I. GIJBELS (1998): “Testing Monotonicity of Regression,” *Journal of Computational and Graphical Statistics*, 7, 489–500. [588]
- BUETTNER, T. (2003): “R&D and the Dynamics of Productivity,” Manuscript, London School of Economics. [586]
- DAWID, A. P. (1979): “Conditional Independence in Statistical Theory,” *Journal of the Royal Statistical Society, Series B*, 41, 1–31. [590]
- DEARDEN, L., S. MACHIN, AND H. REED (1997): “Intergenerational Mobility in Britain,” *Economic Journal*, 107, 47–66. [596]
- EINMAHL, J. H. J., AND I. VAN KEILEGOM (2008): “Specification Tests in Nonparametric Regression,” *Journal of Econometrics*, 143, 88–102. [587]
- EINMAHL, U., AND D. M. MASON (2005): “Uniform in Bandwidth Consistency of Kernel-Type Function Estimators,” *Annals of Statistics*, 33, 1380–1403. [593]
- EKELAND, I., J. J. HECKMAN, AND L. NESHEIM (2004): “Identification and Estimation of Hedonic Models,” *Journal of Political Economy*, 112, S60–S109. [588]
- ELLISON, G., AND S. F. ELLISON (2007): “Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration,” Working Paper, Department of Economics, MIT. [586,587,590]
- ERICSON, R., AND A. PAKES (1995): “Markov-Perfect Industry Dynamics: A Framework for Empirical Work,” *Review of Economic Studies*, 62, 53–82. [586]
- FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall. [597]

- GHOSAL, S., A. SEN, AND A. W. VAN DER VAART (2000): "Testing Monotonicity of Regression," *Annals of Statistics*, 28, 1054–1082. [587-589,593,594,600]
- GIJBELS, I., P. HALL, M. C. JONES, AND I. KOCH (2000): "Tests for Monotonicity of a Regression Mean With Guaranteed Level," *Biometrika*, 87, 663–673. [588]
- HALL, P. (1979): "On the Rate of Convergence of Normal Extremes," *Journal of Applied Probability*, 16, 433–439. [593]
- (1991): "On Convergence Rate of Suprema," *Probability Theory and Related Fields*, 89, 447–455. [593]
- (1993): "On Edgeworth Expansion and Bootstrap Confidence Bands in Nonparametric Curve Estimation," *Journal of the Royal Statistical Society, Series B*, 55, 291–304. [587,588,594]
- HALL, P., AND N. E. HECKMAN (2000): "Testing for Monotonicity of a Regression Mean by Calibrating for Linear Functions," *Annals of Statistics*, 28, 20–39. [587,588]
- HALL, P., R. C. L. WOLFF, AND Q. YAO (1999): "Methods for Estimating a Conditional Distribution Function," *Journal of the American Statistical Association*, 94, 154–163. [589]
- HOPENHAYN, H. A., AND E. C. PRESCOTT (1992): "Stochastic Monotonicity and Stationary Distributions for Dynamic Economics," *Econometrica*, 60, 1387–1406. [586]
- HOROWITZ, J. L. (2001): "The Bootstrap," in *The Handbook of Econometrics*, Vol. 5, ed. by J. J. Heckman and E. Leamer. Amsterdam: North-Holland. [588,594]
- LEE, S., O. LINTON, AND Y.-J. WHANG (2009): "Supplement to 'Testing for Stochastic Monotonicity'," *Econometrica Supplemental Material*, 77, [http://www.econometricsociety.org/ecta/Supmat/7145\\_proofs.pdf](http://www.econometricsociety.org/ecta/Supmat/7145_proofs.pdf); [http://www.econometricsociety.org/ecta/Supmat/7145\\_Data\\_and\\_programs.pdf](http://www.econometricsociety.org/ecta/Supmat/7145_Data_and_programs.pdf). [587,588,591]
- LEVY, H. (2006): *Stochastic Dominance, Investment Decision Making Under Uncertainty* (Second Ed.). Berlin: Springer. [589]
- LINTON, O., E. MAASOUMI, AND Y.-J. WHANG (2005): "Consistent Testing for Stochastic Dominance Under General Sampling Schemes," *Review of Economic Studies*, 72, 735–765. [588]
- LUCAS, R. E., AND N. L. STOKEY (1989): *Recursive Methods in Dynamic Economics*. Cambridge, MA: Harvard University Press. [586]
- MANSKI, C. F. (1997): "Monotone Treatment Response," *Econometrica*, 65, 1311–1334. [598]
- MANSKI, C. F., AND J. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010. [586]
- MINICOZZI, A. L. (2003): "Estimation of Sons' Intergenerational Earnings Mobility in the Presence of Censoring," *Journal of Applied Econometrics*, 18, 291–314. [596]
- NOLAN, D., AND D. POLLARD (1987): "U-Processes: Rates of Convergence," *Annals of Statistics*, 15, 780–799. [587]
- OLLEY, G. S., AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297. [586]
- PAKES, A. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54, 755–784. [586]
- PHILLIPS, P. C. B. (1979): "Normalizing Transformations and Expansions for Functions of Statistics," Research Notes, Birmingham University. [593]
- (1988): "Conditional and Unconditional Statistical Independence," *Journal of Econometrics*, 38, 341–348. [590]
- PITERBARG, V. I. (1996): *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, Translation of Mathematical Monographs, Vol. 148. Providence, RI: American Mathematical Society. [587]
- RIO, E. (1994): "Local Invariance Principles and Their Application to Density Estimation," *Probability Theory and Related Fields*, 98, 21–45. [587]
- SMALL, D., AND Z. TAN (2007): "A Stochastic Monotonicity Assumption for the Instrumental Variables Method," Working Paper, Department of Statistics, Wharton School, University of Pennsylvania. [586]
- SOLON, G. (1992): "Intergenerational Income Mobility in the United States," *American Economic Review*, 82, 393–408. [596]

- (1999): “Intergenerational Mobility in the Labor Market,” in *Handbook of Labor Economics*, Vol. 3A, ed. by O. Ashenfetter and D. Card. Amsterdam: North-Holland, 1761–1800, Chapter 29. [596]
- (2002): “Cross-Country Differences in Intergenerational Earnings Mobility,” *Journal of Economic Perspectives*, 16, 59–66. [596]
- STUTE, W. (1986): “Conditional Empirical Processes,” *Annals of Statistics*, 14, 638–647. [587]
- VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. New York: Springer. [591]

*Dept. of Economics, University College London, Gower Street, London WC1E 6BT, U.K.; l.simon@ucl.ac.uk,*

*Dept. of Economics, London School of Economics, Houghton Street, London WC2A 2AE, U.K.; o.linton@lse.ac.uk, <http://econ.lse.ac.uk/staff/olinton/>,*

*and*

*Dept. of Economics, Seoul National University, Seoul 151-742, Korea; whang@snu.ac.kr; <http://plaza.snu.ac.kr/~whang>.*

*Manuscript received May, 2007; final revision received July, 2008.*