Testing for stochastic monotonicity

Sokbae Lee
Oliver Linton
Yoon-Jae Whang

TESTING FOR STOCHASTIC MONOTONICITY[*]

Sokbae Lee[†], Oliver Linton[‡], and Yoon-Jae Whang[§]

July 28, 2008

Abstract

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.

Key words: Distribution function; Extreme Value Theory; Gaussian Process; Monotonicity.
Journal of Economic Literature Classification: C14, C15.

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) H_0 : For each $y \in \mathcal{Y}$, $F_{Y|X}(y|x) \leq F_{Y|X}(y|x')$ whenever $x \geq x'$ 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. To our best knowledge, 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) vs. incomes (X) at household levels, wages (Y) vs. cognitive skills (X) using individual data, outputs (Y) vs. the stock of capital (X) at the country level, sons' incomes (Y) vs. fathers' incomes (X) using family data, and so on.

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 $Y = Y_{t+1}$ and

$X = Y_t$ 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) uses 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 behaviour and it provides a route to proving the existence of stationary equilibria not requiring smoothness. Hopenhayn and Prescott (1992) argue 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 non-monotone 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 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 a 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 only apply to either one or the other case. For example, Stute (1986) establishes the weak convergence to a Brownian Bridge of a conditional empirical

2

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) establish 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 consider the asymptotics of some functional of a conditional empirical process except that they consider a maximum over a discrete partition of the support of the covariate. See also Einmahl and Van Keilegom (2006). We use some results of Piterbarg (1996) to establish the approximation. These results can be of use elsewhere. See Appendix A.1 on some informal discussion on the proof technique.

One issue with the extreme value limiting distributions is known to be 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) provides 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, e.g. 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) i.e., $\phi(Y) = m(X) + \varepsilon$, where ε is independent of X and both ϕ, 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 in the Appendix.

2 The Test Statistic

This section describes our test statistic. Let $\{(Y_i, X_i) : i = 1, \ldots, n\}$ 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 $Y_i = Y_{t+1}$ and $X_i = Y_t$. 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 θ_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 following U-process:

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

where $K_{h_n}(\cdot) = h_n^{-1}K(\cdot/h_n)$ and $\mathrm{sgn}(x) = 1(x > 0) - 1(x < 0)$. 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 $n \to \infty$,

$$h_n^{-1}EU_n(y, x) \to F_x(y|x) \left(\int \int |u_1 - u_2|K(u_1)K(u_2)du_1du_2 \right) [f_X(x)]^2,$$

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

4

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

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

REMARK 2.1. (i) An alternative class of test statistics is based on explicit estimation of conditional 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 $\widehat{F}_{Y|X}(y|x)$ is some e.g., 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 non-nested 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

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

$q(u) = \int \mathrm{sgn}(u - w)K(w)dw$ and $\widehat{f}_X(x)$ is the kernel density estimator of $f_X(x)$. It can be shown easily that $\tilde{\sigma}_n(x)$ is asymptotically equivalent to $\widehat{\sigma}_n(x)$. In finite samples, $\widehat{\sigma}_n(x)$ may worker 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 $X_i = \psi(W_i, \theta_0) + \varepsilon_i, (i = 1, \ldots, n)$, where ε_i is an unobserved random variable. In this case, our test

does not apply with $\widehat{X}_i = \psi(W_i, \widehat{\theta})$. In order 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 behaviour 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)
$$\widehat{U}_n(y, x) - U_n(y, x) = O_p(n^{-1/2}) \quad \text{and} \quad h_n^{1/2} \widehat{\sigma}_n(x) = O_p(1)$$

uniformly over (y, x) (see Lemmas A.6, A.7, and A.9). Then if $h_n \to 0$,

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

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

$$\tilde{U}_n(y, x) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} [F_{Y|X}(y|X_i) - F_{Y|X}(y|X_j)]\mathrm{sgn}(X_i - X_j)$$
$$\times K_{h_n}(X_i - x)K_{h_n}(X_j - x).$$

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 $\log n/(nh_n^3) \to 0$,

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

6

uniformly over (y, x).

Under the null hypothesis (1), note that

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

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, in order 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) θ_0 is a finite-dimensional parameter and $||\widehat{\theta} - \theta_0|| = O_p(n^{-1/2})$; (g) for each w, $\psi(w, \theta)$ is continuously differentiable with respect to θ; (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 θ_1 and θ_2 in a neighborhood of θ_0, where*

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

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

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

where

(8) $$\lambda = -\frac{6\int q(x)K^2(x)K'(x)dx + \int q^2(x)K(x)K''(x)dx}{\int q^2(x)K^2(x)dx}.$$

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 $h_n \log n \to 0$, $nh_n^3/(\log n) \to \infty$, and $nh_n^2/(\log n)^4 \to \infty$. Then for any x,*

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

7

In particular,

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

REMARK 3.1. It is necessary to compute β_n in (7) to construct a test based on Theorem 3.1. The constant λ 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 $\lambda = 131689/11063 \approx 11.904$ for the biweight kernel $K(u) = (15/16)(1 - u^2)^2 1(|u| \leq 1)$. It is straightforward to show that

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

where $c^* = (8\lambda/\pi)^{1/2}$. Then one can use an approximation to β_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 $\widehat{h}_n$ be a data-dependent sequence such that $\widehat{h}_n/h_n \overset{p}{\to} 1$, 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 $\widehat{h}_n$ is the same as the one 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 α:

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

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

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

where for each n, $F_n(x)$ is the 'distribution function' of the form[1]

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

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,

---

[1] The approach in (12) to defining the critical region is motivated partly by the normalizing transformation approach, Phillips (1979).

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 (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 again general alternatives.

Theorem 3.2. *Assume that $nh_n^3/\log h_n^{-1} \to \infty$. 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 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 $\mathcal{Y} = \{Y_1, Y_2, \ldots, Y_n\}$, $\mathcal{X} = \{0.05, 0.10, \ldots, 0.90, 0.95\}$, and $\tilde{\sigma}_n(x)$ was defined in Remark 2.3. The kernel function was $K(u) = 0.75(1 - u^2)$ for $-1 \le u \le 1$. 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 1 reports results of Monte Carlo experiments using critical values obtained from the asymptotic expansion F_n of the limiting distribution (see (12)) 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 one for $n = 100$, and one 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 1 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 resampling 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 are smaller than either of 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 US 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 parental incomes can give further insights into understanding of the 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.[2]

We use data from the Panel Study of Income Dynamics (PSID), which has been used frequently to study mobility after 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 son's averaged full-time real labor income when age 28 and age 29 and the X variable is the logarithm of parental predicted permanent income.[3]

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 the one used in the Monte Carlo experiment. For each quantile, the bandwidth is chosen by a 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 $\sqrt{n}U_n(y,x)/\widehat{\sigma}_n(x)$ over $\mathcal{Y} \times \mathcal{X}$, where $\mathcal{Y} = \{Y_1, Y_2, \ldots, Y_n\}$, and $\mathcal{X} = [8.48, 10.85]$, where two end points of $\mathcal{X}$ correspond to 1 and 99 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 β_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), respectively. 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

---

[2]As a related paper, Dearden, Machin, and Reed (1997) investigate the intergenerational income mobility in Britain using the quantile transition matrix approach.

[3]Minicozzi (2003) computes sons' average incomes only when both incomes at ages 28 and 29 are available and regards 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 628 original observations of 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.

with respect to Lebesgue measure on $\mathbb{R}^d$. We consider the following hypothesis, which is a multivariate generalization of (1),

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

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

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 semi-monotonicity 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 $\{(Y_i, \mathbf{X}_i) : i = 1, \ldots, n\}$ 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 $\mathbf{u} \equiv (u_1, \ldots, u_d)$, 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 following U-process:

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

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

Again, we define our test statistic as a supremum statistic

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

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

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

Under the null hypothesis (13), note that

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

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, in order 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 following equation:

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

where λ 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 $h_n \log n \to 0$, $nh_n^{3d}/(\log n) \to \infty$, and $nh_n^{d+1}/(\log n)^{2(d+1)} \to \infty$. Then for any x,*

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

In particular,

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

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 $nh_n^{3d}/\log h_n^{-1} \to \infty$.

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.

A   Appendix: Proofs

A.1   Informal Discussion of the Proof Technique

Although the test is easy to implement, proving Theorem 3.1 involves several lengthy steps. Since establishing these steps requires techniques that are not commonly used in econometrics, we now give an informal description of our proof techniques and provide some discussions behind them. Specifically, our proof of Theorem 3.1 consists of the following three steps:

1. The asymptotic approximation of $\widehat{U}_n(y, x)/c_n(x)$ by a Gaussian process (Appendix A.2);

2. The asymptotic approximation of the excursion probability of the maximum of the Gaussian process on a fixed set (Appendix A.3);

3. The asymptotic approximation of the excursion probability of the maximum of the Gaussian process on an increasing set (Appendix A.4).

In particular, in step 1, we show that $\widehat{U}_n(y, x)/c_n(x)$ can be approximated uniformly over (y, x) by $\xi_n[F_Y(y), h_n^{-1}x]$, where $F_Y(\cdot)$ is the c.d.f of Y and ξ_n is a sequence of Gaussian processes $\{\xi_n(u, s) : (u, s) \in [0, 1] \times [0, h_n^{-1}]\}$ with continuous sample paths such that

$$(18) \qquad E[\xi_n(u, s)] = 0, \quad E[\xi_n(u_1, s_1)\xi_n(u_2, s_2)] = [\min(u_1, u_2) - u_1 u_2]\rho(s_1 - s_2),$$

for $u, u_1, u_2 \in [0, 1]$ and $s, s_1, s_2 \in [0, h_n^{-1}]$, where $\rho(\cdot)$ is some known smooth function. See Appendix A.2 for the exact form of $\rho(\cdot)$.

First of all, note that by step 1, taking the supremum of $\widehat{U}_n(y, x)/c_n(x)$ over (y, x) corresponds to taking the supremum of $\xi_n[F_Y(y), h_n^{-1}x]$ over (y, x) asymptotically. Since F_Y is the c.d.f. and $h_n \to 0$, this means that we need to take the supremum of the Gaussian process ξ_n over the product space of a fixed set (in the direction of y) and an increasing set (in the direction of x).

In general, it is expected that the asymptotic distribution of a suitably normalized version of the supremum of a Gaussian process over an increasing set converges to one of extreme value distributions. If the supremum is taken over for Gaussian processes with an one-dimensional parameter, then the corresponding probability theory and applications on statistical problems are well understood. See, for example, see Leadbetter, Lindgren, and Rootzén (1983). However, for Gaussian processes with multi-dimensional parameters (often called Gaussian fields), the probability theory is less developed and applications on statistical problems are rare. Unfortunately, we need to deal with $\xi_n(u, s)$ that has two parameters and approximate the distribution of its supremum over an increasing set. These tasks are steps 2 and 3. The

14

important reference we have used to carry out steps 2 and 3 is Piterbarg (1996), who developed a general theory for approximations of the suprema of Gaussian fields.

Once step 2 is established, then there is a general approximation method to achieve step 3. Thus, Step 2 is the critical step in proving Theorem 3.1. Note that the covariance function of ξ_n in (18) is the product of a Brownian Bridge covariance function and a stationary covariance function. In this paper, we develop a new result for the excursion probability of the maximum of the Gaussian process ξ_n (Theorem A.2). To be specific, the approximating Gaussian process contains both a stationary and a nonstationary part and therefore we need to extend existing results that only apply to either one or the other case. For example, see Section 7 of Piterbarg (1996) for the stationary case and Sections 8 and 9 of Piterbarg (1996) for the nonstationary case, but to our best knowledge, there is no known result regarding our case in the literature.

A.2 Gaussian Process Approximation

Let $f_X(\cdot)$, $F_X(\cdot)$, and $F_Y(\cdot)$, respectively, denote the p.d.f. and c.d.f. of X and c.d.f. of Y. Define
$$\rho(s) = \frac{\int q(z)q(z-s)K(z)K(z-s)dz}{\int q^2(z)K^2(z)dz},$$
where $q(u) = \int \mathrm{sgn}(u-w)K(w)dw$ was defined in the main text. Let $\xi(u,s)$ denote a Gaussian process $\{\xi(u,s) : (u,s) \in [0,1] \times \mathbb{R}\}$ with continuous sample paths such that

$$E[\xi(u,s)] = 0, \quad E[\xi(u_1,s_1)\xi(u_2,s_2)] = [\min(u_1,u_2) - u_1 u_2]\rho(s_1 - s_2),$$

for $u, u_1, u_2 \in [0,1]$ and $s, s_1, s_2 \in \mathbb{R}$. Define $\mathcal{X}_n = [0, 1/h_n]$ and let ξ_n be the restriction of ξ to $[0,1] \times \mathcal{X}_n$.

Theorem A.1. *Let Assumption 3.1 hold. Let h_n satisfy*

$$h_n(\log n)^{1/2} \to 0, \quad nh_n^3 \to \infty, \quad and \quad nh_n^2/(\log n)^2 \to \infty.$$

Then there exists a sequence of Gaussian processes $\{\xi_n(u,s) : (u,s) \in [0,1] \times \mathcal{X}_n\}$ with continuous sample paths such that

$$E[\xi_n(u,s)] = 0, \quad E[\xi_n(u_1,s_1)\xi_n(u_2,s_2)] = [\min(u_1,u_2) - u_1 u_2]\rho(s_1 - s_2),$$

for $u, u_1, u_2 \in [0,1]$ and $s, s_1, s_2 \in \mathcal{X}_n$, and that

$$\sup_{(y,x)\in\mathcal{Y}\times\mathcal{X}} \left| n^{1/2} \frac{\widehat{U}_n(y,x)}{\widehat{\sigma}_n(x)} - \xi_n[F_Y(y), h_n^{-1}x] \right| = O_p\left(n^{-1/2}h_n^{-3/2} + n^{-1/4}h_n^{-1/2}(\log n)^{1/2} + h_n(\log h_n^{-1})^{1/2} \right).$$

15

Proof. The proof of Theorem follows closely Theorem 3.1 of Ghosal, Sen, and van der Vaart (2000). In particular, the theorem can be proved by combining arguments almost identical to those used in the proof of Theorem 3.1 of Ghosal, Sen, and van der Vaart (2000) with Lemmas proved in Section A.6. The only difference here is that because of the estimated X_i's, an additional term of order $O_p\left(n^{-1/2}h_n^{-3/2}\right)$ appears. $\square$

A.3 Asymptotic Behaviour of the Excursion Probability on the Fixed Set

Since the distribution of $\xi_n(u,s)$ does not depend on n, for the purpose of deriving the distribution of the supremum statistic S_n, it suffices to consider the asymptotic behaviour of the excursion probability of the maximum of the Gaussian process $\xi(u,s)$ that has the same covariance function as $\xi_n(u,s)$.

We first consider the asymptotic behaviour of the tail probability of the maximum of $\xi(u,s)$ on a fixed set $[0,1] \times I$, where $I \equiv [0,L]$ is an interval with a fixed length L. Define

$$\Psi(a) = \frac{1}{\sqrt{2\pi}} \int_a^\infty \exp\left(-\frac{1}{2}x^2\right) dx.$$

Theorem A.2. *Let λ denote the quantity defined in Theorem 3.1. In addition, let $I = [0,L]$. Then*

$$\Pr\left(\max_{(u,s)\in[0,1]\times I} \xi(u,s) > a\right) = L\left(\frac{8\lambda}{\pi}\right)^{1/2} a \exp(-2a^2)[1+o(1)]$$

as $a \to \infty$.

The following Lemmas are useful to prove Theorem A.2.

Lemma A.1. *Let $\Pi_\delta = [1/2 - \delta(a), 1/2 + \delta(a)]$, where $\delta(a) = a^{-1}\log a$. Then*

$$\Pr\left(\max_{(u,s)\in[0,1]\times I} \xi(u,s) > a\right) = \Pr\left(\max_{(u,s)\in\Pi_\delta\times I} \xi(u,s) > a\right)[1+o(1)]$$

as $a \to \infty$.

Proof. For all sufficiently large a,

(19)
$$\Pr\left(\max_{(u,s)\in\Pi_\delta\times I} \xi(u,s) > a\right) \leq \Pr\left(\max_{(u,s)\in[0,1]\times I} \xi(u,s) > a\right)$$

$$\leq \Pr\left(\max_{(u,s)\in\Pi_\delta\times I} \xi(u,s) > a\right) + \Pr\left(\max_{(u,s)\in\{[0,1]\backslash\Pi_\delta\}\times I} \xi(u,s) > a\right).$$

Note that

$$E[\xi(u_1,s_1) - \xi(u_2,s_2)]^2 = u_1(1-u_1) + u_2(1-u_2) - 2[\min(u_1,u_2) - u_1u_2]\rho(s_1-s_2).$$

16

Furthermore, by some straightforward manipulation,

$$E[\xi(u_1, s_1) - \xi(u_2, s_2)]^2 \le C\,|u_1 - u_2| + |s_1 - s_2|$$

for some constant C. Thus, Assumption **E3** of Piterbarg (1996, p.118) is satisfied. Then since

$$\max_{(u,s)\in\{[0,1]\setminus\Pi_\delta\}\times I}\sigma^2(u,s) \le 1/4 - \delta(a)^2,$$

by Theorem 8.1 of Piterbarg (1996, p.119), there exists a constant C such that

(20) $\quad \Pr\left(\max_{(u,s)\in\{[0,1]\setminus\Pi_\delta\}\times I}\xi(u,s) > a\right) \le C\operatorname{mes}(\{[0,1]\setminus\Pi_\delta\}\times I)\,a^4\Psi\left(\dfrac{a}{[1/4-\delta(a)^2]^{1/2}}\right).$

Note that by (D.8) of Piterbarg (1996, p.15), as $a \to \infty$,

$$a^4\Psi\left(\frac{a}{[1/4-\delta(a)^2]^{1/2}}\right) \sim \frac{1}{\sqrt{2\pi}}a^3\exp\left(\frac{-a^2/2}{1/4-\delta(a)^2}\right),$$

where $A \sim B$ stands for $A/B \to 1$. Also, for some fixed interior point $\bar s \in I$, we have

$$\Pr\left(\xi(1/2,\bar s) > a\right) = \Psi(a/2) \sim \frac{2}{\sqrt{2\pi}}a^{-1}\exp\left(\frac{-a^2/2}{4}\right).$$

Then it is easy to show that as $a \to \infty$, the probability on the left-hand side of (20) converges to zero at a rate of $\exp\left[-2a^2 + O(\log a)\right]$ and $\Pr\left(\xi(1/2,\bar s) > a\right)$ converges to zero at a rate of $\exp\left[-a^2/8 - O(\log a)\right]$. Thus, the probability on the left-hand side of (20) converges to zero faster than $\Pr\left(\xi(1/2,\bar s) > a\right)$. Since $\Pr\left(\xi(1/2,\bar s) > a\right) \le \Pr\left(\max_{(u,s)\in\Pi_\delta\times I}\xi(u,s) > a\right)$,

$$\Pr\left(\max_{(u,s)\in\{[0,1]\setminus\Pi_\delta\}\times I}\xi(u,s) > a\right) = o\left[\Pr\left(\max_{(u,s)\in\Pi_\delta\times I}\xi(u,s) > a\right)\right].$$

Then the lemma follows immediately from (19). $\qquad\square$

Let $\sigma^2(u,s) = u(1-u)$ and $r[(u_1,s_1),r(u_2,s_2)] = [\min(u_1,u_2)-u_1u_2]\rho(s_1-s_2)$, respectively, denote the variance and covariance functions of $\xi(u,s)$.

Lemma A.2. *As $u \to 1/2$,*

(21) $$\sigma^2(u,s) = \frac{1}{4} - \left(u - \frac{1}{2}\right)^2[1+o(1)]$$

Furthermore, as $(u_1,u_2) \to (1/2,1/2)$ and $|s_1 - s_2| \to 0$,

(22)
$$\begin{aligned}
r[(u_1,s_1),r(u_2,s_2)] &= \frac{1}{4} - \frac{1}{2}|u_1-u_2|\,[1+o(1)] - \frac{\lambda}{8}(s_1-s_2)^2\,[1+o(1)] \\
&\quad - \frac{1}{2}\left(u_1-\frac{1}{2}\right)^2[1+o(1)] - \frac{1}{2}\left(u_2-\frac{1}{2}\right)^2[1+o(1)].
\end{aligned}$$

17

Proof. The first result (21) follows easily from a second-order Taylor series expansion of the variance of $\xi(u, s)$ with respect to u. We now consider the second result (22). In view of the proof of Theorem 9.2 of Piterbarg (1996, p.138), note that as $(u_1, u_2) \to (1/2, 1/2)$,

$$
(23) \qquad \frac{\min(u_1, u_2) - u_1 u_2}{\sqrt{u_1(1-u_1)u_2(1-u_2)}} = 1 - \frac{1}{2} \frac{|u_1 - u_2|}{\sqrt{u_1(1-u_1)u_2(1-u_2)}} + o\left(|u_1 - u_2|\right).
$$

Note that by (4.9) of Ghosal, Sen, and van der Vaart (2000),

$$
(24) \qquad \rho(s_1 - s_2) = 1 - \frac{\lambda(s_1 - s_2)^2}{2} + o\left(|s_1 - s_2|^2\right),
$$

as $|s_1 - s_2| \to 0$. As in (21), a Taylor series expansion of $\sigma(u, s)$ around $u = 1/2$ gives

$$
\sigma(u, s) = \frac{1}{2} - \left(u - \frac{1}{2}\right)^2 [1 + o(1)], \quad \text{as } u \to \frac{1}{2}
$$

for any $s \in I$. Thus, we have

$$
(25) \qquad \sqrt{u_1(1-u_1)u_2(1-u_2)} = \frac{1}{4} - \frac{1}{2}\left(u_1 - \frac{1}{2}\right)^2 [1 + o(1)] - \frac{1}{2}\left(u_2 - \frac{1}{2}\right)^2 [1 + o(1)],
$$

as $(u_1, u_2) \to (1/2, 1/2)$. Then the lemma follows from combining (23) and (24) with (25). $\quad\square$

Let $\varepsilon > 0$ be a fixed constant. Define Gaussian processes $\psi_1^-(u)$ and $\psi_1^+(u)$ such that

$$
\psi_1^-(u) = \frac{\zeta_1^-(u)}{2^{3/2}[1 + 4(1+\varepsilon)(u-0.5)^2]} \quad \text{and} \quad \psi_1^+(u) = \frac{\zeta_1^+(u)}{2^{3/2}[1 + 4(1-\varepsilon)(u-0.5)^2]}
$$

where $\zeta_1^-(u)$ and $\zeta_1^+(u)$ are Gaussian stationary processes with zero means and the covariance functions $r_1^-(u) = \exp\left[-4(1-\varepsilon)|u|\right]$ and $r_1^+(u) = \exp\left[-4(1+\varepsilon)|u|\right]$. In addition, define mean-zero stationary Gaussian processes $\psi_2^-(s)$ and $\psi_2^+(s)$ such that they are independent of $\psi_1^-(u)$ and $\psi_1^+(u)$ and have the the covariance functions of the form

$$
r_2^-(s) = \frac{1}{8}\left[1 - \lambda(1-\varepsilon)s^2 + o(s^2)\right],
$$

$$
r_2^+(s) = \frac{1}{8}\left[1 - \lambda(1+\varepsilon)s^2 + o(s^2)\right],
$$

respectively. Finally, define

$$
\psi^-(u, s) = \psi_1^-(u) + \psi_2^-(s) \quad \text{and} \quad \psi^+(u, s) = \psi_1^+(u) + \psi_2^+(s).
$$

Lemma A.3. *Let $\varepsilon > 0$ be any fixed, arbitrarily small, constant. Then for all sufficiently large a,*

$$
\Pr\left(\max_{(u,s)\in\Pi_\delta \times I} \psi^-(u, s) > a\right) \leq \Pr\left(\max_{(u,s)\in\Pi_\delta \times I} \xi(u, s) > a\right) \leq \Pr\left(\max_{(u,s)\in\Pi_\delta \times I} \psi^+(u, s) > a\right).
$$

Proof. As noted in the proofs of Theorems D.4 and 8.2 of Piterbarg (1996, p.23 and p.133), the lemma follows from Lemma A.2 and the fact that the distribution of the maximum is monotone with respect to the variance and the Slepian inequality (see, for example, Theorem C.1 of Piterbarg (1996, p.6)). □

Lemma A.4. *Let $\varepsilon > 0$ be any fixed, arbitrarily small, constant. As $a \to \infty$,*

$$(26) \qquad \Pr\left(\max_{u \in \Pi_\delta} 2^{3/2}\psi_1^-(u) > a\right) = 2^{1/2}\frac{(1-\varepsilon)}{(1+\varepsilon)^{1/2}}\exp(-a^2/2)[1+o(1)],$$

$$(27) \qquad \Pr\left(\max_{u \in \Pi_\delta} 2^{3/2}\psi_1^+(u) > a\right) = 2^{1/2}\frac{(1+\varepsilon)}{(1-\varepsilon)^{1/2}}\exp(-a^2/2)[1+o(1)].$$

Proof. This lemma can be proved by one of results given in the proof of Theorem D.4 of Piterbarg (1996, p.21). In particular, using the notation used in the proof of of Theorem D.4 of Piterbarg (1996), the excursion probability of $2^{3/2}\psi_1^-(u)$ can be obtained by the result of Case 1 with $\alpha = 1$, $\beta = 2$, $b = 4(1+\varepsilon)$, and $d = 4(1-\varepsilon)$. It follows from the second display on page 22 of Piterbarg (1996) that as $a \to \infty$,

$$\Pr\left(\max_{u \in \Pi_\delta} 2^{3/2}\psi_1^-(u) > a\right) = \frac{H_1\Gamma(1/2)[4(1-\varepsilon)]}{[4(1+\varepsilon)]^{1/2}}a\Psi(a)[1+o(1)],$$

where H_1 is the Pickands' constant with $\alpha = 1$ (defined on pages 13 and 16 of Piterbarg (1996)) and $\Gamma(\cdot)$ is the Gamma function. Note that $\Gamma(1/2) = \sqrt{\pi}$. Furthermore, by (9.6) of Piterbarg (1996, p.138), $H_1 = 1$ and by (D.8) of Piterbarg (1996, p.15),

$$a\Psi(a) \sim (2\pi)^{-1/2}\exp(-a^2/2)$$

as $a \to \infty$. Therefore, (26) follows immediately. The excursion probability of $2^{3/2}\psi_1^+(u)$ can be obtained analogously. □

Lemma A.5. *Let $\varepsilon > 0$ be any fixed, arbitrarily small, constant. As $a \to \infty$,*

$$(28) \qquad \Pr\left(\max_{s \in I} 2^{3/2}\psi_2^-(s) > a\right) = \frac{[(\lambda/2)(1-\varepsilon)]^{1/2}L}{\pi}\exp(-a^2/2)[1+o(1)],$$

$$(29) \qquad \Pr\left(\max_{s \in I} 2^{3/2}\psi_2^+(s) > a\right) = \frac{[(\lambda/2)(1+\varepsilon)]^{1/2}L}{\pi}\exp(-a^2/2)[1+o(1)].$$

Proof. Recall that $I = [0, L]$. By Theorem D.2 of Piterbarg (1996, p.16) and a simple scaling of $\psi_2^-(u)$,

$$\Pr\left(\max_{s \in [0,L]} 2^{3/2}\psi_2^-(s) > a\right) = H_2 L^* a\Psi(a)[1+o(1)]$$

where H_2 is the Pickands' constant with $\alpha = 2$ and $L^* = [\lambda(1-\varepsilon)]^{1/2}L$. By (F.4) of Piterbarg (1996, p.31), $H_2 = 1/\sqrt{\pi}$. Then (28) follows immediately. The excursion probability of $2^{3/2}\psi_2^+(u)$ can be obtained similarly. □

Proof of Theorem A.2. Let $\varepsilon > 0$ be any fixed, arbitrarily small, constant. Note that $\psi^-(u,s)$ and $\psi^+(u,s)$ are convolutions of $\psi_1^-(u)$ and $\psi_2^-(s)$ and of $\psi_1^+(u)$ and $\psi_2^+(s)$, respectively. Then an application of Lemma 8.6 of Piterbarg (1996, p.128) with Lemmas A.4 and A.5 gives

$$(30) \qquad \Pr\left(\max_{(u,s)\in\Pi_\delta\times I} 2^{3/2}\psi^-(u,s) > a\right) = L\frac{(1-\varepsilon)^{3/2}}{(1+\varepsilon)^{1/2}}\left(\frac{\lambda}{\pi}\right)^{1/2} a\exp(-a^2/4)[1+o(1)],$$

$$(31) \qquad \Pr\left(\max_{(u,s)\in\Pi_\delta\times I} 2^{3/2}\psi^+(u,s) > a\right) = L\frac{(1+\varepsilon)^{3/2}}{(1-\varepsilon)^{1/2}}\left(\frac{\lambda}{\pi}\right)^{1/2} a\exp(-a^2/4)[1+o(1)].$$

Then as $a \to \infty$, by Lemma A.1,

$$\Pr\left(\max_{(u,s)\in[0,1]\times I} 2^{3/2}\xi(u,s) > a\right) = L\left(\frac{\lambda}{\pi}\right)^{1/2} a\exp(-a^2/4)[1+o(1)]$$

since the choice of ε can be made arbitrarily small and the constants on the right-hand sides of (30) and (31) are continuous at $\varepsilon = 0$. Therefore, the theorem follows immediately. $\square$

A.4 Asymptotic Behaviour of the Excursion Probability on the Increasing Set

Theorem A.3. *For any x,*

$$\Pr\left(4\beta_n\left\{\max_{(u,s)\in[0,1]\times\mathcal{X}_n}\xi(u,s) - \beta_n\right\} < x\right) = \exp\left\{-\exp\left(-x - \frac{x^2}{8\beta_n^2}\right)\left[1 + \frac{x}{4\beta_n^2}\right]\right\} + o(1).$$

where β_n is defined in (7).

Proof of Theorem A.3. This theorem can be proved using arguments similar to those used in the proof of Theorem G.1 of Piterbarg (1996). Note that the covariance function of $\xi(u,s)$, that is $r[(u_1,s_1), r(u_2,s_2)]$, has compact support and in particular it is zero when $|s_1 - s_2| > 2$. Define an increasing sequence m_n such that $m_n \to \infty$ but $m_n h_n \to 0$ as $n \to \infty$. That is, m_n converges to infinity slower than h_n^{-1}. Further, define sequences of sets

$$I_k = \left[k(m_n h_n)^{-1}, (k+1)(m_n h_n)^{-1} - 2\right],$$
$$J_k = \left[(k+1)(m_n h_n)^{-1} - 2, (k+1)(m_n h_n)^{-1}\right],$$

for $k = 0, 1, \ldots, m_n - 1$. Then we have

$$(32)$$
$$\Pr\left(\max_{(u,s)\in[0,1]\times\mathcal{X}_n}\xi(u,s) < a\right) = \Pr\left(\max_{(u,s)\in[0,1]\times\left[\bigcup_k I_k\right]}\xi(u,s) < a\right)$$
$$- \Pr\left(\max_{(u,s)\in[0,1]\times\left[\bigcup_k I_k\right]}\xi(u,s) < a, \max_{(u,s)\in[0,1]\times\left[\bigcup_k J_k\right]}\xi(u,s) \geq a\right).$$

We first consider the first probability on the right-hand side of (32). Let $c^* = \left(\frac{8\lambda}{\pi}\right)^{1/2}$. For each x, choose $a_n = \beta_n + x/(4\beta_n)$, where β_n is the largest solution to the following equation:

$$(33) \qquad\qquad h_n^{-1} c^* \beta_n \exp(-2\beta_n^2) = 1.$$

Since I_k's are separated by the diameter of the support and the distribution of $\xi(u,s)$ is stationary in the direction of s, it follows from Theorem A.2 that

$$
\begin{aligned}
&\Pr\left(\max_{(u,s)\in[0,1]\times[\bigcup_k I_k]} \xi(u,s) < a_n\right) \\
&= \left[1 - \Pr\left(\max_{(u,s)\in[0,1]\times I_0} \xi(u,s) \geq a_n\right)\right]^{m_n} \\
&= \exp\left(m_n \log\left[1 - \Pr\left(\max_{(u,s)\in[0,1]\times I_0} \xi(u,s) \geq a_n\right)\right]\right) \\
&= \exp\left(-m_n\Pr\left(\max_{(u,s)\in[0,1]\times I_0} \xi(u,s) \geq a_n\right)\right) + O\left(m_n\left[\Pr\left(\max_{(u,s)\in[0,1]\times I_0} \xi(u,s) \geq a_n\right)\right]^2\right) \\
&= \exp\left\{-m_n[(m_n h_n)^{-1} - 2]c^* a_n \exp(-2a_n^2)[1 + o(1)]\right\} + O(h_n),
\end{aligned}
$$

so that

$$(34) \qquad \Pr\left(\max_{(u,s)\in[0,1]\times[\bigcup_k I_k]} \xi(u,s) < a_n\right) = \exp\left\{-\exp\left(-x - \frac{x^2}{8\beta_n^2}\right)\left[1 + \frac{x}{4\beta_n^2}\right]\right\} + o(1).$$

Now consider the second probability on the right-hand side of (32). Note that again using Theorem A.2 and the fact that the distribution of $\xi(u,s)$ is stationary in the direction of s,

$$
\begin{aligned}
&\Pr\left(\max_{(u,s)\in[0,1]\times[\bigcup_k I_k]} \xi(u,s) < a_n, \max_{(u,s)\in[0,1]\times[\bigcup_k J_k]} \xi(u,s) \geq a_n\right) \\
&\leq \Pr\left(\max_{(u,s)\in[0,1]\times[\bigcup_k J_k]} \xi(u,s) \geq a_n\right) \\
&\leq m_n\Pr\left(\max_{(u,s)\in[0,1]\times J_1} \xi(u,s) \geq a_n\right) \\
&= m_n\Pr\left(\max_{(u,s)\in[0,1]\times[0,2]} \xi(u,s) \geq a_n\right) \\
&= O(m_n h_n) = o(1).
\end{aligned}
$$

This and (34) together prove the theorem. $\qquad\qquad\square$

A.5   Proofs of Theorems 3.1 and 3.2

Proof of Theorem 3.1. Since $\beta_n[n^{-1/2}h_n^{-3/2} + n^{-1/4}h_n^{-1/2}(\log n)^{1/2} + h_n(\log n)^{1/2}] \to 0$, the main theorem 3.1 is an immediate consequence of Theorems A.1 and A.3. $\qquad\square$

Proof of Theorem 3.2. The theorem can be proved by arguments similar to those used to prove Theorem 5.1 of Ghosal, Sen, and van der Vaart (2000). In fact, when $F_x(y|x) > 0$ for some (y, x), S_n is of order $O_p(n^{1/2}h_n^{3/2})$ and the consistency follows from the restriction that $nh_n^3/\log h_n^{-1} \to \infty$. $\qquad\square$

A.6 Lemmas for Proving Theorem A.1

Define

$$V_n(y, x, \theta) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} [1(Y_i \le y) - 1(Y_j \le y)]\mathrm{sgn}[\psi(W_i, \theta) - \psi(W_j, \theta)]$$
$$\times K_{h_n}[\psi(W_i, \theta) - x]K_{h_n}[\psi(W_j, \theta) - x],$$

so that $\widehat{U}_n(y, x) = V_n\left(y, x, \widehat{\theta}\right)$. Also, since $F_x(y|x) \equiv 0$, define the projection of $V_n(y, x, \theta)$ by

$$\widehat{V}_n(y, x, \theta) = 2n^{-1} \sum_{i=1}^{n} [1(Y_i \le y) - F(y)]$$
$$\times \int \mathrm{sgn}[\psi(W_i, \theta) - \psi(\tilde{w}, \theta)]K_{h_n}[\psi(\tilde{w}, \theta) - x]dF_W(\tilde{w}) \, K_{h_n}[\psi(W_i, \theta) - x].$$

Lemma A.6. *Let Θ denote a neighborhood of θ_0.*

$$\sup_{(y,x,\theta)\in\mathcal{Y}\times\mathcal{X}\times\Theta} \left| V_n(y, x, \theta) - \widehat{V}_n(y, x, \theta) \right| = O_p\left(n^{-1}h_n^{-2}\right).$$

Proof. The proof is similar to that of Lemma 3.1 of Ghosal, Sen, and van der Vaart (2000). Hence, we will only indicate the differences. Consider a class of functions $\mathcal{M} = \{m_{(y,x,\theta)} : (y, x, \theta) \in \mathcal{Y} \times \mathcal{X} \times \Theta\}$, where

$$m_{(y,x,\theta)}((y_1, w_1), (y_2, w_2)) = [1(y_1 \le y) - 1(y_2 \le y)]\mathrm{sgn}[\psi(w_1, \theta) - \psi(w_2, \theta)]$$
$$\times K_{h_n}[\psi(w_1, \theta) - x]K_{h_n}[\psi(w_2, \theta) - x].$$

This class is contained in the product of the classes

$$\mathcal{M}_1 = \{1(y_1 \le y) - 1(y_2 \le y) : y \in \mathcal{Y}\}$$
$$\mathcal{M}_2 = \left\{K\left(\frac{\psi(w_1, \theta) - x}{h_n}\right) : (x, \theta) \in \mathcal{X} \times \Theta\right\}$$
$$\mathcal{M}_3 = \left\{K\left(\frac{\psi(w_2, \theta) - x}{h_n}\right) : (x, \theta) \in \mathcal{X} \times \Theta\right\}$$
$$\mathcal{M}_4 = \{h_n^{-2}\mathrm{sgn}[\psi(w_1, \theta) - \psi(w_2, \theta)]1\{|\psi(w_1, \theta) - \psi(w_2, \theta)| \le 2h_n\} : \theta \in \Theta\}.$$

Since θ is finite-dimensional and K is of bounded variation, $\mathcal{M}$ is a VC-class with the envelope function Ch_n^{-2} with some positive finite constant C, by Lemmas 2.6.15 and 2.6.18 of van der

Vaart and Wellner (1996). Then using Theorem 2.6.7 of van der Vaart and Wellner (1996) and following the proof of Lemma 3.1 of Ghosal, Sen, and van der Vaart (2000), we have, for some finite constant C,

$$E\left[\sup_{(y,x,\theta)\in\mathcal{Y}\times\mathcal{X}\times\Theta}\left|V_n(y,x,\theta)-\widehat{V}_n(y,x,\theta)\right|\right]\le Cn^{-1}h_n^{-2},$$

which gives the conclusion of the lemma. □

Lemma A.7.

$$\sup_{(y,x)\in\mathcal{Y}\times\mathcal{X}}\left|\widehat{U}_n(y,x)-\widehat{V}_n(y,x,\theta_0)\right|=O_p\left(n^{-1/2}\right).$$

Proof. Note that by Assumption 3.1 (h),

$$\left|\widehat{V}_n(y,x,\widehat{\theta})-\widehat{V}_n(y,x,\theta_0)\right|$$

$$=\left|2n^{-1}\sum_{i=1}^{n}[1(Y_i\le y)-F(y)]\right.$$

$$\times\left\{\int\mathrm{sgn}[\psi(W_i,\widehat{\theta})-\psi(\tilde{w},\widehat{\theta})]K_{h_n}[\psi(\tilde{w},\widehat{\theta})-x]dF_W(\tilde{w})\ K_{h_n}[\psi(W_i,\widehat{\theta})-x]\right.$$

$$\left.\left.-\int\mathrm{sgn}[\psi(W_i,\theta_0)-\psi(\tilde{w},\theta_0)]K_{h_n}[\psi(\tilde{w},\theta_0)-x]dF_W(\tilde{w})\ K_{h_n}[\psi(W_i,\theta_0)-x]\right\}\right|$$

$$\le C\left[\left\|\widehat{\theta}-\theta_0\right\|n^{-1}\sum_{i=1}^{n}K_{h_n}[\psi(W_i,\widehat{\theta})-x]+n^{-1}\sum_{i=1}^{n}\left\{K_{h_n}[\psi(W_i,\widehat{\theta})-x]-K_{h_n}[\psi(W_i,\theta_0)-x]\right\}\right]$$

for some positive constant $C<\infty$, which is independent of (y,x). Also, note that using the standard empirical process method (for example, van der Vaart and Wellner, 1996), it is straightforward to show that for a $n^{-1/2}$ neighborhood Θ_n of θ_0,

$$\sup_{(x,\theta)\in\mathcal{X}\times\Theta_n}n^{-1}\sum_{i=1}^{n}K_{h_n}[\psi(W_i,\theta)-x]=O_p(1),$$

$$\sup_{(x,\theta)\in\mathcal{X}\times\Theta_n}n^{-1}\left|\sum_{i=1}^{n}\left\{K_{h_n}[\psi(W_i,\theta)-x]-K_{h_n}[\psi(W_i,\theta_0)-x]\right\}\right|=O_p\left(n^{-1/2}\right).$$

Then the lemma follows from the root-n-consistency of $\widehat{\theta}$ and Lemma A.6 since $\widehat{U}_n(y,x)=V_n\left(y,x,\widehat{\theta}\right)$. □

Define

$$\phi_{n,y,x}(Y,X)=2[1(Y\le y)-F_Y(y)]\int\mathrm{sgn}(X-\tilde{x})K_{h_n}(\tilde{x}-x)dF_X(\tilde{x})K_{h_n}(X-x).$$

Lemma A.8. *There exists a sequence of Gaussian processes $G_n(\cdot)$, indexed by $\mathcal{Y} \times \mathcal{X}$, with continuous sample paths and with*

$$E[G_n(y,x)] = 0, \quad for\ (y,x) \in \mathcal{Y} \times \mathcal{X},$$

$$E[G_n(y_1,x_1)G_n(y_2,x_2)] = E[\phi_{n,y_1,x_1}(Y,X)\phi_{n,y_2,x_2}(Y,X)],$$

for (y_1,x_1) and $(y_2,x_2) \in \mathcal{Y} \times \mathcal{X}$, such that

$$\sup_{(y,x)\in\mathcal{Y}\times\mathcal{X}} \left| n^{1/2}\widehat{V}_n(y,x,\theta_0) - G_n(y,x) \right| = O\left(n^{-1/4}h_n^{-1}(\log n)^{1/2} \right) \quad a.s.$$

Proof. As in the proof of Lemma 3.2 of Ghosal, Sen, and van der Vaart (2000), we use Theorem 1.1 of Rio (1994). Since it can be proved using arguments identical to those used to prove Lemma 3.2 of Ghosal, Sen, and van der Vaart (2000), we will only highlight the differences. To apply Rio's theorem, we rewrite $\varphi_{n,y,x}(Y,X)$ as

$$\phi_{n,y,x}(Y,X) = 2[1(U \le u) - u] \int \operatorname{sgn}(X - \tilde{x})K_{h_n}(\tilde{x} - x)dF_X(\tilde{x})K_{h_n}(X - x)$$

$$\equiv \varphi_{n,u,x}(U,X),$$

where $U = F_Y(Y)$ and $u = F_Y(y)$. Then U is uniformly distributed in $[0,1] \equiv \mathcal{U}$. Thus, Theorem 1.1 of Rio (1994) can be applied to a normalized empirical process associated with $\varphi_{n,u,x}(U,X)$. First, we verify that the class of functions $(v,t) \mapsto h_n\varphi_{n,u,x}(v,t)$, indexed by $(u,x) \in \mathcal{U} \times \mathcal{X}$, is uniformly of bounded variation (UBV). By the definition of Rio (1994), it suffices to show that

$$\sup_{(u,x)\in\mathcal{U}\times\mathcal{X}} \sup_{g\in\mathcal{D}_2([0,1]^2)} \left(\int_{\mathbb{R}^2} h_n\varphi_{n,u,x}(v,t)\ \operatorname{div}\ g(v,t)\ dv\,dt/\, \|g\|_\infty \right) < \infty,$$

where $\mathcal{D}_2([0,1]^2)$ denotes the space of C^∞ functions with values in $\mathbb{R}^2$ and with compact support included in $[0,1]^2$, div denotes the divergence, and $\|g\|_\infty = \sup_{(v,t)\in\mathbb{R}^2} \|g(v,t)\|$ with $\|\cdot\|$ being the usual Euclidean norm. To do so, note that for any $g(v,t) \equiv (g_v(v,t), g_t(v,t))$,

$$\int_{\mathbb{R}^2} \varphi_{n,u,x}(v,t)\ \operatorname{div}\ g(v,t)\ dv\,dt$$

$$= \int_{\mathbb{R}^2} 2[1(v \le u) - u] \int \operatorname{sgn}(t - \tilde{x})K_{h_n}(\tilde{x} - x)dF_X(\tilde{x})K_{h_n}(t - x) \left[\frac{\partial g_v(v,t)}{\partial v} + \frac{\partial g_t(v,t)}{\partial t} \right]\ dv\,dt$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} 2[1(v \le u) - u]\frac{\partial g_v(v,t)}{\partial v}\ dv \int \operatorname{sgn}(t - \tilde{x})K_{h_n}(\tilde{x} - x)dF_X(\tilde{x})K_{h_n}(t - x)\ dt$$

$$+ \int_{\mathbb{R}^2} 2[1(v \le u) - u] \int \operatorname{sgn}(t - \tilde{x})K_{h_n}(\tilde{x} - x)dF_X(\tilde{x})K_{h_n}(t - x)\ \frac{\partial g_t(v,t)}{\partial t}\ dv\,dt.$$

Then it is straightforward to verify that

$$\sup_{g\in\mathcal{D}_2([0,1]^2)} \left(\int_{\mathbb{R}^2} \varphi_{n,u,x}(v,t)\ \operatorname{div}\ g(v,t)\ dv\,dt/\, \|g\|_\infty \right) = O\left(h_n^{-1} \right)$$

24

uniformly over $(u,x) \in \mathcal{U} \times \mathcal{X}$. This implies that the class of functions $\{h_n\varphi_{n,u,x} : (u,x) \in \mathcal{U} \times \mathcal{X}\}$ satisfies the UBV condition of Rio (1994). Furthermore, it is also straightforward to verify that

$$\sup_{g \in \mathcal{D}_2([a,b]^2)} \left(\int_{\mathbb{R}^2} \varphi_{n,u,x}(v,t) \text{ div } g(v,t) \, dv \, dt / \|g\|_\infty \right) = O\left(h_n^{-1}[b-a]\right)$$

uniformly over $(u,x) \in \mathcal{U} \times \mathcal{X}$. This implies that the class of functions $\{h_n\varphi_{n,u,x} : (u,x) \in \mathcal{U} \times \mathcal{X}\}$ also satisfies the LUBV condition of Rio (1994). We now verify that the class of functions $\{h_n\varphi_{n,u,x} : (u,x) \in \mathcal{U} \times \mathcal{X}\}$ is a VC class. The function $h_n\varphi_{n,u,x}$ is bounded by a constant uniformly in $(u,z) \in \mathcal{U} \times \mathcal{X}$ and is obtained by taking an average of

$$2h_n[1(v \le u) - 1(\tilde{u} \le u)]\text{sgn}(\tilde{x} - t)K_{h_n}(\tilde{x} - t)K_{h_n}(t - x)$$

over $(\tilde{u}, \tilde{x})$. Then it is easy to show that $\{h_n\varphi_{n,u,x} : (u,x) \in \mathcal{U} \times \mathcal{X}\}$ is a VC class by using arguments similar to those used in the proof of Lemma 3.2 of Ghosal, Sen, and van der Vaart (2000, in particular equation 8.5). Finally, by applying Theorem 1.1 of Rio (1994), there exists a sequence of centered Gaussian processes $G_n(u,x)$ with covariance

$$E[G_n(u_1,x_1)G_n(u_2,x_2)] = E[\varphi_{n,u_1,x_1}(U,X)\varphi_{n,u_2,x_2}(U,X)].$$

By switching back to the original variable Y and its corresponding index y, we obtain the desired result. $\qquad\square$

Define

$$\sigma_n^2(x) = 4\int \left[\int \text{sgn}(\bar{x} - \tilde{x})K_{h_n}(\tilde{x} - x)dF_X(\tilde{x})K_{h_n}(\bar{x} - x)\right]^2 dF_X(\bar{x})$$

and

$$\sigma^2(x) = 4\left[\int q^2(u)K^2(u)du\right][f_X(x)]^3.$$

Lemma A.9.

$$(a)\sup_{x \in \mathcal{X}} \left|h_n\sigma_n^2(x) - \sigma^2(x)\right| = o(1).$$

$$(b)\lim_{n \to \infty} \inf \, h_n \inf_{x \in \mathcal{X}} \sigma_n^2(x) > 0.$$

$$(c)\sup_{x \in \mathcal{X}} \left|\widehat{\sigma}_n^2(x) - \sigma_n^2(x)\right| = O_p\left(n^{-1/2}h_n^{-2}\right).$$

Proof. Parts (a) and (b) of the lemma follow directly from Lemma 3.3 (a)-(b) of Ghosal, Sen, and van der Vaart (2000). To prove part (c) of the lemma, note that $\widehat{\sigma}_n^2(x)$ depends on the estimated X_i. To deal with this, let $\tilde{\sigma}_n^2(x,\theta)$ be the same as $\widehat{\sigma}_n^2(x)$ except that $\widehat{X}_i$ is replaced

25

by $\psi(W_i, \theta)$. As in the proof of Lemma A.6, modifying the proof of Lemma 3.3 of Ghosal, Sen, and van der Vaart (2000) gives

$$\sup_{x \in \mathcal{X}} \sup_{\theta \in \Theta} \left| \tilde{\sigma}_n^2(x, \theta) - E \tilde{\sigma}_n^2(x, \theta) \right| = O_p \left(n^{-1/2} h_n^{-2} + n^{-1} h_n^{-3} + n^{-3/2} h_n^{-4} \right),$$

where Θ is a neighborhood of θ_0. Then part (c) follows from the restriction on h_n and the fact that $E \tilde{\sigma}_n^2(x, \theta)$ is Lipschitz continuous with respect to θ. $\qquad \square$

Lemma A.10. *For the sequence of Gaussian processes $\{G_n(y, x) : (y, x) \in \mathcal{Y} \times \mathcal{X}\}$ obtained in Lemma A.8, there corresponds a sequence of Gaussian processes $\{\xi_n(u, s) : (u, s) \in [0, 1] \times \mathcal{X}_n\}$ with continuous sample paths such that*

$$E[\xi_n(u, s)] = 0, \quad E[\xi_n(u_1, s_1)\xi_n(u_2, s_2)] = [\min(u_1, u_2) - u_1 u_2]\rho(s_1 - s_2),$$

for $u, u_1, u_2 \in [0, 1]$ and $s, s_1, s_2 \in \mathcal{X}_n$, where

$$\sup_{(y,x) \in \mathcal{Y} \times \mathcal{X}} \left| \frac{G_n(y, x)}{\sigma_n(x)} - \xi_n[F_Y(y), h_n^{-1} x] \right| = O_p \left(h_n \sqrt{\log h_n^{-1}} \right)$$

Proof. Let $\mathcal{G}_n$ denote the class of functions $\{g_{n,u,x} : (u, x) \in \mathcal{U} \times \mathcal{X}\}$, where $g_{n,u,x}(U, X) = \varphi_{n,u,x}(U, X)/\sigma_n(x)$. Also, let $\tilde{\mathcal{G}}_n$ denote the class of functions $\{\tilde{g}_{n,u,x} : (u, x) \in \mathcal{U} \times \mathcal{X}\}$, where

$$\tilde{g}_{n,u,x}(U, X) = \tilde{\varphi}_{n,u,x}(U, X)/\tilde{\sigma}_{n,x}(X),$$

$$\tilde{\varphi}_{n,u,x}(U, X) = [1(U \leq u) - u] \int \operatorname{sgn}(X - \tilde{x}) \, K_{h_n}(\tilde{x} - x) d\tilde{x} K_{h_n}(X - x),$$

$$\tilde{\sigma}_{n,x}(X) = \left[\int \left(\int \operatorname{sgn}(\bar{x} - \tilde{x}) K_{h_n}(\tilde{x} - x) d\tilde{x} \right)^2 \left[K_{h_n}(\bar{x} - x) \right]^2 d\bar{x} \right]^{1/2} [f_X(X)]^{1/2}.$$

As explained in Remark 8.3 of Ghosal, Sen, and van der Vaart (2000), it is possible to extend Lemma A.8 in that there exists a sequence of Gaussian bridges, say $\{B_n(g) : g \in \mathcal{G}_n \cup \tilde{\mathcal{G}}_n\}$, with

$$E[B_n(g)] = 0, \ E[B_n(g_1)B_n(g_2)] = \operatorname{cov}(g_1, g_2)$$

for all $g, g_1, g_2 \in \mathcal{G}_n \cup \tilde{\mathcal{G}}_n$ and with continuous sample paths with respect to the L_2-metric such that

$$G_n(u, x) = \sigma_n(x)B_n(\varphi_{n,u,x}),$$

where $G_n(u, x)$ is defined in the proof of Lemma A.8. Now let $\tilde{\xi}_n(u, x) = B_n(\tilde{g}_{n,u,x})$ and $\gamma_n(u, x) = G_n(u, x)/\sigma_n(x) - \tilde{\xi}_n(u, x)$. As in the proof of Lemma 3.4 of Ghosal, Sen, and van der Vaart (2000), note that $\gamma_n(u, x)$ is a mean zero Gaussian process with

$$E[\gamma_n(u_1, x_1)\gamma_n(u_2, x_2)] = E[(g_{n,u_1,x_1} - \tilde{g}_{n,u_1,x_1})(g_{n,u_2,x_2} - \tilde{g}_{n,u_2,x_2})].$$

Then the lemma can be proved using identical arguments used in the proof of Lemma 3.4 of Ghosal, Sen, and van der Vaart (2000). $\qquad \square$

A.7 Proof of Theorem 6.1

This theorem can be proved using arguments similar to those used in the proof of Theorem 3.1. In particular, the following lemmas can be proved, whose proofs are omitted here for brevity, and then the desired result follows.

Recall that

$$\rho(s) = \frac{\int q(z)q(z-s)K(z)K(z-s)dz}{\int q^2(z)K^2(z)dz},$$

where $q(u) = \int \operatorname{sgn}(u-w)K(w)dw$. Let $\boldsymbol{\xi}(u,\mathbf{s})$ denote a Gaussian process $\{\boldsymbol{\xi}(u,\mathbf{s}) : (u,\mathbf{s}) \in [0,1] \times \mathbb{R}^d\}$ with continuous sample paths such that

$$E[\boldsymbol{\xi}(u,\mathbf{s})] = 0, \quad E[\boldsymbol{\xi}(u_1,\mathbf{s}_1)\boldsymbol{\xi}(u_2,\mathbf{s}_2)] = [\min(u_1,u_2) - u_1 u_2] \prod_{j=1}^{d} \rho(s_{1j} - s_{2j}),$$

for $u, u_1, u_2 \in [0,1]$ and $\mathbf{s}, \mathbf{s}_1 \equiv (s_{11}, \ldots, s_{1d}), \mathbf{s}_2 \equiv (s_{21}, \ldots, s_{2d}) \in \mathbb{R}^d$. Define $\mathcal{X}_n = [0, 1/h_n]^d$ and let $\boldsymbol{\xi}_n$ be the restriction of $\boldsymbol{\xi}$ to $[0,1] \times \mathcal{X}_n$.

Lemma A.11. *Let Assumption 6.1 hold. Let h_n satisfy*

$$h_n(\log n)^{1/2} \to 0, \quad nh_n^{3d} \to \infty, \quad and \quad nh_n^{d+1}/(\log n)^{d+1} \to \infty.$$

Then there exists a sequence of Gaussian processes $\{\boldsymbol{\xi}(u,\mathbf{s}) : (u,\mathbf{s}) \in [0,1] \times \mathcal{X}_n\}$ with continuous sample paths such that

$$E[\boldsymbol{\xi}_n(u,\mathbf{s})] = 0, \quad E[\boldsymbol{\xi}_n(u_1,\mathbf{s}_1)\boldsymbol{\xi}_n(u_2,\mathbf{s}_2)] = [\min(u_1,u_2) - u_1 u_2] \prod_{j=1}^{d} \rho(s_{1j} - s_{2j}),$$

for $u, u_1, u_2 \in [0,1]$ and $\mathbf{s}, \mathbf{s}_1 \equiv (s_{11}, \ldots, s_{1d}), \mathbf{s}_2 \equiv (s_{21}, \ldots, s_{2d}) \in \mathcal{X}_n$, and that

$$\sup_{(y,\mathbf{x}) \in \mathcal{Y} \times \mathcal{X}} \left| n^{1/2} \frac{\widehat{\mathbf{U}}_n(y,\mathbf{x})}{\widehat{\mathbf{s}}_n(x)} - \boldsymbol{\xi}_n[F_Y(y), h_n^{-1}\mathbf{x}] \right|$$
$$= O_p\left[n^{-1/2}h_n^{-3d/2} + n^{-1/2(d+1)}h_n^{-1/2}(\log n)^{1/2} + h_n(\log n)^{1/2} \right].$$

Lemma A.12. *Let λ denote the quantity defined in Theorem 3.1 and let $\mathbf{I} \equiv [0,L]^d$ be a rectangle with a fixed volume L^d. Then*

$$\Pr\left(\max_{(u,\mathbf{s}) \in [0,1] \times \mathbf{I}} \boldsymbol{\xi}(u,\mathbf{s}) > a \right) = L^d 2^{-(d-1)} \left(\frac{8\lambda}{\pi} \right)^{d/2} a^d \exp(-2a^2)[1 + o(1)]$$

as $a \to \infty$.

Lemma A.13. *For any x,*

$$\Pr\left(4\mathbf{b}_n \left\{ \max_{(u,s) \in [0,1] \times \mathcal{X}_n} \boldsymbol{\xi}(u,\mathbf{s}) - \mathbf{b}_n \right\} < x \right) = \exp\left\{ -\exp\left(-x - \frac{x^2}{8\mathbf{b}_n^2} \right) \left[1 + \frac{x}{4\mathbf{b}_n^2} \right]^d \right\} + o(1).$$
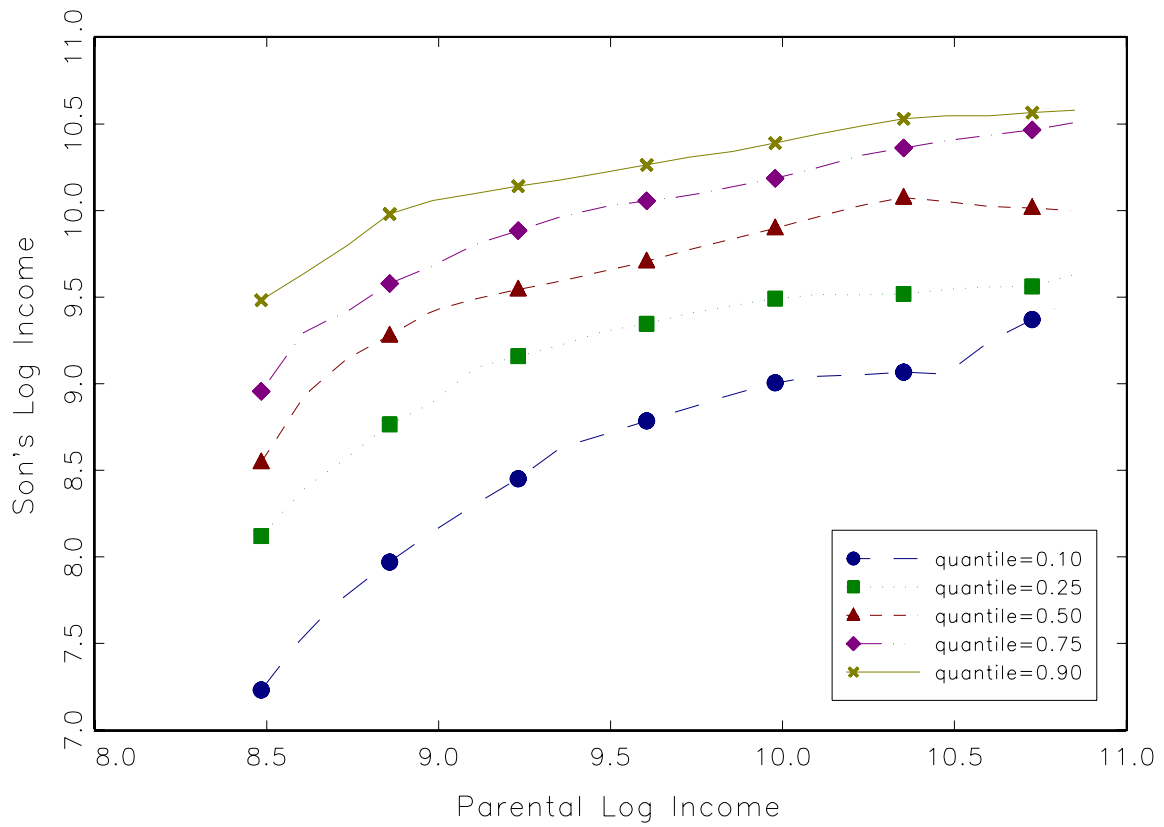
where $\mathbf{b}_n$ is defined in (16).

References

[1] 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.

[2] Bickel, P.J. and M. Rosenblatt (1973): "On some global measures of the deviations of density function estimates." *Annals of Statistics,* 1, 1071-1095.

[3] 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,* 75:2, 323-363.

[4] Bowman, A. W., M. C. Jones and I. Gijbels (1998): "Testing monotonicity of regression." *Journal of Computational and Graphical Statistics,* 7, 489-500.

[5] Buettner, T., (2003): "R&D and the Dynamics of Productivity," Manuscript, London School of Economics.

[6] Dearden, L., S. Machin, and H. Reed (1997): "Intergenerational Mobility in Britain," *Economic Journal,* 107:440, 47-66.

[7] Dawid, A.P. (1979): "Conditional independence in Statistical Theory," *Journal of the Royal Statistical Society, Series B.* 41, 1-31.

[8] Einmahl, J.H.J., and I. Van Keilegom (2006): "Goodness-of-Fit Tests in Nonparametric Regression." Tilburg University Technical Report no. 2006-79.

[9] Einmahl, U. and D.M. Mason (2005): "Uniform in bandwidth consistency of kernel-type function estimators," *Annals of Statistics,* 33, 1380-1403.

[10] Ekeland, I., J.J. Heckman, and L. Nesheim (2004): "Identification and estimation of Hedonic Models." *Journal of Political Economy,* 112, S60-S109.

[11] Ellison, G. and S. F. Ellison (2007): "Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration." MIT Department of Economics, Working paper.

[12] Ericson, R. and A. Pakes (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies,* 62, 53-82.

[13] Fan, J. and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications* (London: Chapman & Hall).

[14] Gijbels, I., P. Hall, M. C. Jones and I. Koch (2000): "Tests for monotonicity of a regression mean with guaranteed level." *Biometrika,* 87:3, 663-673.

[15] GHOSAL, S., A. SEN, AND A. W. VAN DER VAART (2000): "Testing monotonicity of regression." *Annals of Statistics,* 4, 1054-1082.

[16] HALL, P. (1979): "On the rate of convergence of normal extremes," *Journal of Applied Probability,* 16, 433-439

[17] HALL, P. (1991): "On convergence rate of suprema," *Probability Theory and Related Fields,* 89, 447-455.

[18] HALL, P. (1993): "On Edgeworth Expansion and Bootstrap Confidence Bands in Nonparametric Curve Estimation," *Journal of the Royal Statistical. Society, Ser. B.* 55, 291-304.

[19] 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.

[20] 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.

[21] HOPENHAYN, H.A. AND E.C. PRESCOTT (1992): "Stochastic Monotonicity and Stationary Distributions for Dynamic Economics," *Econometrica,* 60, 1387-1406.

[22] HOROWITZ, J.L. (2001). "The Bootstrap." in *The Handbook of Econometrics*, Vol 5. Eds J.J. Heckman and E. Leamer (2001). North Holland, Amsterdam.

[23] LEADBETTER, M.R., LINDGREN, G., AND H. ROOTZÉN (1983): *Extremes and related properties of random sequences and processes.* Springer, New York.

[24] LEVY, H., (2006): *Stochastic Dominance, Investment Decision Making under Uncertainty.* 2nd edition. Springer, Berlin.

[25] LINTON, O., E. MAASOUMI, AND Y.-J. WHANG (2005): "Consistent Testing for Stochastic Dominance under General Sampling Schemes," *Review of Economic Studies*, 72:3, 735-765.

[26] LUCAS, R.E. AND N.L. STOKEY (1989): *Recursive Methods in Dynamic Economics.* Harvard University Press, Cambridge MA.

[27] MANSKI, C. F (1997): "Monotone Treatment Response," *Econometrica*, 65:6, 1311-1334.

[28] MANSKI, C. F AND J. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68:4, 997-1010.

[29] MINICOZZI, A.L. (2003): "Estimation of sons' intergenerational earnings mobility in the presence of censoring," *Journal of Applied Econometrics*, 18:3, 291-314.

[30] NOLAN, D., AND D. POLLARD (1987): "U-processes: Rates of convergence." *Annals of Statistics,* 15, 780-799.

[31] OLLEY, G.S. AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica,* 64, 1263-1297.

[32] PAKES, A. AND R. ERICSON (1998): "Empirical implications of alternative models of firm dynamics," *Journal of Economic Theory* 79, 1-46.

[33] PAKES, A. (1986): "Patents as Options: Some estimates of the Value of Holding European Patent Stocks." *Econometrica,* 54, 755-784.

[34] PHILLIPS, P.C.B. (1979): "Normalizing transformations and expansions for functions of statistics." Research notes, Birmingham University.

[35] PHILLIPS, P.C.B. (1988): "Conditional and unconditional statistical independence." *Journal of Econometrics,* 38, 341-348.

[36] PITERBARG (1996): *Asymptotic Methods in the Theory of Gaussian Processes and Fields.* Translation of Mathematical Monographs vol 148. American Mathematical Society, Providence, Rhode Island.

[37] RIO, E., (1994): "Local invariance principles and their application to density estimation." *Probability Theory and Related Fields,* 98, 21-45.

[38] SOLON, G. (1992): "Intergenerational Income Mobility in the United States," *American Economic Review,* 82:3, 393-408.

[39] SOLON, G. (1999): "Intergenerational mobility in the labor market," in *Handbook of Labor Economics,* Volume 3, Part 1, Ch. 29, pp.1761-1800.

[40] SOLON, G. (2002): "Cross-Country Differences in Intergenerational Earnings Mobility, *Journal of Economic Perspectives,* Vol. 16, No. 3, pp.59-66.

[41] 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.

[42] STUTE, W. (1986): "Conditional Empirical Processes." *Annals of Statistics,* 14, 638-647.

[43] STUTE, W. (1996): "On almost sure convergence of conditional empirical distribution functions." *Annals of Probability,* 14, 891-901.

[44] VAN DER VAART, A. W. AND WELLNER, J. A. (1996): *Weak Convergence and Empirical Processes.* Springer, New York.

Figure 1: Intergenerational Income Mobility

Local Linear Quantile Regression Estimates

Note: This figure shows local linear quantile regression estimates of sons' log incomes on parental log incomes.

Table 1. *Simulation Results*

Using critical values obtained from
the asymptotic expansion F_n of the limiting distribution
(1500 replications in each experiment)

Sample Size	Bandwidth			
	$h_n = 0.4$	$h = 0.5$	$h = 0.6$	$h = 0.7$
Rejection proportions when the null hypothesis is true:				
$n = 50$	0.014	0.021	0.025	0.030
$n = 100$	0.028	0.033	0.034	0.034
$n = 200$	0.025	0.031	0.036	0.033
$n = 500$	0.032	0.039	0.033	0.037
Rejection proportions when the null hypothesis is false:				
$n = 50$	0.687	0.762	0.771	0.760
$n = 100$	0.976	0.988	0.989	0.977
$n = 200$	1.000	1.000	1.000	1.000

Using critical values obtained from
the type I extreme value distribution
(1500 replications in each experiment)

Sample Size	Bandwidth			
	$h_n = 0.4$	$h = 0.5$	$h = 0.6$	$h = 0.7$
Rejection proportions when the null hypothesis is true:				
$n = 50$	0.009	0.017	0.013	0.017
$n = 100$	0.022	0.024	0.022	0.021
$n = 200$	0.015	0.021	0.022	0.021
$n = 500$	0.021	0.021	0.022	0.023
Rejection proportions when the null hypothesis is false:				
$n = 50$	0.618	0.693	0.697	0.694
$n = 100$	0.966	0.976	0.983	0.965
$n = 200$	1.000	1.000	1.000	1.000

Using bootstrap critical values with 500 bootstrap samples
(500 replications in each experiment)

Sample Size	Bandwidth			
	$h_n = 0.4$	$h = 0.5$	$h = 0.6$	$h = 0.7$
Rejection proportions when the null hypothesis is true:				
$n = 50$	0.064	0.062	0.046	0.058
Rejection proportions when the null hypothesis is false:				
$n = 50$	0.814	0.872	0.880	0.856