cemmap

centre for microdata methods and practice

An adaptive test of stochastic monotonicity

Denis Chetverikov
Daniel Wilhelm
Dongwoo Kim

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP24/18

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

An Adaptive Test of Stochastic Monotonicity*

Denis Chetverikov† Daniel Wilhelm‡ Dongwoo Kim§

Abstract

We propose a new nonparametric test of stochastic monotonicity which adapts to the unknown smoothness of the conditional distribution of interest, possesses desirable asymptotic properties, is conceptually easy to implement, and computationally attractive. In particular, we show that the test asymptotically controls size at a polynomial rate, is non-conservative, and detects local alternatives that converge to the null with the fastest possible rate. Our test is based on a data-driven bandwidth value and the critical value for the test takes this randomness into account. Monte Carlo simulations indicate that the test performs well in finite samples. In particular, the simulations show that the test controls size and may be significantly more powerful than existing alternative procedures.

*This version: April 1, 2018. We thank Ivan Canay and Whitney Newey for useful comments.

†Department of Economics, University of California at Los Angeles, 315 Portola Plaza, Bunche Hall, Los Angeles, CA 90024, USA; E-Mail address: chetverikov@econ.ucla.edu.

‡Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom; E-Mail address: d.wilhelm@ucl.ac.uk. The author gratefully acknowledges financial support from the ESRC Centre for Microdata Methods and Practice at IFS (RES-589-28-0001) and the European Research Council (ERC-2014-CoG-646917-ROMIA and ERC-2015-CoG-682349).

§Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom; E-Mail address: dongwoo.kim.13@ucl.ac.uk.

1 Introduction

Monotone relationships play a significant role in economic models, and therefore developing tests of such relationships is an important task for econometric research. In this paper, we propose a new nonparametric test of the hypothesis that two random variables satisfy the stochastic monotonicity condition. Such a test is useful in many economic applications, for example for testing monotone IV assumptions (e.g. Kasy (2014), Hoderlein, Holzmann, Kasy, and Meister (2016), Chetverikov and Wilhelm (2017), Wilhelm (2017)) and for testing identifying assumptions (e.g. Matzkin (1994), Lewbel and Linton (2007), Banerjee, Mukherjee, and Mishra (2009)).[1] More generally, stochastic monotonicity plays an important role in industrial organization (e.g. Ellison and Ellison (2011)), in stochastic dynamic programming (e.g. Stokey and Lucas Jr. (1989), Ericson and Pakes (1995), Olley and Pakes (1996)), and in finance (e.g. Richardson, Richardson, and Smith (1992), Boudoukh, Richardson, Smith, and Whitelaw (1999), Patton and Timmermann (2010)), among many other fields of economics.

Consider two continuous random variables X and Y, both supported on $[0, 1]$. In this paper, we are interested in testing the null of stochastic monotonicity,

$$H_0 : F_{Y|X}(y|x') \geq F_{Y|X}(y|x'') \text{ for all } y, x', x'' \in (0, 1) \text{ with } x' \leq x'', \tag{1}$$

against the alternative

$$H_a : F_{Y|X}(y|x') < F_{Y|X}(y|x'') \text{ for some } y, x', x'' \in (0, 1) \text{ with } x' \leq x''. \tag{2}$$

We propose a new nonparametric test of (1) against (2) that possesses favorable properties relative to existing approaches. First, to the best of our knowledge it is the first test that is shown to be adaptive. This means that the test adapts to the unknown smoothness level of the functions $x \mapsto F_{Y|X}(y|x)$ through a data-driven bandwidth choice. For comparison, the implementation of non-adaptive tests requires the user to specify a bandwidth value, which is undesirable because the test results may be sensitive to the particular value that is chosen. Moreover, the non-adaptive test may have low power if the bandwidth value provided by the user is not appropriate for a particular data-generating process, and, in addition, if the user performs some search over different bandwidth values to be used in the non-adaptive test, the resulting procedure may not control size, even in large samples. Second, we show that our test is asymptotically controlling size and is non-conservative, i.e. it has limiting rejection probability not larger than the nominal level for all data-generating processes in the null and equal to the nominal level for some data-generating

[1] Chetverikov and Wilhelm (2017) have already applied our proposed procedure for testing whether their monotone IV assumption holds in the context of estimating gasoline demand functions.

processes in the null. In fact, we show that the probability of rejecting the null under the null can exceed the nominal level at most by a polynomial order, which we refer to as polynomial size control. Third, we show that the test is consistent against all fixed and against local alternatives that converge to the null with the fastest possible rate. Our critical values are computed through a simple multiplier bootstrap procedure which delivers the polynomial size control without employing any higher-order corrections as is necessary in other existing approaches based on the limit distribution, e.g. Lee, Linton, and Whang (2009). Fourth, we show in simulations that our test not only controls size, but can be significantly more powerful than existing alternative approaches. Finally, our test is very simple to implement and is computationally attractive. It only requires a nonparametric estimator of the conditional distribution that is computed once on the whole sample and does not need to be re-computed on the bootstrap samples. Importantly, our test is robust with respect to the choice of the tuning parameter underlying this nonparametric estimator, in the sense that varying the tuning parameter only yields second order changes in the rejection probabilities. We provide an R implementation of the test at `https://github.com/dongwookim1984`.

There are several alternative approaches in the literature for testing (1) against (2). Our test statistic is based on the differences of the conditional cdf for different values of the conditioning variable X and is therefore most closely related to the one proposed in Lee, Linton, and Whang (2009). An important difference is that we take the maximum over the bandwidth value to achieve adaptivity whereas they let the user specify a particular bandwidth value. In consequence, our critical value is computed in a different fashion, using a multiplier bootstrap procedure. Delgado and Escanciano (2012) and Seo (2016) construct a test statistic by comparing the empirical copula of (X, Y) with its least concave majorant. Lee, Song, and Whang (2013) and Hsu, Liu, and Shi (2016) propose tests of functional inequalities of which testing the null of stochastic monotonicity is a special case. Stochastic monotonicity implies the weaker concept of regression monotonicity, i.e. monotonicity of $x \mapsto \mathrm{E}[Y|X = x]$, and some testing approaches for this hypothesis are similar to those of the former (e.g. Ghosal, Sen, and Vaart (2000), Chetverikov (2012), and Delgado and Escanciano (2013)). The approach closest to ours is Chetverikov (2012), but there are several important differences between his and this paper. First, we test a different, stronger null hypothesis that requires an additional maximum over values of y. Second, our test statistic is different and leads to a test that is substantially easier to implement because we do not require the nonparametric estimation of his studentization factor, the conditional variance function.

2 The Test

In this section, we introduce our new test of the null of stochastic monotonicity based on an i.i.d. sample $(X_i, Y_i)_{i=1}^n$ from the distribution of the pair (X, Y). Throughout the paper, we assume that the random variables X and Y are normalized so that they both have support $[0,1]$. Let $K\colon \mathbb{R} \to \mathbb{R}$ be a kernel function and, for a bandwidth value $h > 0$, define $K_h(x) := h^{-1}K(x/h)$, $x \in \mathbb{R}$. Suppose H_0 is satisfied. Then, by the law of iterated expectations,

$$\mathrm{E}\Big[(1\{Y_i \le y\} - 1\{Y_j \le y\})\mathrm{sign}(X_i - X_j)K_h(X_i - x)K_h(X_j - x)\Big] \le 0 \tag{3}$$

for all $x, y \in (0,1)$ and $i, j = 1, \dots, n$. Denoting

$$K_{ij,h}(x) := \mathrm{sign}(X_i - X_j)K_h(X_i - x)K_h(X_j - x), \quad x \in \mathbb{R},$$

taking the sum of the left-hand side in (3) over $i, j = 1, \dots, n$, and rearranging give

$$\mathrm{E}\left[\sum_{i=1}^n 1\{Y_i \le y\}\sum_{j=1}^n (K_{ij,h}(x) - K_{ji,h}(x))\right] \le 0,$$

or, equivalently,

$$\mathrm{E}\left[\sum_{i=1}^n k_{i,h}(x)1\{Y_i \le y\}\right] \le 0, \tag{4}$$

where

$$k_{i,h}(x) := \sum_{j=1}^n (K_{ij,h}(x) - K_{ji,h}(x)) = 2\sum_{j=1}^n K_{ij,h}(x), \quad x \in \mathbb{R}.$$

Our test is based on the observation that under H_0, (4) holds for all $x \in (0,1)$ and $y \in (0,1)$. To define the test statistic T, let

$$h_{\max} := 1, \quad h_{\min} := 1/n^{1-\delta}, \quad \text{for some } \delta \in (0, 2/3],$$

and

$$\mathcal{B}_n := \left\{h_{\max}u^l\colon l = 0, 1, 2, \dots, \left[\frac{\log(h_{\max}/h_{\min})}{\log(1/u)}\right]\right\}, \quad \text{for some } u \in (0,1)$$

be a collection of bandwidth values, where the notation $[a]$ denotes the largest integer that is smaller than or equal to a. Here, $\mathcal{B}_n$ forms a geometric grid on the interval $[h_{\min}, h_{\max}]$ with the geometric step u. Also, let

$$\mathcal{X}_n := \{X_1, \dots, X_n\}$$

and

$$\mathcal{Y}_n := \Big\{l/n\colon l = 1, \dots, n-1\Big\}.$$

We define our test statistic by

$$T := \max_{(x,y,h)\in\mathcal{X}_n\times\mathcal{Y}_n\times\mathcal{B}_n} \frac{\sum_{i=1}^n k_{i,h}(x)1\{Y_i \le y\}}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}}. \tag{5}$$

The statistic T is most closely related to that in Lee, Linton, and Whang (2009). The main difference is that we take the maximum with respect to the set of bandwidth values $h \in \mathcal{B}_n$ to let the data choose the best possible bandwidth value and to achieve adaptivity of the test.

We now discuss the construction of a critical value for the test. Suppose that we would like to have a test of level (approximately) $\alpha \in (0, e^{-1})$. As demonstrated by Lee, Linton, and Whang (2009), the derivation of the asymptotic distribution of T is complicated even when $\mathcal{B}_n$ is a singleton. Moreover, when $\mathcal{B}_n$ is not a singleton, it is generally unknown whether T converges to some non-degenerate asymptotic distribution, even after an appropriate normalization. We avoid these complications by employing a multiplier bootstrap critical value. Specifically, let $e_1, \dots, e_n$ be an i.i.d. sequence of $N(0,1)$ random variables that are independent of the data and let $\widehat{F}_{Y|X}(y|x)$ be an estimator of $F_{Y|X}(y|x)$. We then define a bootstrap test statistic by

$$T^b := \max_{(x,y,h)\in\mathcal{X}_n\times\mathcal{Y}_n\times\mathcal{B}_n} \frac{\sum_{i=1}^n e_i k_{i,h}(x)(1\{Y_i \le y\} - \widehat{F}_{Y|X}(y|X_i))}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}}$$

and the critical value[2] $c(\alpha)$ as

$$c(\alpha) := (1-\alpha) \text{ conditional quantile of } T^b \text{ given the data.}$$

We reject H_0 if and only if $T > c(\alpha)$.

We emphasize how simple and computationally attractive the implementation of this test is. The test statistic itself is just the maximum of a long vector of sums of observations. The bootstrap statistic requires the computation of the nonparametric estimator $\widehat{F}_{Y|X}$ evaluated at the grid of values for y and the observed values X_i for the conditioning variable. However, this estimator has to be computed only once on the whole sample and each bootstrap iteration only introduces new draws of multipliers $e_1, \dots, e_n$ to the sum of the numerator. We provide an R implementation of the test at `https://github.com/dongwookim1984`.

[2]In the terminology of the moment inequalities literature, $c(\alpha)$ can be considered a "one-step" or "plug-in" critical value. Using similar ideas as in Chetverikov (2012), we could also consider two-step or even multi-step (stepdown) critical values. For brevity of the paper, however, we do not consider these options here.

Remark 2.1 (Testing First-Order Stochastic Dominance)**.** If X were binary rather than continuous, the null of stochastic monotonicity would reduce to that of first-order stochastic dominance of $F_{Y|X}(\cdot|0)$ by $F_{Y|X}(\cdot|1)$. Our test could be adapted to this case by removing $k_{i,h}(x)$ from the numerator, removing the denominator, and maximizing only over y. The test statistic then would be identical to that in Barrett and Donald (2003). □

3 Large Sample Properties of the Test

In this section, we derive asymptotic properties of the test proposed in Section 2. First we show that the test asymptotically controls size and is non-conservative. We also show that the probability of rejecting the null under the null can exceed the nominal level α at most by a polynomial order. Then we demonstrate that the test is consistent against all fixed and against local alternatives that converge to the null at the fastest possible rate.

We start our analysis in this section by providing the list of required regularity conditions.

Assumption 3.1 (Kernel)**.** *The kernel function* $K\colon \mathbb{R} \to \mathbb{R}$ *is such that (i)* $K(x) > 0$ *for all* $x \in (-1, 1)$*, (ii)* $K(x) = 0$ *for all* $x \notin (-1, 1)$*, (iii)* K *is continuous, and (iv)* $\int_{-\infty}^{\infty} K(x)dx = 1$*.*

Here, we assume that the kernel function K has bounded support, is continuous, and is strictly positive on the support. The last condition excludes higher-order kernels but allows us to perform search (i.e. take the maximum) over a large set of bandwidth values $\mathcal{B}_n$.

Assumption 3.2 (Joint Distribution of X and Y)**.** *(i) The distribution of* X *is absolutely continuous with respect to the Lebesgue measure on* $[0, 1]$ *with the pdf* f_X *satisfying* $c_X \le f_X(x) \le C_X$ *for all* $x \in (0, 1)$ *and some constants* $0 < c_X < C_X < \infty$*. (ii) The conditional cdf* $F_{Y|X}(y|x)$ *is such that* $c_\epsilon \le F_{Y|X}(\epsilon|x) \le F_{Y|X}(1-\epsilon|x) \le C_\epsilon$ *for all* $x \in (\epsilon, 1-\epsilon)$ *and some constants* $0 < c_\epsilon < C_\epsilon < 1$ *and* $0 < \epsilon < 1/2$*.*

This is a weak regularity condition requiring, in particular, the support of the random variable X to be $[0, 1]$ and the density of X to be bounded from above and away from zero on the support. The second part of this condition requires that for all x that are not too close to the boundary of the support of X, a non-negligible mass of the conditional distribution of Y given $X = x$ is concentrated on the interval $[\epsilon, 1-\epsilon]$.

Assumption 3.3 (Estimator of $F_{Y|X}(y|x)$)**.** *The estimator* $\widehat{F}_{Y|X}(y|x)$ *of* $F_{Y|X}(y|x)$ *satisfies*

$$\mathrm{P}\left(\max_{(x,y)\in\mathcal{X}_n\times\mathcal{Y}_n} |\widehat{F}_{Y|X}(y|x) - F_{Y|X}(y|x)| > C_F n^{-c_F}\right) \le C_F n^{-c_F}$$

for some constants $c_F, C_F > 0$.

This is a mild high-level condition implying uniform consistency of an estimator $\widehat{F}_{Y|X}(y|x)$ of $F_{Y|X}(y|x)$ over $(x,y) \in \mathcal{X}_n \times \mathcal{Y}_n$ with a polynomial rate of convergence. In Appendix A, we demonstrate that this assumption holds for the kernel estimator

$$\widehat{F}_{Y|X}(y|x) := \frac{\sum_{i=1}^n 1\{Y_i \le y\} K_b(X_i - x)}{\sum_{i=1}^n K_b(X_i - x)}, \quad x, y \in [0,1], \tag{6}$$

if we set the bandwidth value $b = b_n = 1/\sqrt{n}$ as long as Assumptions 3.1 and 3.2 are satisfied and the functions $x \mapsto F_{Y|X}(y|x)$ are Lipschitz-continuous. For a more general treatment providing conditions underlying Assumption 3.3, we refer an interested reader to Härdle, Janssen, and Serfling (1988).

It is important to notice that our test is robust with respect to the choice of the bandwidth value b for the nonparametric estimator $\widehat{F}_{Y|X}(y|x)$ in (6). In particular, varying the bandwidth value b will affect the rejection probability of the test only in the second order. Thus, although it is possible in principle to use a data-driven method for selecting the bandwidth value b that would yield an estimator $\widehat{F}_{Y|X}(y|x)$ with the fastest possible rate of convergence, there is no need to do so, and simply letting $b = b_n = 1/\sqrt{n}$ would give similar results.[3]

We are now able to state our formal results. The first theorem shows that our test asymptotically controls size and is not conservative:

Theorem 3.1 (Polynomial Size Control)**.** *Let Assumptions 3.1, 3.2, and 3.3 be satisfied. In addition, assume that $\alpha \in (0, e^{-1})$. If H_0 holds, then*

$$\mathrm{P}\left(T > c(\alpha)\right) \le \alpha + Cn^{-c}. \tag{7}$$

If the functions $x \mapsto F_{Y|X}(y|x)$ are constant for all $y \in (0,1)$, then

$$\left|\mathrm{P}\left(T > c(\alpha)\right) - \alpha\right| \le Cn^{-c}. \tag{8}$$

In both (7) and (8), c and C are constants that depend only on $c_F, C_F, c_X, C_X, c_\epsilon, C_\epsilon, u, \delta, \epsilon$, and the kernel function K.

The result (7) implies that our test asympotically controls size. The result (8) in turn strengthens this statement by showing that the rejection probability for some data-generating processes in the null is asympotically equal to the nominal level α, so the test is not conservative. Furthermore, the probability of rejecting H_0 when H_0 is satisfied

[3] Of course, if the support of X is $[c, C]$ for some constants $c < C$ rather than $[0,1]$, an appropriate bandwidth value would be $b = (C - c)/\sqrt{n}$.

can exceed the nominal level α only by a term that is polynomially small in n. We refer to this phenomenon as *polynomial size control*. As explained in Lee, Linton, and Whang (2009), when $\mathcal{B}_n$ is a singleton, convergence of T to the limit distribution is logarithmically slow. For this reason, Lee, Linton, and Whang (2009) used higher-order corrections derived in Piterbarg (1996) to obtain polynomial size control. Theorem 3.1 shows that the multiplier bootstrap also leads to the polynomial size control, even though no higher-order corrections are required. To prove Theorem 3.1, we rely on the high-dimensional CLT and bootstrap results in Chernozhukov, Chetverikov, and Kato (2013, 2017).

The constants c and C in (7) and (8) depend on the data generating process only via constants and the kernel function appearing in Assumptions 3.1, 3.2, and 3.3. Therefore, inequalities (7) and (8) hold uniformly over all data-generating processes satisfying these assumptions with the same constants. In this sense, our test provides uniform size control.

Remark 3.1 (Weak Condition on the Bandwidth Values)**.** As we set $h_{\min} = 1/n^{1-\delta}$ for some $\delta \in (0,1)$, our theorem requires

$$\frac{1}{nh} \leq C_h n^{-c_h} \tag{9}$$

for all $h \in \mathcal{B}_n$ and some constants $c_h, C_h > 0$, which is considerably weaker than the analogous condition in Lee, Linton, and Whang (2009), who require $1/(nh^3) \to 0$, up to logarithmic terms. As follows from the proof of the theorem, the multiplier bootstrap distribution approximates the conditional distribution of the test statistic given $\mathcal{X}_n = \{X_1, \dots, X_n\}$. Conditional on $\mathcal{X}_n$, the denominator in the definition of T is fixed, and does not require any approximation. Instead, we could try to approximate the denominator of T by its probability limit. This is done in Ghosal, Sen, and Vaart (2000) using the theory of Hoeffding projections (in a different setting) but they require the condition $1/nh^2 \to 0$, which is also stronger than our condition (9). □

Our second result in this section concerns the ability of our test to detect fixed models in the alternative H_a.

Theorem 3.2 (Consistency)**.** *Let Assumptions 3.1, 3.2, and 3.3 be satisfied and assume that $(x, y) \mapsto F_{Y|X}(y|x)$ is continuously differentiable. If H_a holds with*

$$\frac{\partial}{\partial x} F_{Y|X}(y^*|x^*) > 0, \quad \textit{for some } x^*, y^* \in (0,1)$$

then

$$\mathrm{P}\left(T > c(\alpha)\right) \to 1 \ \textit{as } n \to \infty. \tag{10}$$

This theorem shows that our test is consistent against any model in H_a (with smooth $(x,y) \mapsto F_{Y|X}(y|x)$), which is typically considered a necessary condition for the test to be useful. Note, however, that the result (10) only shows that if the sample size n is large enough, the probability of rejecting the null when the null does not hold will be close to one. The result does not specify, on the other hand, how large the sample size n has to be in order for the rejection probability to be close one. We therefore complement the result in Theorem 3.2 by deriving the rate of consistency of our test against local alternatives.

To this end, we introduce the triangular array $\{(X_{i,n}, Y_{i,n})\colon i = 1, \dots, n\}_{n\geq 1}$, where for each $n \geq 1$, $(X_{i,n}, Y_{i,n})_{i=1}^n$ is an i.i.d. sample from the distribution of the pair (X^n, Y^n), and the distribution of (X^n, Y^n) can vary with n. Let

$$F_{Y|X,n}(y|x) := \mathrm{P}_n(Y^n \leq y \mid X^n = x), \quad x, y \in [0,1],$$

denote the conditional cdf of the distribution of Y^n given X^n.

Assumption 3.4 (Smoothness)**.** *For all $n \geq 1$, $(x,y) \mapsto F_{Y|X,n}(y|x)$ is twice continuously differentiable and*

$$\left|\frac{\partial}{\partial x} F_{Y|X,n}(y|x)\right| \leq C_L, \quad \text{for all } x, y \in [0,1],$$
$$\left|\frac{\partial^2}{\partial x \partial y} F_{Y|X,n}(y|x)\right| \leq C_L, \quad \text{for all } x, y \in [0,1],$$
$$\left|\frac{\partial}{\partial x} F_{Y|X,n}(y|x_2) - \frac{\partial}{\partial x} F_{Y|X,n}(y|x_1)\right| \leq C_L |x_2 - x_1|^\beta, \quad \text{for all } x_1, x_2, y \in [0,1],$$

for all $n \geq 1$ and some constants $0 < C_L < \infty$ and $\beta \in (0,1)$.

The first two conditions in this assumption require the sequence of conditional cdfs $F_{Y|X,n}(y|x)$ to have a bounded first derivative with respect to x and cross derivative, where the bound is independent of n. The third condition requires the derivative of $F_{Y|X,n}(y|x)$ with respect to x to be Hölder continuous in x with constant and exponent that are independent of n.

Theorem 3.3 (Rate of Consistency)**.** *Let Assumptions 3.1, 3.2, and 3.3 be satisfied for all $n \geq 1$ with the same constants c_F, c_F, c_X, C_X, c_ϵ, C_ϵ, and ϵ, and the same kernel K, where we replace (X, Y) in Assumption 3.2 by (X^n, Y^n) and $F_{Y|X}$ in Assumption 3.3 by $F_{Y|X,n}$. In addition, suppose that Assumption 3.4 holds. If for all n and some sequence of positive constants $(\ell_n)_{n\geq 1}$ such that $(\log n/n)^{\beta/(2\beta+3)} = o(\ell_n)$ and*

$$\frac{\partial}{\partial x} F_{Y|X,n}(y^*|x^*) > \ell_n, \quad \text{for some } x^*, y^* \in (0,1),$$

then

$$\mathrm{P}\left(T > c(\alpha)\right) \to 1 \text{ as } n \to \infty. \tag{11}$$

This theorem shows that our test is consistent against local alternatives for which the size of the deviation of $\partial F_{Y|X,n}(y|x)/\partial x$ from zero converges to zero at a rate slower than $(\log n/n)^{\beta/(2\beta+3)}$. Using the standard arguments, e.g. in Dümbgen and Spokoiny (2001), it is possible to show that this is the fastest possible rate with which the alternatives can converge to the null if we hope to be able to detect them (in the minimax sense). We therefore conclude that our test is consistent against the alternatives that converge to the null with the fastest possible rate.

Remark 3.2 (Testing First-Order Stochastic Dominance)**.** As indicated in Remark 2.1, one could modify our test to accommodate the case in which X is binary, leading to a test of first-order stochastic dominance. The results of this section then imply that our test, which would be equivalent to that in Barrett and Donald (2003), satisfies polynomial size control and polynomial rate of consistency. These desirable properties were not shown in Barrett and Donald (2003). □

4 Simulations

In this section, we describe a simulation experiment which illustrates the finite sample performance of our test and compare it to other alternatives. The design is based on Delgado and Escanciano (2012) and the 2014 working paper version of Seo (2016).[4] We simulate 1,000 Monte Carlo samples of sizes 100, 200 and 300 from the following four data generating processes:

N1: $Y_i = U_i$

A1: $Y_i = -0.1X_i + U_i$

A2: $Y_i = -0.1X_i^2 + U_i$

A3: $Y_i = 0.2X_i - \beta\exp(-250(X_i - 0.5)^2) + U_i$

where $\beta = 0.2$, X_i is uniformly distributed on the unit interval, and U_i is drawn from $N(0, 0.1^2)$. By construction, X_i and U_i are independent of each other for all i. In the model $N1$, Y_i and X_i are independent, so the null hypothesis holds. Models $A2$ and $A3$ are models in the alternative hypothesis for which the null is violated at every conditioning value of X_i. Model $A4$, on the other hand, is an alternative that deviates from the null only locally. The right panel of Figure 1 shows the conditional mean function.

[4]The only reason why we base our simulations on the 2014 working paper version of Seo (2016) rather than the 2016 version is that the former contains simulation results for various sample sizes whereas the latter only reports results for very small ones ($n = 70$ and $n = 120$).

Model	n	CWK	S-L^1	S-L^2	S-L^∞	DE	$LLW_{0.5}$	$LLW_{0.6}$	$LLW_{0.7}$
N1	100	0.051 (0.175)	0.055	0.048	0.049	0.046	0.034	0.035	0.036
	200	0.050 (0.127)	0.054	0.053	0.053	0.052	0.031	0.034	0.033
	300	0.064 (0.129)	0.062	0.053	0.044	0.042	0.036	0.039	0.039
A1	100	0.412 (0.574)	0.877	0.828	0.653	0.634	0.408	0.542	0.612
	200	0.667 (0.740)	0.988	0.980	0.911	0.880	0.749	0.853	0.908
	300	0.885 (0.860)	0.999	1.000	0.995	0.980	0.911	0.964	0.980
A2	100	0.418 (0.580)	0.874	0.806	0.620	0.599	0.469	0.587	0.651
	200	0.693 (0.738)	0.990	0.981	0.938	0.906	0.805	0.892	0.925
	300	0.902 (0.843)	1.000	1.000	0.995	0.981	0.938	0.972	0.983
A3	100	0.339 (0.144)	0.003	0.030	0.154	0.032	0.012	0.013	0.022
	200	0.690 (0.138)	0.003	0.054	0.304	0.157	0.014	0.009	0.014
	300	0.929 (0.143)	0.004	0.178	0.539	0.382	0.012	0.009	0.009

Table 1: Rejection probabilities of our test and the tests of Seo (2016), Delgado and Escanciano (2012), and Lee, Linton, and Whang (2009). The nominal size is 0.05. The values in the parentheses of CWK are the optimal bandwidths chosen by the test, averaged over the simulation samples. The results for the other tests besides CWK are taken from the 2014 working paper version of Seo (2016).

For the implementation of our test, we choose the Epanechnikov kernel for K and construct the set of bandwidth values $\mathcal{B}_n$ using $u = 2/3$. The number of elements in $\mathcal{B}_n$ are 7, 8, and 8, for the three sample sizes $n = 100$, $n = 200$, and $n = 300$, respectively. To estimate the conditional cdf of Y given X, we use the estimator defined in (6) with $b = n^{-1/2}$. The multiplier bootstrap critical values are computed based on 200 bootstrap samples with Gaussian multipliers and nominal size of the test is chosen to be 0.05.

Table 1 shows the empirical rejection frequencies of various tests in each of the four models and each of the three sample sizes. "CWK" refers to our new test. The values in parentheses are the optimal bandwidths that our test chooses (i.e. the bandwidth value at which our test statistic achieves the maximum), averaged over the simulation samples. "S-L^1", "S-L^2", and "S-L^∞" refer to the L^1-, L^2-, and L^∞-versions of Seo (2016)'s test, "DE" to Delgado and Escanciano (2012)'s test, and "$LLW_{0.5}$", "$LLW_{0.6}$", and "$LLW_{0.7}$" to the test of Lee, Linton, and Whang (2009) using the bandwidth values 0.5, 0.6, and 0.7, respectively.

Like all other tests ours also controls size well across all dgps and sample sizes, although LLW appears somewhat conservative. Under the alternative models A1–A3, the performance of the tests differs. Under A1 and A2, all tests generally perform well in the sense that their rejection frequencies are around 0.4 for the small sample size and are closer to one as the sample size increases. Seo (2016)'s test clearly dominates all other tests at small sample sizes with significantly higher rejection frequencies, especially the L^1-, L^2- versions of the test. Under the alternative A3, the relative behavior of the tests is very different. The rejection frequencies of our test are very similar to those for the other alternatives of about 0.3, 0.7, and 0.9 as the sample size increases, but all other tests perform significantly worse, struggling to detect the local deviation from the null. Most of the rejection frequencies of the other tests are close to zero. Only DE and S-L^∞ start rejecting at the larger sample sizes, but their rejection frequencies still remain far below those of our test. To understand why our test performs similarly well for global as well as local alternatives, it is helpful to consider the bandwidths that our test automatically chooses. They are relatively large for alternatives A1 and A2 (around 0.6 to 0.9), but substantially smaller for A3 (around 0.1). The small bandwidth in model A3 is the reason for why our test is able to detect the local deviation. On the other hand, tests such as LLW employ a fixed, user-chosen bandwidth which leads to power against some (A1 and A2), but not against other (A3) alternatives.

To further investigate the power properties of our test, we generate 1,000 MC samples of size $n = 100$ from variations of $A3$ in which we vary β. The right panel of Figure 1 shows how increasing β increases the size of the local deviation from the null. All other aspects of the dgp A3 remain the same as above. The left panel of Figure 1 shows the empirical rejection frequencies of our test compared to LLW with a variety of different bandwidths as a function β. As the size of the local deviation from the null increases all tests reject more frequently. However, the power of the LLW test varies substantially as we vary the bandwidth from 0.1 to 0.6, with the power being largest for the bandwidth of 0.2, but it is significantly less powerful than ours for all bandwidth values and all values of β.

In conclusion, our test controls size and is able to adapt well to the smoothness of the conditional cdf of $Y|X$, which allows it to perform similarly well against all alternatives, while the other tests considered here perform well against some, but not all alternatives.

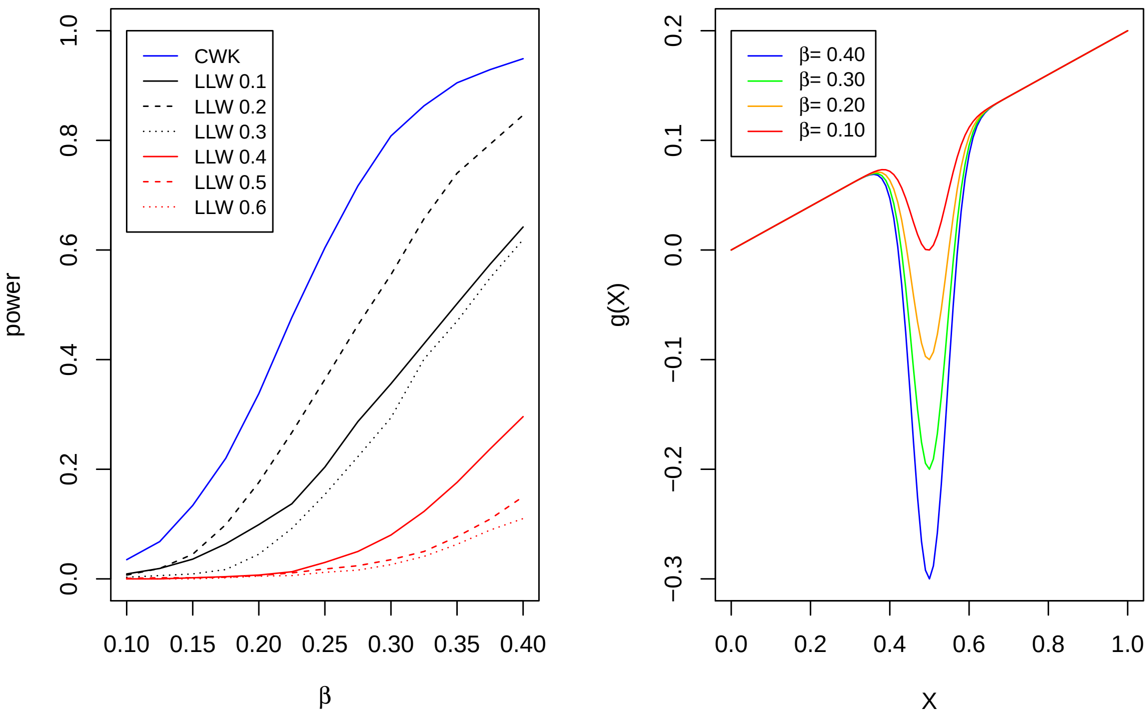

Figure 1: The power curves under various degrees of local deviation from the null.

A Verification of Assumption 3.3

In this section, we verify Assumption 3.3. For concreteness, we work with kernel estimators but the same assumption can be verified for other nonparametric estimators as well. For the same kernel function $K\colon \mathbb{R} \to \mathbb{R}$ as that used in Sections 2 and 3 and a bandwidth value $b > 0$, define $K_b(x) := b^{-1}K(x/b)$, $x \in \mathbb{R}$. The kernel estimator of $F_{Y|X}(y|x)$ is then given by

$$\widehat{F}_{Y|X}(y|x) := \frac{\sum_{i=1}^n 1\{Y_i \le y\}K_b(X_i - x)}{\sum_{i=1}^n K_b(X_i - x)}, \quad x, y \in [0,1].$$

Also, as in Section 2, define $\mathcal{X}_n := \{X_1, \dots, X_n\}$ and $\mathcal{Y}_n := \{l/n\colon l = 1, \dots, n-1\}$. To verify Assumption 3.3, we will impose Assumptions 3.1 and 3.2 and assume that the functions $x \mapsto F_{Y|X}(y|x)$ are Lipschitz-continuous. The latter condition is very mild, and some form of it is typically imposed in the literature on nonparametric estimation. The following lemma shows that the kernel estimator $\widehat{F}_{Y|X}(y|x)$ satisfies Assumption 3.3 under the aforementioned conditions as long as the bandwidth b does not go to zero too quickly.

Lemma A.1 (Verification of Assumption 3.3)**.** *Let Assumptions 3.1 and 3.2 be satisfied and assume that*

$$|F_{Y|X}(y|x_2) - F_{Y|X}(y|x_1)| \le C_L|x_2 - x_1|, \quad \textit{for all } x_1, x_2, y \in [0,1]$$

for some constant $C_L > 0$. Also, assume that the bandwidth value b is such that $\log n \le nb$. Then there exists a constant $C > 0$ depending only on c_X, C_X, C_L, and the kernel function K such that

$$\mathrm{P}\left(\max_{(x,y)\in\mathcal{X}_n\times\mathcal{Y}_n} |\widehat{F}_{Y|X}(y|x) - F_{Y|X}(y|x)| > C\left(b + \sqrt{\frac{\log n}{nb}}\right)\right) \le \frac{1}{n},$$

which implies that Assumption 3.3 holds if $b = b_n$ is set so that $b \le Cn^{-c}$ and $\log n/(nb) \le Cn^{-c}$ for some constants $c, C > 0$. In particular, we can set $b = b_n = 1/\sqrt{n}$.

Proof. By Assumption 3.1, K is a continuous function with support $[-1,+1]$, and so there exists a constant $C_K > 0$ such that $K(x) \le C_K$ for all $x \in \mathbb{R}$. Therefore, using the change of variables formula, we obtain that for all $x \in [0,1]$,

$$\mathrm{E}[K_b(X-x)^2] = \frac{1}{b^2}\int_{-\infty}^{+\infty} K\left(\frac{s-x}{b}\right)^2 f_X(s)ds = \frac{1}{b}\int_{-1}^{+1} K(t)^2 f_X(x+tb)dt \le \frac{2C_XC_K^2}{b}$$

by Assumptions 3.1 and 3.2. Hence, by Bernstein's inequality,

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^n K_b(X_i - x) - \mathrm{E}[K_b(X-x)]\right| > t\right) \le 2\exp\left(\frac{-t^2/2}{\mathrm{E}[K_b(X-x)^2]/n + C_Kt/(nb)}\right)$$
$$\le 2\exp\left(\frac{-nbt^2/2}{2C_XC_K^2 + C_Kt}\right).$$

Thus, given that $\log n \le nb$, there exists a constant C_1 depending only on C_X and C_K such that

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^n K_b(X_i - x) - \mathrm{E}[K_b(X-x)]\right| > C_1\sqrt{\frac{\log n}{nb}}\right) \le \frac{1}{2n^3}.$$

Further, observe that for all $x, y \in [0,1]$,

$$1\{Y \le y\}K_b(X-x) \le K_b(X-x) \le C_K/b$$

almost surely, and so by the same argument,

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^n 1\{Y_i \le y\}K_b(X_i - x) - \mathrm{E}[1\{Y \le y\}K_b(X-x)]\right| > C_1\sqrt{\frac{\log n}{nb}}\right) \le \frac{1}{2n^3}.$$

Next, for all $x, y \in [0,1]$,

$$\begin{aligned}
&\Big|\mathrm{E}[1\{Y \le y\}K_b(X-x)] - F_{Y|X}(y|x)\mathrm{E}[K_b(X-x)]\Big| \\
&\quad = \Big|\mathrm{E}[F_{Y|X}(y|X)K_b(X-x)] - F_{Y|X}(y|x)\mathrm{E}[K_b(X-x)]\Big| \\
&\quad \le \Big|\mathrm{E}[(F_{Y|X}(y|X) - F_{Y|X}(y|x))K_b(X-x)]\Big| \le C_L b\mathrm{E}[K_b(X-x)] \le 2C_L C_X C_K b
\end{aligned}$$

by the Lipschitz property of the function $x \mapsto F_{Y|X}(y|x)$.

Now, given that $|\mathcal{X}_n \times \mathcal{Y}_n| = n(n-1) \le n^2$, combining the presented inequalities and using the union bound shows that for all $(x,y) \in \mathcal{X}_n \times \mathcal{Y}_n$,

$$\widehat{F}_{Y|X}(y|x) = \frac{F_{Y|X}(y|x)\mathrm{E}[K_b(X-x)] + N_{x,y}}{\mathrm{E}[K_b(X-x)] + D_{x,y}},$$

where $N_{x,y}$ and $D_{x,y}$ are random variables such that

$$\mathrm{P}\left(\max_{(x,y)\in\mathcal{X}_n\times\mathcal{Y}_n} |N_{x,y}| \vee |D_{x,y}| > C_2\left(b + \sqrt{\frac{\log n}{nb}}\right)\right) \le \frac{1}{n},$$

where $C_2 > 0$ is a constant depending only C_X, C_L, and the kernel function K. Thus, given that $\mathrm{E}[K_b(X-x)] \ge c$ for all $x \in \mathcal{X}_n$ and some constant $c > 0$ depending only on c_X and the kernel function K, the asserted claim follows. Q.E.D.

B Preliminary Lemmas

In this section, c and C are understood as sufficiently small and large constants, respectively, whose values may change at each appearance. The constants c and C can be chosen to depend only on $c_F, C_F, c_X, C_X, c_\epsilon, C_\epsilon, u, \delta, \epsilon$, and the kernel function K. Also, denote

$$\mathcal{S}_n := \left\{(x,y,h) \in \mathcal{X}_n \times \mathcal{Y}_n \times \mathcal{B}_n \colon \frac{\sum_{i=1}^n k_{i,h}(x)^2 F_{Y|X}(y|X_i)(1 - F_{Y|X}(y|X_i))}{\sum_{i=1}^n k_{i,h}(x)^2} \ge \frac{1}{\log^2 n}\right\}$$

and let $\mathcal{S}_n^c$ be the complement of $\mathcal{S}_n$ relative to $\mathcal{X}_n \times \mathcal{Y}_n \times \mathcal{B}_n$. Moreover, for all $\gamma \in (0,1)$, let $c_0(\gamma)$ be the $(1-\gamma)$ conditional quantile of

$$T_0^b := \max_{(x,y,h)\in\mathcal{S}_n} \frac{\sum_{i=1}^n e_i k_{i,h}(x)(1\{Y_i \le y\} - \widehat{F}_{Y|X}(y|X_i))}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}}$$

given the data and let $\bar{c}_0(\gamma)$ be the $(1-\gamma)$ conditional quantile of

$$\bar{T}_0^b := \max_{(x,y,h)\in\mathcal{S}_n} \frac{\sum_{i=1}^n e_i k_{i,h}(x)(1\{Y_i \le y\} - F_{Y|X}(y|X_i))}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}}$$

given the data. Below we state and prove several lemmas that are useful for the proof of Theorem 3.1.

Lemma B.1. *Under Assumptions 3.1 and 3.2,*

$$\max_{1\le i\le n} \max_{(x,h)\in\mathcal{X}_n\times\mathcal{B}_n} \frac{|\sqrt{n} k_{i,h}(x)|}{(\sum_{l=1}^n k_{l,h}(x)^2)^{1/2}} \le B_n := (Cn^{1-\delta})^{1/2}$$

with probability at least $1 - Cn^{-c}$.

Proof of Lemma B.1. The claim follows from the proof of Theorem 4.2 in Chetverikov (2012), which can be seen by noting that in the notation of Chetverikov (2012),

$$A_n = \max_{1\le i\le n} \max_{(x,h)\in\mathcal{X}_n\times\mathcal{B}_n} \frac{|k_{i,h}(x)|}{(\sum_{l=1}^n k_{l,h}(x)^2)^{1/2}}$$

if we set $k = 0$ and $\sigma_i^2 = 1$ for all $i = 1, \dots, n$ there. Also, in the proof of Theorem 4.2 in Chetverikov (2012), all statements "with probability $1 - o(1)$" can be replaced by "with probability at least $1 - Cn^{-c}$". Q.E.D.

Lemma B.2. *Under Assumptions 3.1 and 3.2,*

$$\max_{(x,y,h)\in\mathcal{S}_n^c} \frac{|\sum_{i=1}^n k_{i,h}(x)(1\{Y_i \le y\} - F_{Y|X}(y|X_i))|}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}} \le \frac{C}{\sqrt{\log n}}$$

with probability at least $1 - Cn^{-c}$.

Proof of Lemma B.2. Note that for any $(x,y,h) \in \mathcal{S}_n^c$,

$$\sum_{i=1}^n \mathrm{E}\left[\frac{k_{i,h}(x)^2(1\{Y_i \le y\} - F_{Y|X}(y|X_i))^2}{\sum_{l=1}^n k_{l,h}(x)^2} \mid \mathcal{X}_n\right]$$
$$= \frac{\sum_{i=1}^n k_{i,h}(x)^2 F_{Y|X}(y|X_i)(1 - F_{Y|X}(y|X_i))}{\sum_{l=1}^n k_{l,h}(x)^2} < \frac{1}{\log^2 n}$$

by the definition of the set $\mathcal{S}_n^c$. Thus, given that

$$|1\{Y_i \le y\} - F_{Y|X}(y|X_i)| \le 2$$

for all $i = 1, \dots, n$ and $y \in \mathcal{Y}_n$ almost surely, applying Bernstein's inequality conditional on $\mathcal{X}_n$ and using Lemma B.1 and the union bound shows that for any $t > 0$,

$$\mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n^c}\left|\frac{|\sum_{i=1}^n k_{i,h}(x)(1\{Y_i \le y\} - F_{Y|X}(y|X_i))|}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}}\right| > t\right) \le 2\exp\left(C\log n - \frac{-t^2/2}{\log^{-2} n + 2B_n t/\sqrt{n}}\right) + Cn^{-c}$$

since $\log|\mathcal{S}_n^c| \le C\log n$. Substituting $t = C/\sqrt{\log n}$ and $B_n = (Cn^{1-\delta})^{1/2}$ in this inequality gives the asserted claim. Q.E.D.

Lemma B.3. *Under Assumptions 3.1, 3.2, and 3.3,*

$$\max_{(x,y,h)\in\mathcal{S}_n^c}\frac{\sum_{i=1}^n k_{i,h}(x)^2(1\{Y_i \le y\} - \widehat{F}_{Y|X}(y|X_i))^2}{\sum_{i=1}^n k_{i,h}(x)^2} \le \frac{C}{\log^2 n}$$

with probability at least $1 - Cn^{-c}$.

Proof of Lemma B.3. Note that

$$\max_{(x,y,h)\in\mathcal{S}_n^c}\frac{\sum_{i=1}^n k_{i,h}(x)^2(1\{Y_i \le y\} - \widehat{F}_{Y|X}(y|X_i))^2}{\sum_{i=1}^n k_{i,h}(x)^2} \tag{12}$$

$$\le \max_{(x,y,h)\in\mathcal{S}_n^c}\frac{2\sum_{i=1}^n k_{i,h}(x)^2(1\{Y_i \le y\} - F_{Y|X}(y|X_i))^2}{\sum_{i=1}^n k_{i,h}(x)^2} \tag{13}$$

$$+ \max_{(x,y,h)\in\mathcal{S}_n^c}\frac{2\sum_{i=1}^n k_{i,h}(x)^2(\widehat{F}_{Y|X}(y|X_i) - F_{Y|X}(y|X_i))^2}{\sum_{i=1}^n k_{i,h}(x)^2} \tag{14}$$

and the term in (14) is bounded from above by

$$2\max_{(x,y)\in\mathcal{X}_n\times\mathcal{Y}_n}|\widehat{F}_{Y|X}(y|x) - F_{Y|X}(y|x)|^2 \le Cn^{-c}$$

with probability at least $1 - Cn^{-c}$ by Assumption 3.3. To bound the term in (13), note that for each $(x,y,h) \in \mathcal{S}_n^c$,

$$\mathrm{E}\left[\frac{\sum_{i=1}^n k_{i,h}(x)^2(1\{Y_i \le y\} - F_{Y|X}(y|X_i))^2}{\sum_{i=1}^n k_{i,h}(x)^2} \mid \mathcal{X}_n\right]$$

$$\mathrm{E}\left[\frac{\sum_{i=1}^n k_{i,h}(x)^2 F_{Y|X}(y|X_i)(1 - F_{Y|X}(y|X_i))}{\sum_{i=1}^n k_{i,h}(x)^2} \mid \mathcal{X}_n\right] < \frac{1}{\log^2 n}$$

by the definition of the set $\mathcal{S}_n^c$, and so by the second part of Lemma D.1, applied conditional on $\mathcal{X}_n$,

$$\mathrm{E}\left[\max_{(x,y,h)\in\mathcal{S}_n^c}\frac{\sum_{i=1}^n k_{i,h}(x)^2(1\{Y_i\le y\}-F_{Y|X}(y|X_i))^2}{\sum_{i=1}^n k_{i,h}(x)^2}\mid\mathcal{X}_n\right]$$
$$\le C\left(\frac{1}{\log^2 n}+\max_{1\le i\le n}\max_{(x,h)\in\mathcal{X}_n\times\mathcal{B}_n}\frac{k_{i,h}(x)^2\log n}{\sum_{l=1}^n k_{l,h}(x)^2}\right)\le C\left(\frac{1}{\log^2 n}+\frac{B_n^2\log n}{n}\right)\le\frac{C}{\log^2 n}$$

with probability at least $1-Cn^{-c}$, where the second and the third inequalities in the second line holds by Lemma B.1 and the definition of B_n, respectively. Hence, applying the first part of Lemma D.1 with $s=1$ and $t=Cn^{-c}$ conditional on $\mathcal{X}_n$ shows that

$$\max_{(x,y,h)\in\mathcal{S}_n^c}\frac{\sum_{i=1}^n k_{i,h}(x)^2(1\{Y_i\le y\}-F_{Y|X}(y|X_i))^2}{\sum_{i=1}^n k_{i,h}(x)^2}\le\frac{C}{\log^2 n}$$

with probability at least $1-Cn^{-c}$. Combining presented inequalities gives the asserted claim. Q.E.D.

Lemma B.4. *Under Assumptions 3.1, 3.2, and 3.3, for all* $\alpha\in(0,e^{-1})$,

$$c(\alpha)\ge c \text{ and } c_0(\alpha)\le c(\alpha)\le c_0(\alpha-1/n)$$

with probability at least $1-Cn^{-c}$.

Proof of Lemma B.4. By Assumption 3.2 and the choice of the sets $\mathcal{X}_n$, $\mathcal{Y}_n$, and $\mathcal{B}_n$, with probability at least $1-Cn^{-c}$, there exists $(\bar{x},\bar{y},\bar{h})\in\mathcal{X}_n\times\mathcal{Y}_n\times\mathcal{B}_n$ such that $\epsilon<\bar{x}-\bar{h}<\bar{x}+\bar{h}<1-\epsilon$ and $\epsilon<\bar{y}<1-\epsilon$, in which case

$$T^b\ge\frac{\sum_{i=1}^n e_i k_{i,\bar{h}}(\bar{x})(1\{Y_i\le\bar{y}\}-\widehat{F}_{Y|X}(\bar{y}|X_i))}{(\sum_{i=1}^n k_{i,\bar{h}}(\bar{x})^2)^{1/2}},$$

and conditional on the data, the random variable on the right-hand side of this inequality is zero-mean Gaussian with variance

$$\frac{\sum_{i=1}^n k_{i,\bar{h}}(\bar{x})^2(1\{Y_i\le\bar{y}\}-\widehat{F}_{Y|X}(\bar{y}|X_i))^2}{\sum_{i=1}^n k_{i,\bar{h}}(\bar{x})^2}$$
$$\ge\min_{i:\,\epsilon<X_i<1-\epsilon}\left(\widehat{F}_{Y|X}(\bar{y}|X_i)^2\wedge(1-\widehat{F}_{Y|X}(\bar{y}|X_i))^2\right)\ge c$$

with probability at least $1-Cn^{-c}$ by Assumptions 3.1, 3.2, and 3.3. This implies the first asserted claim: $c(\alpha)\ge c$ with probability at least $1-Cn^{-c}$.

To prove the second asserted claim, note that $c_0(\alpha)\le c(\alpha)$ almost surely because $T_0^b\le T^b$ almost surely. Therefore, it remains to show that $c(\alpha)\le c_0(\alpha-Cn^{-c})$ with probability at least $1 - Cn^{-c}$. To do so, observe that for each $(x, y, h) \in \mathcal{S}_n^c$, conditional on the data, the random variable

$$\frac{\sum_{i=1}^n e_i k_{i,h}(x)(1\{Y_i \leq y\} - \widehat{F}_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}}$$

is zero-mean Gaussian with variance

$$\frac{\sum_{i=1}^n k_{i,h}(x)^2(1\{Y_i \leq y\} - \widehat{F}_{Y|X}(y|X_i))^2}{\sum_{i=1}^n k_{i,h}(x)^2} \leq \frac{C}{\log^2 n}$$

with probability at least $1 - Cn^{-c}$ uniformly over $(x, y, h) \in \mathcal{S}_n^c$ by Lemma B.3. Hence, by Borell's inequality,

$$\mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n^c} \frac{\sum_{i=1}^n e_i k_{i,h}(x)(1\{Y_i \leq y\} - \widehat{F}_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}} \leq \frac{C}{\sqrt{\log n}} \mid (X_i, Y_i)_{i=1}^n\right) \geq 1 - \frac{1}{n} \tag{15}$$

with probability at least $1 - Cn^{-c}$. Now, on the event $\mathcal{A}_n$ that $c(\alpha) \geq c \geq c/\sqrt{\log n}$ and (15) holds, we have

$$\begin{aligned}
&\mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n} \frac{\sum_{i=1}^n e_i k_{i,h}(x)(1\{Y_i \leq y\} - \widehat{F}_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}} \leq c(\alpha) \mid (X_i, Y_i)_{i=1}^n\right) \\
&\leq \mathrm{P}\left(\max_{(x,y,h)\in\mathcal{X}_n\times\mathcal{Y}_n\times B_n} \frac{\sum_{i=1}^n e_i k_{i,h}(x)(1\{Y_i \leq y\} - \widehat{F}_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}} \leq c(\alpha) \mid (X_i, Y_i)_{i=1}^n\right) \\
&\quad + \mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n^c} \frac{\sum_{i=1}^n e_i k_{i,h}(x)(1\{Y_i \leq y\} - \widehat{F}_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}} > c(\alpha) \mid (X_i, Y_i)_{i=1}^n\right) \\
&\leq 1 - \alpha + \mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n^c} \frac{\sum_{i=1}^n e_i k_{i,h}(x)(1\{Y_i \leq y\} - \widehat{F}_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}} > \frac{c}{\sqrt{\log n}} \mid (X_i, Y_i)_{i=1}^n\right) \\
&\leq 1 - \alpha + 1/n,
\end{aligned}$$

which implies that $c_0(\alpha - 1/n) \geq c(\alpha)$. Thus, since $\mathrm{P}(\mathcal{A}_n) \geq 1 - Cn^{-c}$ by the discussion above, the second asserted claim follows. This completes the proof of the theorem. Q.E.D.

Lemma B.5. *Under Assumption 3.3, for any* $\gamma \in (Cn^{-c}, 1 - Cn^{-c})$,

$$\bar{c}_0(\gamma + Cn^{-c}) \leq c_0(\gamma) \leq \bar{c}_0(\gamma - Cn^{-c})$$

with probability at least $1 - Cn^{-c}$.

Proof of Lemma B.5. To prove the asserted claim, note that

$$|T_0^b - \bar{T}_0^b| \leq \max_{(x,y,h)\in\mathcal{S}_n} \left| \frac{\sum_{i=1}^n e_i k_{i,h}(x)(\widehat{F}_{Y|X}(y|X_i) - F_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}} \right|$$

and that conditional on the data, for each $(x, y, h) \in \mathcal{S}_n$, the random variable under the modulus on the right-hand side of this inequality is zero-mean Gaussian with variance

$$\frac{\sum_{i=1}^n k_{i,h}(x)^2(\widehat{F}_{Y|X}(y|X_i) - F_{Y|X}(y|X_i))^2}{\sum_{i=1}^n k_{i,h}(x)^2} \leq \max_{(x,y)\in\mathcal{X}_n\times\mathcal{Y}_n} |\widehat{F}_{Y|X}(y|x) - F_{Y|X}(y|x)|^2 \leq Cn^{-c}$$

with probability at least $1 - Cn^{-c}$, in which case we have by Borell's inequality that

$$\mathrm{P}\Big(|T_0^b - \bar{T}_0^b| \leq Cn^{-c}|(X_i, Y_i)_{i=1}^n\Big) \geq 1 - \frac{1}{n}.$$

Therefore, for all $\gamma \in (1/n, 1 - 1/n)$,

$$\bar{c}_0(\gamma + 1/n) - Cn^{-c} \leq c_0(\gamma) \leq \bar{c}_0(\gamma - 1/n) + Cn^{-c} \tag{16}$$

with probability at least $1 - Cn^{-c}$. In addition, by Lemma D.2,

$$\bar{c}_0(\gamma - 1/n) + Cn^{-c} \leq \bar{c}_0(\gamma - Cn^{-c})$$

and

$$\bar{c}_0(\gamma + 1/n) - Cn^{-c} \geq \bar{c}_0(\gamma + Cn^{-c})$$

for all $\gamma \in (Cn^{-c}, 1 - Cn^{-c})$. Combining these inequalities with (16) gives the asserted claim. Q.E.D.

C Proofs for Section 2

Proof of Theorem 3.1. In this proof, c and C are understood as sufficiently small and large constants, respectively, whose values may change at each appearance. The constants c and C can be chosen to depend only on $c_F, C_F, c_X, C_X, c_\epsilon, C_\epsilon, u, \delta, \epsilon$, and the kernel function K. Also, recall the definitions of the set $\mathcal{S}_n$ and the critical values $c_0(\gamma)$ and $\bar{c}_0(\gamma)$ appearing in the beginning of Appendix B.

To prove the theorem, suppose that H_0 holds. Then

$$\begin{aligned}
\mathrm{P}(T > c(\alpha)) &= \mathrm{P}\left(\max_{(x,y,h)\in\mathcal{X}_n\times\mathcal{Y}_n\times\mathcal{B}_n} \frac{\sum_{i=1}^n k_{i,h}(x)1\{Y_i \leq y\}}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}} > c(\alpha)\right) \\
&\leq \mathrm{P}\left(\max_{(x,y,h)\in\mathcal{X}_n\times\mathcal{Y}_n\times\mathcal{B}_n} \frac{\sum_{i=1}^n k_{i,h}(x)(1\{Y_i \leq y\} - F_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}} > c(\alpha)\right) \\
&\leq \mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n} \frac{\sum_{i=1}^n k_{i,h}(x)(1\{Y_i \leq y\} - F_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}} > c(\alpha)\right) + Cn^{-c}
\end{aligned} \tag{17}$$

by Lemmas B.2 and B.4. Next,

$$\begin{aligned}
&\mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n}\frac{\sum_{i=1}^n k_{i,h}(x)(1\{Y_i\le y\}-F_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}}>c(\alpha)\right)\\
&\qquad\le \mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n}\frac{\sum_{i=1}^n k_{i,h}(x)(1\{Y_i\le y\}-F_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}}>c_0(\alpha)\right)+Cn^{-c}\\
&\qquad\le \mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n}\frac{\sum_{i=1}^n k_{i,h}(x)(1\{Y_i\le y\}-F_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}}>\bar c_0(\alpha+Cn^{-c})\right)+Cn^{-c}
\end{aligned}$$

by Lemmas B.4 and B.5. Thus, $\mathrm{P}(T>c(\alpha))$ is bounded from above by

$$\mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n}\frac{\sum_{i=1}^n k_{i,h}(x)(1\{Y_i\le y\}-F_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}}>\bar c_0(\alpha+Cn^{-c})\right)+Cn^{-c}.$$

Also, if all functions $x\mapsto F_{Y|X}(y|x)$ are constant, then (17) holds with equality instead of inequality, and so by similar arguments, $\mathrm{P}(T>c(\alpha))$ is bounded from below by

$$\mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n}\frac{\sum_{i=1}^n k_{i,h}(x)(1\{Y_i\le y\}-F_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}}>\bar c_0(\alpha-Cn^{-c})\right)-Cn^{-c}.$$

Therefore, to prove the theorem, it remains to show that for all $\gamma\in(0,1)$,

$$\left|\mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n}\frac{\sum_{i=1}^n k_{i,h}(x)(1\{Y_i\le y\}-F_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}}>\bar c_0(\gamma)\right)-\gamma\right|\le Cn^{-c}. \tag{18}$$

To do so, let $Z=(Z_{x,y,h})_{(x,y,h)\in\mathcal{S}_n}$ be a zero-mean Gaussian random vector such that

$$\begin{aligned}
&\mathrm{E}[Z_{x_1,y_1,h_1}Z_{x_2,y_2,h_2}]\\
&\quad=\mathrm{E}\left[\frac{\sum_{i=1}^n k_{i,h_1}(x_1)k_{i,h_2}(x_2)(1\{Y_i\le y_1\}-F_{Y|X}(y_1|X_i))(1\{Y_i\le y_2\}-F_{Y|X}(y_2|X_i))}{(\sum_{i=1}^n k_{i,h_1}(x_1)^2\sum_{i=1}^n k_{i,h_2}(x_2)^2)^{1/2}}\right]
\end{aligned}$$

for all (x_1,y_1,h_1) and (x_2,y_2,h_2) in $\mathcal{S}_n$. Also, for $\gamma\in(0,1)$, let $c_0^G(\gamma)$ be the $(1-\gamma)$ quantile of $\max_{(x,y,h)\in\mathcal{S}_n}Z_{x,y,h}$. Further, note that by Lemma B.1 and the definition of the set $\mathcal{S}_n$, we have with probability at least $1-Cn^{-c}$ that for all $(x,y,h)\in\mathcal{S}_n$,

$$\sum_{i=1}^n\mathrm{E}\left[\frac{k_{i,h}(x)^2(1\{Y_i\le y\}-F_{Y|X}(y|X_i))^2}{\sum_{l=1}^n k_{l,h}(x)^2}\mid\mathcal{X}_n\right]\ge\frac{1}{\log^2 n}, \tag{19}$$

$$\sqrt{n}\sum_{i=1}^n\mathrm{E}\left[\frac{|k_{i,h}(x)^3(1\{Y_i\le y\}-F_{Y|X}(y|X_i))^3|}{(\sum_{l=1}^n k_{l,h}(x)^2)^{3/2}}\mid\mathcal{X}_n\right]\le 8B_n,$$

$$n\sum_{i=1}^n\mathrm{E}\left[\frac{k_{i,h}(x)^4(1\{Y_i\le y\}-F_{Y|X}(y|X_i))^4}{(\sum_{l=1}^n k_{l,h}(x)^2)^2}\mid\mathcal{X}_n\right]\le 16B_n^2,$$

$$\max_{1\le i\le n}\mathrm{E}\left[\exp\left(\frac{\sqrt{n}|k_{i,h}(x)(1\{Y_i\le y\}-F_{Y|X}(y|X_i))|}{(\sum_{l=1}^n k_{l,h}(x)^2)^{1/2}}\Big/(3B_n)\right)\mid\mathcal{X}_n\right]\le 2.$$

These inequalities imply that it follows from Proposition 4.1 in Chernozhukov, Chetverikov, and Kato (2017) applied conditional on $\mathcal{X}_n$ that with probability $1 - Cn^{-c}$, for all $\gamma \in (Cn^{-c}, 1 - Cn^{-c})$,

$$\bar{c}_0(\gamma + Cn^{-c}) \leq c_0^G(\gamma) \leq \bar{c}_0(\gamma - Cn^{-c}). \tag{20}$$

Also, it follows from Proposition 2.1 in Chernozhukov, Chetverikov, and Kato (2017) applied conditional on $\mathcal{X}_n$ that for all $\gamma \in (0,1)$,

$$\left| \mathrm{P}\left(\max_{(x,y,h)\in\mathcal{S}_n} \frac{\sum_{i=1}^n k_{i,h}(x)(1\{Y_i \leq y\} - F_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}} > c_0^G(\gamma)\right) - \gamma \right| \leq Cn^{-c}. \tag{21}$$

Combining (20) and (21) in turn gives (18) and completes the proof of the theorem (note that a direct application of Propositions 2.1 and 4.1 in Chernozhukov, Chetverikov, and Kato (2017) would require that the left-hand of (19) is bounded by c instead of $1/\log^2 n$ but this does not lead to a problem because we apply these propositions with

$$\frac{\log n(k_{i,h}(x)(1\{Y_i \leq y\} - F_{Y|X}(y|X_i)))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}}$$

instead of

$$\frac{k_{i,h}(x)(1\{Y_i \leq y\} - F_{Y|X}(y|X_i))}{(\sum_{i=1}^n k_{i,h}(x)^2)^{1/2}}$$

and use the fact that $B_n \log n/\sqrt{n} \leq Cn^{-c}$). Q.E.D.

Proof of Theorem 3.2. In this proof, C is a constant that depends on u and δ only but its value can change at each appearance.

To prove the asserted claim, note that since $(x,y) \mapsto F_{Y|X}(y|x)$ is continuously differentiable and there exist $x^* \in (0,1)$ and $y^* \in (0,1)$ such that

$$\frac{\partial}{\partial x} F_{Y|X}(y^*|x^*) > 0,$$

it follows that there exists $\bar{y}^* \in \mathcal{Y}_n$ such that

$$\frac{\partial}{\partial x} F_{Y|X}(\bar{y}^*|x^*) > 0.$$

For this $\bar{y}^*$,

$$T = \max_{(x,y,h)\in\mathcal{X}_n\times\mathcal{Y}_n\times\mathcal{B}_n} \frac{\sum_{i=1}^n k_{i,h}(x)1\{Y_i \leq y\}}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}} \geq \max_{(x,h)\in\mathcal{X}_n\times\mathcal{B}_n} \frac{\sum_{i=1}^n k_{i,h}(x)1\{Y_i \leq \bar{y}^*\}}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}} =: T_{\bar{y}^*}.$$

Further, conditional on the data, the random variables

$$T^b(x,y,h) := \frac{\sum_{i=1}^n e_i k_{i,h}(x)(1\{Y_i \leq y\} - \widehat{F}_{Y|X}(y|X_i))}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}}, \quad (x,y,h) \in \mathcal{X}_n \times \mathcal{Y}_n \times \mathcal{B}_n,$$

are zero-mean Gaussian with variance bounded from above by

$$\frac{\sum_{i=1}^n\left(k_{i,h}(x)(1\{Y_i\le y\}-\widehat{F}_{Y|X}(y|X_i))\right)^2}{\sum_{i=1}^n k_{i,h}(x)^2}$$
$$\le \max_{y\in\mathcal{Y}_n}\max_{1\le i\le n}\left(1\{Y_i\le y\}-\widehat{F}_{Y|X}(y|X_i)\right)^2\le C$$

uniformly over $(x,y,h)\in\mathcal{X}_n\times\mathcal{Y}_n\times\mathcal{B}_n$ with probability at least $1-C_Fn^{-c_F}$ by Assumption 3.3. Therefore, with the same probability, $c(\alpha)\le C(\log n)^{1/2}$ since $c(\alpha)$ is the $(1-\alpha)$ conditional quantile of T^b given the data, $T^b=\max_{(x,y,h)\in\mathcal{X}_n\times\mathcal{Y}_n\times\mathcal{B}_n}T^b(x,y,h)$, and $p:=|\mathcal{X}_n\times\mathcal{Y}_n\times\mathcal{B}_n|$, the number of elements of the set $\mathcal{X}_n\times\mathcal{Y}_n\times\mathcal{B}_n$, satisfies $\log p\le C\log n$. Thus, the growth rate of the critical value $c(\alpha)$ satisfies the same upper bound, with a possibly different constant C, as if we were testing monotonicity of only one regression function, $x\mapsto \mathrm{E}[1\{Y\le\bar{y}^*\}|X=x]$, and using the bootstrap test statistic

$$T^b_{\bar{y}^*}=\max_{(x,h)\in\mathcal{X}_n\times\mathcal{B}_n}T^b(x,\bar{y}^*,h)$$

to simulate the critical value. Hence, the asserted claim follows from the same arguments as those given in the proof of Theorems 3.2 and 4.2 in Chetverikov (2012), which show that

$$\mathrm{P}\left(T_{\bar{y}^*}\le C\sqrt{\log n}\right)\to 0 \text{ as } n\to\infty$$

for any constant $C>0$. This completes the proof of the theorem. Q.E.D.

Proof of Theorem 3.3. Note that $(\log n/n)^{\beta/(2\beta+3)}=o(\ell_n)$ implies that $n\ell_n\to\infty$ as $n\to\infty$. Hence, given that

$$\left|\frac{\partial^2}{\partial x\partial y}F_{Y|X,n}(y|x)\right|\le C_L,\quad \text{for all } x,y\in[0,1],$$

and that there exist $x^*,y^*\in(0,1)$ such that

$$\frac{\partial}{\partial x}F_{Y|X,n}(y^*|x^*)>\ell_n,$$

it follows that there exists $\bar{y}^*\in\mathcal{Y}_n$ such that

$$\frac{\partial}{\partial x}F_{Y|X,n}(\bar{y}^*|x^*)>\widetilde{\ell}_n:=\ell_n/2.$$

Like in the proof of Theorem 3.2, for this $\bar{y}^*$, we have

$$T\ge T_{\bar{y}^*}:=\max_{(x,h)\in\mathcal{X}_n\times\mathcal{B}_n}\frac{\sum_{i=1}^n k_{i,h}(x)1\{Y_i\le\bar{y}^*\}}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}},$$

and, in addition, we have that $c(\alpha) \leq C(\log n)^{1/2}$ with probability at least $1 - C_F n^{-c_F}$, where C is a constant that depends only on δ and u. Therefore, since $\widetilde{\ell}_n$ satisfies $(\log n/n)^{\beta/(2\beta+3)} = o(\widetilde{\ell}_n)$, the asserted claim now follows from the same arguments as those given in the proof of Theorems 3.4 and 4.2 in Chetverikov (2012). Here, we note that our choice of $h_{\min} = 1/n^{1-\delta}$ with $\delta \in (0, 2/3]$ ensures that the collection of bandwidth values $\mathcal{B}_n$ contains at least one bandwidth value of order $(\log n/n)^{1/(2\beta+3)}$, which is required in the arguments of Chetverikov (2012). Q.E.D.

D Technical Lemmas

Lemma D.1. *Let $X_1, \dots, X_n$ be independent random vectors in $\mathbb{R}^p$ with $p \geq 2$ such that $X_{ij} \geq 0$ for all $i = 1, \dots, n$ and $j = 1, \dots, p$. Define*

$$M := \max_{1 \leq i \leq n} \max_{1 \leq j \leq p} X_{ij}.$$

Then for any $s \geq 1$ and $t > 0$,

$$\mathrm{P}\left(\max_{1 \leq j \leq p} \sum_{i=1}^{n} X_{ij} \geq 2\mathrm{E}\Big[\max_{1 \leq j \leq p} \sum_{i=1}^{n} X_{ij} \Big] + t \right) \leq K_1 \mathrm{E}[M^s]/t^s. \tag{22}$$

where K_1 is a constant depending only on s. In addition,

$$\mathrm{E}\left[\max_{1 \leq j \leq p} \sum_{i=1}^{n} X_{ij} \right] \leq K_2 \left(\max_{1 \leq j \leq p} \mathrm{E}\left[\sum_{i=1}^{n} X_{ij} \right] + \mathrm{E}[M] \log p \right), \tag{23}$$

where K_2 is a universal constant.

Proof. See Lemma E.4 in Chernozhukov, Chetverikov, and Kato (2017) and Lemma 9 in Chernozhukov, Chetverikov, and Kato (2015) for the proof of (22) and (23), respectively. Q.E.D.

Lemma D.2. *Let $Z = (Z_1, \dots, Z_p)'$ be a zero-mean Gaussian random vector in $\mathbb{R}^p$ with $\sigma_j^2 := \mathrm{E}[Z_j^2] > 0$ for all $j = 1, \dots, p$. Denote $\underline{\sigma} := \min_{1 \leq j \leq p} \sigma_j$. Then for all $\epsilon > 0$ and $x = (x_1, \dots, x_p)' \in \mathbb{R}^p$, we have*

$$\mathrm{P}(Z \leq x + \epsilon) - \mathrm{P}(Z \leq x) \leq \frac{\epsilon}{\underline{\sigma}}(\sqrt{2 \log p} + 2), \tag{24}$$

where $x + \epsilon = (x_1 + \epsilon, \dots, x_p + \epsilon)'$.

Proof. See Lemma A.1 in Chernozhukov, Chetverikov, and Kato (2017). Q.E.D.

References

Banerjee, M., D. Mukherjee, and S. Mishra (2009): "Semiparametric binary regression models under shape constraints with an application to Indian schooling data," *Journal of Econometrics*, 149(2), 101 – 117.

Barrett, G. F., and S. G. Donald (2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, 71(1), 71–104.

Boudoukh, J., M. Richardson, T. Smith, and R. F. Whitelaw (1999): "Ex Ante Bond Returns and the Liquidity Preference Hypothesis," *The Journal of Finance*, 54(3), 1153–1167.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013): "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors," *The Annals of Statistics*, 41(6), 2786–2819.

——— (2015): "Comparison and anti-concentration bounds for maxima of Gaussian random vectors," *Probability Theory and Related Fields*, 162.

——— (2017): "Central Limit Theorems and Bootstrap in High Dimensions," *The Annals of Probability*, 45.

Chetverikov, D. (2012): "Testing Regression Monotonicity in Econometric Models," Discussion paper.

Chetverikov, D., and D. Wilhelm (2017): "Nonparametric Instrumental Variable Estimation Under Monotonicity," *Econometrica*, 85(4), 1303–1320.

Delgado, M. A., and J. C. Escanciano (2012): "Distribution-free tests of stochastic monotonicity," *Journal of Econometrics*, 170(1), 68–75.

——— (2013): "Conditional Stochastic Dominance Testing," *Journal of Business & Economic Statistics*, 31(1), 16–28.

Dümbgen, L., and V. Spokoiny (2001): "Multiscale Testing of Qualitative Hypotheses," *Annals of Statistics*, 29, 124–152.

Ellison, G., and S. F. Ellison (2011): "Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration," *American Economic Journal: Microeconomics*, 3(1), 1–36.

ERICSON, R., AND A. PAKES (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *The Review of Economic Studies*, 62(1), 53–82.

GHOSAL, S., A. SEN, AND A. W. V. D. VAART (2000): "Testing Monotonicity of Regression," *The Annals of Statistics*, 28(4), 1054–1082.

HÄRDLE, W., P. JANSSEN, AND R. SERFLING (1988): "Strong Uniform Consistency Rates for Estimators of Conditional Functionals," *The Annals of Statistics*, 16(4), 1428–1449.

HODERLEIN, S., H. HOLZMANN, M. KASY, AND A. MEISTER (2016): "Erratum Instrumental Variables with Unrestricted Heterogeneity and Continuous Treatment," *The Review of Economic Studies*, forthcoming.

HSU, Y.-C., C.-A. LIU, AND X. SHI (2016): "Testing Generalized Regression Monotonicity," Discussion paper.

KASY, M. (2014): "Instrumental Variables with Unrestricted Heterogeneity and Continuous Treatment," *The Review of Economic Studies*, 81(4), 1614.

LEE, S., O. LINTON, AND Y.-J. WHANG (2009): "Testing for Stochastic Monotonicity," *Econometrica*, 77(2), 585–602.

LEE, S., K. SONG, AND Y.-J. WHANG (2013): "Testing functional inequalities," *Journal of Econometrics*, 172(1), 14 – 32.

LEWBEL, A., AND O. LINTON (2007): "Nonparametric Matching and Efficient Estimators of Homothetically Separable Functions," *Econometrica*, 75(4), 1209–1227.

MATZKIN, R. L. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, pp. 2523–2558. Elsevier Science B.V.

OLLEY, G. S., AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297.

PATTON, A. J., AND A. TIMMERMANN (2010): "Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts," *Journal of Financial Economics*, 98(3), 605 – 625.

PITERBARG, V. (1996): *Asymptotic Methods in the Theory of Gaussian Processes and Fields*. American Mathematical Society, Providence, RI.

Richardson, M., P. Richardson, and T. Smith (1992): "The monotonicity of the term premium: Another look," *Journal of Financial Economics*, 31(1), 97 – 105.

Seo, J. (2016): "Tests of Stochastic Monotonicity with Improved Size and Power Properties," Discussion paper.

Stokey, N. L., and R. E. Lucas Jr. (1989): *Recursive Methods in Economic Dynamics*. Harvard University Press.

Wilhelm, D. (2017): "Testing for the Presence of Measurement Error," Discussion paper.