Econometric Theory, **37**, 2021, 495–536.
doi:10.1017/S0266466620000225

AN ADAPTIVE TEST OF STOCHASTIC MONOTONICITY*

DENIS CHETVERIKOV
University of California at Los Angeles
DANIEL WILHELM
University College London
DONGWOO KIM
Simon Fraser University

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 nonconservative, and detects certain smooth 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, under some alternatives, is significantly more powerful than existing procedures.

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 et al., 2016; Chetverikov and Wilhelm, 2017; Wilhelm, 2019) 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

*We thank Ivan Canay and Whitney Newey for useful comments. Daniel Wilhelm 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). Address correspondence to Daniel Wilhelm, Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom; e-mail: d.wilhelm@ucl.ac.uk.

[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.

© Cambridge University Press 2020

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

and Pakes, 1996), and in finance (e.g., Richardson, Richardson, and Smith, 1992; Boudoukh et al., 1999; Patton and Timmermann, 2010), among many other fields of economics.

Consider two continuous random variables X and Y, both supported on $[0,1]$. 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

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

where $F_{Y|X}$ denotes the cdf of Y given X, that is, $F_{Y|X}(y|x) := \mathrm{P}(Y \leq y \mid X = x)$ for all $x, y \in [0,1]$.

We propose a new nonparametric test of (1) against (2) with attractive properties. First, our test controls asymptotic size and is nonconservative, that is, its limiting rejection probability does not exceed the nominal level uniformly over all data-generating processes in the null (satisfying mild regularity conditions) and is equal to the nominal level for some of them. In fact, we show that the probability of rejecting the null under the null can exceed the nominal level at most by a term that converges to zero with a polynomial rate, which we refer to as the polynomial size control. Second, our test is consistent against all smooth fixed alternatives. Third, our test is rate-optimal against sequences of alternatives in smoothness classes $\mathcal{M}_\beta$, where $\beta \in (0,1]$ denotes the smoothness parameter, and $\mathcal{M}_\beta$ is the set of all distributions of the pair (X,Y) satisfying some mild regularity conditions and such that

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

for some given constant C_L; see Section 3 for details on the definition of $\mathcal{M}_\beta$. Here, rate-optimality against sequences of alternatives in $\mathcal{M}_\beta$ means that there exist a constant $\kappa \in (0,1)$ and a sequence of positive constants $\{r_n\}_{n \geq 1}$ converging to zero such that our test is uniformly consistent against all sequences of alternatives in $\mathcal{M}_\beta$ separated from the null at least by r_n (in a certain metric) and for any test whose asymptotic size does not exceed its nominal level, there exists a sequence of alternatives in $\mathcal{M}_\beta$ separated from the null at least by κr_n such that the asymptotic power of the test against this sequence of alternatives also does not exceed its nominal level. Fourth, our test is adaptive to the set of classes $\{\mathcal{M}_\beta\}_{\beta \in (0,1]}$ meaning that the test is rate-optimal against $\mathcal{M}_\beta$ for each $\beta \in (0,1]$ and implementing the test does not require specifying the value of β.

For comparison, the implementation of nonadaptive tests often requires the user to specify a value of some smoothing parameter, such as a bandwidth value, which is undesirable because the test results may be sensitive to the particular value that is chosen. Moreover, nonadaptive tests may have low power if the smoothing parameter 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

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

values, the resulting procedure may not control size, even in large samples. To the best of our knowledge, our test is the first test of stochastic monotonicity that is shown to be adaptive.

Our test is also 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. 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 a locally weighted version of Kendall's tau and is thus most closely related to the one proposed in Lee, Linton, and Whang (2009). In both cases, the weights are determined by some kernel function K and a bandwidth value h but an important difference is that we take the maximum over many different values of h whereas they let the user specify a particular value. This gives us an advantage in terms of power. In particular, like ours, their test is also rate-optimal against sequences of alternatives in the smoothness classes $\mathcal{M}_\beta$ but achieving rate-optimality using their test requires providing a β-dependent bandwidth value h, and so their test cannot be adaptive. Thus, our test can be viewed as an adaptive version of theirs. Our critical value is also different since it has to take into account the fact that we perform a search over different bandwidth values.

Delgado and Escanciano (2012) and Seo (2018) construct a test statistic by comparing the empirical copula of (X, Y) with its partial least concave majorant. We show that these tests are not rate-optimal against sequences of alternatives in the smoothness classes $\mathcal{M}_\beta$. A practical consequence of this is that the tests have low finite-sample power against some smooth alternatives. The tests are, however, consistent against certain sequences of $n^{-1/2}$-alternatives, which we call "wide" alternatives because they deviate from the null over an interval of conditioning values of X. Our test is not consistent against such wide alternatives. Similar trade-offs have been noted in other contexts (e.g., Fan, 1996; Fan and Li, 2000; Horowitz and Spokoiny, 2001).

Hsu, Liu, and Shi (2019) and Lee, Song, and Whang (2018) propose tests of functional inequalities of which testing the null of stochastic monotonicity is a special case.

Also, stochastic monotonicity implies the weaker concept of regression monotonicity, that is, monotonicity of the function $x \mapsto \mathrm{E}[Y \mid X = x]$, and there are several papers in the literature that develop tests of regression monotonicity, for example, Ghosal, Sen, and Vaart (2000), Delgado and Escanciano (2013), Lee et al. (2013), and Chetverikov (2012). Our method for testing stochastic monotonicity is most closely related to the method of testing regression monotonicity in Chetverikov (2012), but there are several important differences between that and this paper. First, his method can be used to test the hypothesis that the function $x \mapsto \mathrm{E}[1\{Y \le y\} \mid X = x]$ is decreasing for any given $y \in (0, 1)$ but we test a different, stronger hypothesis that $x \mapsto \mathrm{E}[1\{Y \le y\} \mid X = x]$ is decreasing simultaneously

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

for all $y \in (0,1)$. Second, we use a different normalization factor making our test substantially easier to implement. Third, we have to deal with values of y that are close to the boundary of the support of Y, and doing so requires us to develop some new results on distributional approximations for the maxima of sums of high-dimensional random vectors. In particular, we extend the results in Chernozhukov, Lee, and Rosen (2013) and Chernozhukov, Chetverikov, and Kato (2017) by relaxing their condition that the variance of each component of the random vectors is bounded away from zero. Instead, we only require that the variance of at least one component is bounded away from zero. This extension may be of substantial independent interest.

Finally, we emphasize that our paper is not the first to take the maximum over many different bandwidth values in the test statistic to improve power properties of a test. For example, the same idea appeared in Horowitz and Spokoiny (2001), who were concerned with testing parametric regression models against general nonparametric alternatives, Dümbgen and Spokoiny (2001), who were concerned with testing regression function shape restrictions, and Armstrong (2015), who was concerned with testing hypotheses about a nonparametric regression function at a point. There are also many papers on testing moment inequalities that use the same idea, including Andrews and Shi (2013), Armstrong (2014), Armstrong and Chan (2016), and Chetverikov (2018).

The rest of the paper is organized as follows. In Section 2, we describe our new test. In Section 3, we present our main results on the properties of the test. In Section 4, we discuss new results on the distributional approximations for the maxima of sums of high-dimensional random vectors, which are used to establish the results in Section 3. In Section 5, we carry out a small simulation study to shed some light on the performance of our test in finite samples and also to compare its power with that of other tests in the literature. In the Appendix, we provide all the proofs.

2. THE TEST

In this section, we introduce our new test of stochastic monotonicity based on an i.i.d. sample (X_i, Y_i), $i = 1, \dots, n$, from the distribution of the pair (X, Y). Throughout the paper, we assume that the random variables X and Y are normalized to have support $[0,1]$.[2]

Let $K\colon \mathbb{R} \to \mathbb{R}$ be a continuous (kernel) function with support $[-1,1]$ such that $K(x) > 0$ for all $x \in (-1,1)$. For all (bandwidth) values $h > 0$, define

$$K_h(x) := h^{-1}K(x/h), \quad \text{for all } x \in \mathbb{R}.$$

[2]Subsequently, it will become clear that our test is invariant to strictly increasing transformations of Y, so assuming its support to be $[0,1]$ is a normalization. Also, the subsequent results can easily be adapted to the case in which X has support $[\underline{x}, \overline{x}]$ for some finite constants $\underline{x} < \overline{x}$. For simplicity of the presentation, however, we keep the assumption of the support of both variables being $[0,1]$.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

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

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

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

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

or, equivalently,

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

where

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

Our test is based on the observation that, under smoothness of $F_{Y|X}(y|\cdot)$ for all $y \in [0,1]$, the null H_0 is equivalent to (4) holding for all $x, y \in [0,1]$ and all $h \in (0,1)$. To define the test statistic T, let $h_{\max} := 1$, $h_{\min} := 1/\sqrt{n}$ and

$$\mathcal{H} := \left\{h = h_{\max} u^l : h \ge h_{\min} \text{ and } l = 0, 1, 2, \dots\right\}, \quad \text{for some } u \in (0,1),$$

be a collection of bandwidth values. Here, $\mathcal{H}$ forms a geometric grid on the interval $[h_{\min}, h_{\max}]$ with geometric step u and expands as n gets large. Also, let

$$\mathcal{X} := \{X_1, \dots, X_n\} \quad \text{and} \quad \mathcal{Y} := \{Y_1, \dots, Y_n\}. \tag{5}$$

We define our test statistic as

$$T := \max_{(x,y,h) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{H}} \frac{\sum_{i=1}^{n} k_{i,h}(x) 1\{Y_i \le y\}}{\left(\sum_{i=1}^{n} k_{i,h}(x)^2\right)^{1/2}}. \tag{6}$$

The statistic T is most closely related to that in Lee et al. (2009). The main difference is that we take the maximum over bandwidth values $h \in \mathcal{H}$ 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 with asymptotic level $\alpha \in (0, 1/2)$. As demonstrated by Lee et al. (2009), the derivation of the asymptotic distribution of T is complicated even when $\mathcal{H}$ is a singleton. Moreover, when $\mathcal{H}$ is not a singleton, it is generally unknown whether T converges to some nondegenerate asymptotic distribution,

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

even after an appropriate normalization. We avoid these complications by employing a multiplier bootstrap critical value. Specifically, for $x, y \in [0,1]$, let

$$\widehat{F}_{Y|X}(y|x) := \frac{\sum_{i=1}^n 1\{Y_i \le y\} K_b(X_i - x)}{\sum_{i=1}^n K_b(X_i - x)} \tag{7}$$

be an estimator of $F_{Y|X}(y|x)$, where we set $b := n^{-2/3}$ (other specifications, e.g., $b = n^{-1/2}$ are also possible). Also, let $e_1, \dots, e_n$ be an i.i.d. sequence of $N(0,1)$ random variables that are independent of the data. We then define a bootstrap test statistic as

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

and the critical value[3] $c(\alpha)$ as

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

We reject H_0 if and only if $T > c(\alpha)$. Since Y_i enters the test statistic and the critical values only through the indicators $1\{Y_i \le y\}$ and through the set $\mathcal{Y}$, and since the maximum is taken over $y \in \mathcal{Y}$, the test is invariant to strictly monotone transformations of Y_i.

We emphasize how simple and computationally straightforward the implementation of this test is. In particular, the bootstrap test statistic requires computing the nonparametric estimator $\widehat{F}_{Y|X}$ but this has to be done only once, using the original sample, and should not be repeated on each bootstrap sample. Also, when n is large, one can obtain nearly identical results by using coarser grids instead of $\mathcal{X}$ and $\mathcal{Y}$ both in the test statistic and in the bootstrap test statistic to decrease the number of elements over which the maxima are taken in order to reduce the computational burden.

To conclude this section, we note that implementing our test requires the choice of only two parameters: the kernel function K and the size of the geometric step u in the collection of bandwidth values u. For our simulations in Section 5, we used the Epanechnikov kernel function K and we set $u = 2/3$. This combination of parameters seems to work well in practice. Also, our test is robust with respect to the choice of the bandwidth value b. 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 $\widehat{F}_{Y|X}$ with the fastest possible rate of convergence (in some norm), there is no need to do so, and simply setting $b = n^{-2/3}$ gives similar results.[4] Finally, we note that it may be possible to use other estimators $\widehat{F}_{Y|X}$ of $F_{Y|X}$.

[3] 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 those in Chetverikov (2012), we could also consider two-step or even multistep (stepdown) critical values. For brevity of the paper, however, we do not consider these options here.

[4] Of course, if the support of X is $[\underline{x}, \bar{x}]$ for some constants $\underline{x} < \bar{x}$ rather than $[0,1]$, an appropriate bandwidth value would be $b = (\bar{x} - \underline{x})n^{-2/3}$. In this case, we would also have to use $h_{\max} = \bar{x} - \underline{x}$ and $h_{\min} = (\bar{x} - \underline{x})/\sqrt{n}$.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

3. LARGE SAMPLE PROPERTIES OF THE TEST

In this section, we derive size and power properties of the test proposed in Section 2. First, we show that the test controls asymptotic size and is nonconservative. We also show that the probability of rejecting the null under the null can exceed the nominal level α at most by a term that converges to zero with a polynomial rate. Then we demonstrate that the test is consistent against all fixed smooth alternatives and study the rate of consistency against two types of local alternatives. In particular, we show that our test is rate-optimal and adaptive against sequences of alternatives in the smoothness classes $\mathcal{M}_\beta$, $\beta \in (0,1]$, to be formally defined below. The proofs for this section rely on new results on the distributional approximations for the maxima of sums of high-dimensional random vectors. Since these may be of independent interest, we present them in more detail in Section 4.

We start our analysis by providing the list of required regularity conditions. Let $C_X, C_L \in [1,\infty)$, $\epsilon \in (0,1/2)$, and $\beta \in (0,1]$ be some constants.

Assumption 3.1 (Distribution of X). The distribution of X is absolutely continuous with respect to the Lebesgue measure on $[0,1]$ with the pdf f_X satisfying $1/C_X \leq f_X(x) \leq C_X$ for all $x \in (0,1)$.

This is a weak regularity condition. It requires 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. Conditions of this form are often used in the nonparametric analysis and the theory can easily be adapted to the case in which the support of X is $[\underline{x},\bar{x}]$ for some finite constants $\underline{x} < \bar{x}$.

Assumption 3.2 (Nondegeneracy). The conditional cdf $F_{Y|X}$ satisfies $\epsilon \leq F_{Y|X}(y|x) \leq 1-\epsilon$ for some $x, y \in (0,1)$.

This is another weak regularity condition. It holds if there exists at least one $x \in (0,1)$ such that the conditional distribution of Y given $X = x$ has at least two points on its support. We note that it is trivial to test (1) against (2) if this condition does not hold.

Assumption 3.3 (Smoothness). The conditional cdf $F_{Y|X}$ is such that

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

This is our key smoothness condition. It requires the derivative of $x \mapsto F_{Y|X}(y|x)$ to be Hölder continuous in x with exponent $\beta \in (0,1]$ and constant $C_L \in [1,\infty)$ for all $y \in (0,1)$. Note that here we implicitly assume that the derivate $(\partial/\partial x) F_{Y|X}(y|x)$ exists for all $x, y \in (0,1)$.

Let $\mathcal{M}_\beta$ be the set of all distributions on $[0,1]^2$ such that if (X,Y) has a distribution from $\mathcal{M}_\beta$, then Assumptions 3.1, 3.2, and 3.3 are satisfied. Thus, $\mathcal{M}_\beta$ is the smoothness class appearing in Section 1. Also, let $\mathcal{M}_{\beta,0}$ denote the set of all

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

distributions in $\mathcal{M}_\beta$ such that if (X, Y) has a distribution from $\mathcal{M}_{\beta,0}$, then X and Y are independent. Thus, all distributions in $\mathcal{M}_{\beta,0}$ satisfy (1).

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. If H_0 holds, then*

$$\mathrm{P}(T > c(\alpha)) \leq \alpha + Cn^{-c}. \tag{8}$$

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

$$|\mathrm{P}(T > c(\alpha)) - \alpha| \leq Cn^{-c}. \tag{9}$$

In both (8) and (9), $c, C \in (0, \infty)$ are constants depending only on C_X, C_L, ϵ, β, u, and the kernel function K.

The result (8) implies that our test controls asymptotic size. The result (9) in turn strengthens this statement by showing that the rejection probability for some data-generating processes[5] 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 can exceed the nominal level α only by a term that is polynomially small in n. We refer to this phenomenon as *the polynomial size control*. As explained in Lee et al. (2009), when $\mathcal{H}$ is a singleton, convergence of T to the limit distribution is logarithmically slow. For this reason, Lee et al. (2009) used higher-order corrections derived in Piterbarg (1996) to obtain the polynomial size control. Theorem 3.1 shows that the multiplier bootstrap also leads to the polynomial size control, without requiring higher-order corrections.

The constants c and C in (8) and (9) depend on the data generating process only via constants appearing in Assumptions 3.1, 3.2, and 3.3, as well as the constant u and the kernel function K. Therefore, inequalities (8) and (9) hold uniformly over all data-generating processes satisfying these assumptions with the same constants, and so our test provides uniform control of the asymptotic size. In other words, we have

$$\sup_{M \in \mathcal{M}_\beta} \mathrm{P}_M(T > c(\alpha)) \leq \alpha + Cn^{-c} \quad \text{and} \quad \sup_{M \in \mathcal{M}_{\beta,0}} |\mathrm{P}_M(T > c(\alpha)) - \alpha| \leq Cn^{-c},$$

where P_M denotes the probability when (X_i, Y_i), $i = 1, \ldots, n$, is a random sample from the distribution $M \in \mathcal{M}_\beta$.

To prove Theorem 3.1, we apply the high-dimensional CLT and bootstrap results in Chernozhukov et al. (2013) and Chernozhukov, Chetverikov, and Kato (2017).

[5] Theorem 3.1 establishes (9) only for data-generating processes in which Y and X are independent. This result is somewhat weaker than the corresponding result in Seo (2018), who shows that her test possesses an asymptotic null rejection probability equal to nominal size for a larger set of data-generating processes.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

However, the application is not straightforward because we would like to apply these results conditional on $\{X_i\}_{i=1}^n$, but in this case, the variance of the random variables

$$\frac{\sum_{i=1}^n k_{i,h}(x)1\{Y_i \le y\}}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}}$$

appearing in the definition of the test statistic T in (6) gets arbitrarily close to zero for values of y in $\mathcal{Y}$ that are close to the boundary of the support of Y, and the results of Chernozhukov et al. (2013, 2017) require all random variables, over which the maximum is taken, to have variance bounded away from zero. One possible way to deal with this issue would be to truncate the values in $\mathcal{Y}$ in the test statistic T that are too close to either zero or one but that would require introducing additional tuning parameters, which is undesirable. Instead, we extend the results of Chernozhukov et al. (2013, 2017) by relaxing the condition that all random variables, over which the maximum is taken, should have variance bounded away from zero and requiring that only one of these random variables has variance bounded away from zero. We present the extension in Section 4 and use it to prove Theorem 3.1.

Next, we investigate power properties of our test. First, we show that our test is consistent against all fixed smooth alternatives:

THEOREM 3.2 (Consistency against fixed alternatives). *Let Assumption 3.1 be satisfied. In addition, assume that the functions* $x \mapsto F_{Y|X}(y|x)$ *are continuously differentiable for all* $y \in (0,1)$. *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

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

Second, we derive the rate of consistency against sequences of alternatives of two types: wide and smooth. We define the *sequence of wide alternatives* as any sequence $\{M_n\}_{n\ge1}$ of distributions of the pair (X,Y), indexed by the sample size n, such that if (X,Y) is distributed according to M_n, then its conditional cdf satisfies

$$\frac{\partial}{\partial x}F_{Y|X}(y|x) \ge r_n, \quad \text{for some } y \in (0,1) \text{ and all } x \in (x_{n,l}, x_{n,r}), \tag{11}$$

and, moreover, Assumption 3.1 holds, where $\{r_n\}_{n\ge1}$ is a sequence of strictly positive constants and $\{(x_{n,l}, x_{n,r})\}_{n\ge1}$ is a sequence of intervals in $[0,1]$ with asymptotically nonvanishing length, that is, satisfying $\liminf_{n\to\infty}(x_{n,r} - x_{n,l}) > 0$. In turn, the *sequence of smooth alternatives* is any sequence $\{M_n\}_{n\ge1}$ of distributions in $\mathcal{M}_\beta$ such that if (X,Y) is distributed according to M_n, then its conditional cdf $F_{Y|X}(y|x)$ satisfies

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

$$\frac{\partial}{\partial x} F_{Y|X}(y|x) \geq r_n, \quad \text{for some } x, y \in (0,1), \tag{12}$$

where, again, $\{r_n\}_{n \geq 1}$ is a sequence of strictly positive constants.

THEOREM 3.3 (Consistency against wide alternatives). *Let Assumption 3.1 be satisfied and let $x_l, x_r \in (0,1)$ be some constants such that $x_l < x_r$. There exist constants $c, C \in (0, \infty)$, not necessarily the same as those appearing in Theorem 3.1, depending only on $x_r - x_l$, C_X, u, and the kernel function K such that if*

$$\frac{\partial}{\partial x} F_{Y|X}(y|x) \geq \sqrt{\frac{C \log n}{n}}, \quad \textit{for some } y \in (0,1) \textit{ and all } x \in (x_l, x_r), \tag{13}$$

then

$$\mathrm{P}(T > c(\alpha)) \geq 1 - Cn^{-c}. \tag{14}$$

This theorem implies that our test is consistent against any sequence of wide alternatives if r_n in (11) satisfies $r_n \sqrt{n/\log n} \to \infty$ as $n \to \infty$. This rate is not optimal. Indeed, it can be shown that the tests in Delgado and Escanciano (2012) and Seo (2018), for example, are consistent against any sequence of wide alternatives if r_n in (11) satisfies $r_n \sqrt{n} \to \infty$. However, since the optimal rate cannot be faster than the parametric rate $1/\sqrt{n}$, meaning that no test can be uniformly consistent against sequences of wide alternatives if r_n in (11) satisfies $r_n \sqrt{n} \to 0$ as $n \to \infty$, it follows that our test is rate-optimal against sequences of wide alternatives up to the arguably small logarithmic factor $\sqrt{\log n}$.

THEOREM 3.4 (Consistency against smooth alternatives). *Let Assumptions 3.1 and 3.3 be satisfied. There exist constants $c, C \in (0, \infty)$, not necessarily the same as those appearing in previous theorems, depending only on C_X, C_L, β, u, and the kernel function K such that if*

$$\frac{\partial}{\partial x} F_{Y|X}(y|x) \geq \left(\frac{C \log n}{n} \right)^{\beta/(2\beta+3)}, \quad \textit{for some } x, y \in (0,1), \tag{15}$$

then

$$\mathrm{P}(T > c(\alpha)) \geq 1 - Cn^{-c}. \tag{16}$$

This theorem implies that our test is consistent against any sequence of smooth alternatives if r_n in (12) is given by $r_n = C(\log n/n)^{\beta/(2\beta+3)}$ for a certain constant C. The rate r_n appearing here is optimal in the sense that there exists a universal constant $c \in (0, \infty)$ such that for any test whose asymptotic size does not exceed its nominal level, there exists a sequence of smooth alternatives satisfying (12) with $r_n = c(\log n/n)^{\beta/(2\beta+3)}$ such that the asymptotic power of the test against this sequence of alternatives also does not exceed its nominal level. In other words, our test is rate-optimal against sequences of alternatives in the smoothness class $\mathcal{M}_\beta$.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

To show rate-optimality, we will rely upon the arguments similar to those in Dümbgen and Spokoiny (2001). We will need the following additional notation. Let $\varphi = \varphi(X_1, Y_1, \ldots, X_n, Y_n)$ be a test, that is, φ is a function of the data with values in $\{0, 1\}$. The test φ rejects the null if and only if $\varphi = 1$. Also, let $\mathrm{E}_M[\varphi]$ denote the rejection probability for the test φ when (X_i, Y_i), $i = 1, \ldots, n$, is a random sample from the distribution $M \in \mathcal{M}_\beta$. Finally, for all $x, y \in [0, 1]$ and $M \in \mathcal{M}_\beta$, let $F^M_{Y|X}(y|x) := \mathrm{P}_M(Y \leq y \mid X = x)$ where (X, Y) is distributed according to M.

THEOREM 3.5 (Lower bound). *For any $\gamma \in (0, 1)$ and any test φ satisfying*

$$\sup_{M \in \mathcal{M}_{\beta,0}} \mathrm{E}_M[\varphi] \leq \gamma,$$

there exists $M \in \mathcal{M}_\beta$ such that

$$\frac{\partial}{\partial x} F^M_{Y|X}(y|x) \geq c \left(\frac{\log n}{n} \right)^{\beta/(2\beta+3)}, \quad \text{for some } x, y \in (0, 1), \tag{17}$$

and

$$\mathrm{E}_M[\varphi] \leq \gamma + Cn^{-c}, \tag{18}$$

where $c, C \in (0, \infty)$ are universal constants, not necessarily the same as those appearing in previous theorems.

Applying Theorem 3.5 for each n with $\gamma = \gamma_n \to \alpha$ and using Theorem 3.4 implies that our test is indeed rate-optimal against sequences of smooth alternatives, that is, sequences of alternatives in the smoothness class $\mathcal{M}_\beta$. Moreover, since implementing our test does not require specifying β, it follows that our test is adaptive to the set of smoothness classes $\{\mathcal{M}_\beta\}_{\beta \in (0, 1]}$.

Remark 3.1 (Comparison with Lee et al., 2009 and Lee et al., 2018). As discussed in Section 1, our test can be regarded as an adaptive version of the test of Lee et al. (2009), with the main difference being that we maximize the test statistic over multiple bandwidth values h whereas they use only a single bandwidth. As a consequence, their test is also rate-optimal against sequences of alternatives in $\mathcal{M}_\beta$ but achieving this rate-optimality requires specifying a β-dependent bandwidth value h. In fact, if β is known, it is actually preferred to use their test because their test does not have to blow up the critical value to account for the search over multiple bandwidth values and so has better power. However, when β is unknown, which is typically the case in practice, we recommend using our test since it automatically provides a choice of the bandwidth value leading to good power. A similar comment applies to Lee et al. (2018) who also use a single bandwidth. □

Remark 3.2 (Comparison with Delgado and Escanciano, 2012 and Seo, 2018). The advantages of the tests developed in Delgado and Escanciano (2012) and Seo

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

(2018) are that they have better power relative to our test against sequences of wide alternatives, as discussed below Theorem 3.3, and that they are invariant to monotonic, continuous transformations of X. In addition, Delgado and Escanciano (2012)'s test does not involve tuning parameters. In turn, our test has better power against sequences of smooth alternatives. In particular, one can show that, unlike ours, the tests by Delgado and Escanciano (2012) and Seo (2018) are not rate-optimal against sequences of alternatives in $\mathcal{M}_\beta$. To do so, suppose that $X \sim U[0,1]$ and $Z \sim N(0,1)$ are independent random variables and that

$$Y = \Phi\left(Z - h^{1+\beta} g_0\left(\frac{X-h}{h}\right)\right),$$

where g_0 is the function defined in the proof of Theorem 3.5, $h \in (0,1)$ is some constant, and Φ is the cdf of the $N(0,1)$ distribution. Then

$$F_{Y|X}(y|x) = \Phi\left(\Phi^{-1}(y) + h^{1+\beta} g_0\left(\frac{x-h}{h}\right)\right), \quad \text{for all } x, y \in (0,1),$$

and, by the proof of Theorem 3.5, the distribution of the pair (X,Y) belongs to $\mathcal{M}_\beta$ and the conditional cdf $F_{Y|X}$ satisfies (12) with

$$r_n = (h/2)^\beta \phi(1), \tag{19}$$

where ϕ is the pdf of the $N(0,1)$ distribution. Moreover, the copula of the distribution of the pair (X,Y),

$$C(u,v) := \int_0^v F_{Y|X}(F_Y^{-1}(u)|x)dx, \quad \text{for all } u, v \in (0,1),$$

where F_Y denotes the cdf of Y, satisfies

$$C(u,2h) \le 2h\left(F_Y^{-1}(u) + h^{1+\beta}\right), \quad \text{for all } u \in (0,1).$$

Hence, denoting by $v \mapsto \widetilde{C}(u,v)$ the least concave majorant of the function $v \mapsto C(u,v)$, for all $u \in [0,1]$, it follows, after some simple algebra, that

$$\sup_{(u,v)\in[0,1]^2} |\widetilde{C}(u,v) - C(u,v)| \le 4h^2 \cdot h^{1+\beta} = 4h^{3+\beta}. \tag{20}$$

On the other hand, Thm. 4.3 in Seo (2018) shows that the tests of Delgado and Escanciano (2012) and Seo (2018) have trivial power against sequences of alternatives with copulas $C^{(n)}$ satisfying

$$\sqrt{n} \sup_{(u,v)\in[0,1]^2} |\widetilde{C}^{(n)}(u,v) - C^{(n)}(u,v)| \to 0 \quad \text{as } n \to \infty. \tag{21}$$

Combining (19), (20), and (21) and using $h = h_n = (C \log n/n)^{1/(2\beta+3)}$, for any constant $C \in (0,\infty)$, shows that the tests of Delgado and Escanciano (2012) and Seo (2018) are not rate-optimal against sequences of alternatives in $\mathcal{M}_\beta$. □

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Remark 3.3 (Comparison with Hsu et al., 2019). One can show that the test by Hsu et al. (2019) also has better power relative to our test against sequences of wide alternatives, as discussed below Theorem 3.3. On the other hand, since it is based on inference techniques by Andrews and Shi (2013), we conjecture that similar calculations as those in Appendix K of Chernozhukov et al. (2013) can be used to show that Hsu et al. (2019)'s test is not rate-optimal against sequences of alternatives in $\mathcal{M}_\beta$. A formal proof is out of scope for the present paper. □

Remark 3.4 (Multivariate X). A multivariate extension of our results to the case when X_i is a vector is possible by following arguments in Sect. 6 of Lee et al. (2009). □

Remark 3.5 (Higher-order stochastic monotonicity). It is straightforward to adapt our test and subsequent formal results for testing the null of higher-order stochastic monotonicity as in Ex. 2.5 of Hsu et al. (2019). □

Remark 3.6 (Time series data). It would be possible to extend our results to the case of dependent data following similar arguments as in Sect. 7 of Chernozhukov et al. (2014). □

4. MAXIMA OF SUMS OF HIGH-DIMENSIONAL RANDOM VECTORS WITHOUT NONVANISHING VARIANCE ASSUMPTION

In this section, we present new results on distributional approximations for the maxima of sums of high-dimensional random vectors, which are central to the derivations of the large sample properties in Section 3 and may be of independent interest.

To motivate the setup consider the following statistic:

$$T_0 := \max_{(x,y,h)\in\mathcal{X}\times\mathcal{Y}_d\times\mathcal{H}} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\sqrt{n}k_{i,h}(x)(1\{Y_i \le y\} - F_{Y|X}(y|X_i))}{\left(\sum_{l=1}^{n} k_{l,h}(x)^2\right)^{1/2}},$$

where $\mathcal{Y}_d \subset (0,1)$ is a certain set independent of the data (see Lemma A.1). This statistic is similar to our test statistic T in (6) except that $\mathcal{Y}$ is replaced by $\mathcal{Y}_d$ and the sum has been centered at mean zero. Now, index the elements over which we take the maximum by $j = 1,\dots,p$, where p denotes the number of elements in $\mathcal{X}\times\mathcal{Y}_d\times\mathcal{H}$, and $j \mapsto (x_j, y_j, h_j)$ is a one-to-one mapping from $\{1,\dots,p\}$ to $\mathcal{X}\times\mathcal{Y}_d\times\mathcal{H}$, so T_0 can be written as

$$T_0 = \max_{1\le j\le p} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\sqrt{n}k_{i,h_j}(x_j)(1\{Y_i \le y_j\} - F_{Y|X}(y_j|X_i))}{\left(\sum_{l=1}^{n} k_{l,h_j}(x_j)^2\right)^{1/2}}. \tag{22}$$

It is fairly easy to see that, under the null, $T \le T_0$ with probability at least $1 - n^{-1}$. Therefore, for the approximation of quantiles of T, we are interested in approximating the distribution of the statistic T_0 when p is allowed to be large, potentially much larger than the sample size n.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

More generally, in this section, we will therefore consider maxima of sums of high-dimensional random vectors of the form

$$S_n := \max_{1\le j\le p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ij},$$

where $Z_1, \ldots, Z_n$ are independent p-dimensional random vectors with mean zero, that is, $\mathrm{E}[Z_{ij}] = 0$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, p$, where Z_{ij} denotes the jth component of Z_i. A number of previous papers considered Gaussian and bootstrap approximations of the distribution of S_n under the condition that

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[Z_{ij}^2] \ge 1 \quad \text{for all } j = 1, \ldots, p; \tag{23}$$

see Belloni et al. (2018) for a review. However, these results are not appropriate for the purposes of this paper since we necessarily have

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[Z_{ij}^2] \text{ arbitrarily close to 0 for some } j = 1, \ldots, p$$

when Z_{ij} is the ith summand in (22). In this section, we therefore seek to relax (23) and obtain Gaussian and bootstrap approximations of the distribution of S_n with (23) replaced by

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[Z_{ij}^2] \ge 1 \quad \text{for some } j = 1, \ldots, p. \tag{24}$$

Clearly, (24) is substantially weaker than (23), and so our results in this section may be of independent interest.

Let $\{B_n\}_{n\ge 1}$ be a sequence of positive constants, possibly growing to infinity as n gets large. In addition to (24), we will impose the following conditions:

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[|Z_{ij}|^{2+k}] \le B_n^k \quad \text{for all } j = 1, \ldots, p \text{ and } k = 1, 2, \tag{25}$$

$$\mathrm{E}[\exp(|Z_{ij}|/B_n)] \le 2 \quad \text{for all } i = 1, \ldots, n \text{ and } j = 1, \ldots, p. \tag{26}$$

These two conditions are the same as those in the previous papers; see Belloni et al. (2018) for details. Also, let $Z_1^g, \ldots, Z_n^g$ be independent Gaussian p-dimensional random vectors with mean zero and covariance $\mathrm{E}[(Z_i^g)(Z_i^g)'] = \mathrm{E}[Z_i Z_i']$ for all $i = 1, \ldots, n$. Finally, let

$$S_n^g := \max_{1\le j\le p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ij}^g$$

be a Gaussian analog of S_n. Our first theorem in this section gives a Gaussian approximation of the distribution of S_n:

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

THEOREM 4.1 (Gaussian approximation). *Let (24), (25), and (26) be satisfied, and let $p \geq 3$. Then for any $x_0 > 0$,*

$$\sup_{x \geq x_0} \left| \mathrm{P}(S_n \leq x) - \mathrm{P}(S_n^g \leq x) \right| \leq \left(\frac{CB_n^2 \log^{10}(pn)}{n} \right)^{1/6},$$

where C is a constant depending only on x_0.

This theorem gives an approximation of the distribution of S_n by the distribution of S_n^g but this approximation is typically infeasible because implementing it would require knowing the matrix

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[(Z_i^g)(Z_i^g)'] = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[Z_i Z_i'],$$

which is unknown in most applications, including ours. We therefore consider a further approximation by the multiplier bootstrap. Specifically, let $e_1, \ldots, e_n$ be i.i.d. $N(0,1)$ random variables that are independent of $Z_1, \ldots, Z_n$, and let

$$S_n^e := \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i Z_{ij} \tag{27}$$

be a multiplier bootstrap analog of S_n. For example, in the context of our test of stochastic monotonicity, S_n^e corresponds to the bootstrap statistic

$$T_0^e := \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i \frac{\sqrt{n} k_{i,h_j}(x_j)(1\{Y_i \leq y_j\} - F_{Y|X}(y_j|X_i))}{\left(\sum_{l=1}^{n} k_{l,h_j}(x_j)^2 \right)^{1/2}}.$$

Our second theorem in this section gives a bootstrap approximation of the distribution of S_n^g, and hence of S_n:

THEOREM 4.2 (Bootstrap approximation). *Let (24), (25), and (26) be satisfied, and let $p \geq 3$. Then for any $x_0 > 0$,*

$$\sup_{x \geq x_0} \left| \mathrm{P}(S_n^g \leq x) - \mathrm{P}(S_n^e \leq x \mid (Z_i)_{i=1}^n) \right| \leq \left(\frac{CB_n^2 \log^{10}(pn)}{n} \right)^{1/6}, \tag{28}$$

with probability at least $1 - 4/(pn)$, where C is a constant depending only on x_0.

Theorems 4.1 and 4.2 taken together give the bootstrap approximation of the distribution of S_n. However, these two theorems do not immediately provide conditions ensuring that

$$|\mathrm{P}(S_n > c_n^e(\alpha)) - \alpha| \text{ is close to 0,}$$

where $c_n^e(\alpha)$ is the $(1-\alpha)$th quantile of the conditional distribution of S_n^e given $(Z_i)_{i=1}^n$. The reason is that $c_n^e(\alpha)$ is random and possibly correlated with S_n, but

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Theorems 4.1 and 4.2 only provide approximations of probabilities $\mathrm{P}(S_n > x)$ for nonrandom $x \in \mathbb{R}$. Theorem 4.3 below provides the additional step to go from a nonrandom value of x to $c_n^e(\alpha)$. In addition, since $F_{Y|X}$ is unknown and needs to be estimated, the bootstrap statistic T_0^e above is infeasible and we will construct a feasible bootstrap procedure with vectors Z_i replaced by corresponding estimators $\widehat{Z}_i$, $i = 1, \ldots, n$.

Let $\{\kappa_n\}_{n\geq 1}$ and $\{\zeta_n\}_{n\geq 1}$ be sequences of positive constants converging to zero as n gets large. Also, let $\widehat{Z}_1, \ldots, \widehat{Z}_n$ be estimators of $Z_1, \ldots, Z_n$ such that the following condition holds:

$$\mathrm{P}\left(\max_{1\leq j\leq p} \frac{1}{n}\sum_{i=1}^{n} (\widehat{Z}_{ij} - Z_{ij})^2 > \zeta_n^2\right) < \kappa_n. \tag{29}$$

Define

$$\widehat{S}_n^e := \max_{1\leq j\leq p} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i \widehat{Z}_{ij}, \tag{30}$$

where we assume that $e_1, \ldots, e_n$ are independent of $\widehat{Z}_1, \ldots, \widehat{Z}_n$, and let $c_n(\alpha)$ be the $(1-\alpha)$th quantile of $\widehat{S}_n^e$ given $(\widehat{Z}_i)_{i=1}^n$. For example, in the context of the derivations of the large sample results in Section 3, $\widehat{Z}_{ij}$ is similar to the ith summand in (22) except that $F_{Y|X}$ is replaced by an estimator.

Our third, and final, result in this section provides conditions under which $c_n(\alpha)$ can be used to approximate the $(1-\alpha)$th quantile of S_n:

THEOREM 4.3 (Bootstrap critical values). *Let (24), (25), (26), and (29) be satisfied, and let $p \geq 3$. Then for any $\alpha_0 \in (0, 1/2)$,*

$$\sup_{\alpha\in(0,\alpha_0)} \left|\mathrm{P}(S_n > c_n(\alpha)) - \alpha\right| \leq C\left(\left(\frac{B_n^2 \log^{10}(pn)}{n}\right)^{1/6} + \zeta_n \log^{3/2}(pn)\right) + \kappa_n, \tag{31}$$

where C is a constant depending only on α_0.

Remark 4.1 (Relation to the literature). Theorem 4.3 is related to Thm. 3.2 in Chernozhukov et al. (2013) and Thm. 4.3 in Chernozhukov et al. (2015). The main improvement of our result is that we only impose (24), which is substantially weaker than the condition (23) that the other two papers impose. On the other hand, the cost of this improvement is that the bound in (31) contains

$$\frac{B_n^2 \log^{10}(pn)}{n} \quad \text{instead of} \quad \frac{B_n^2 \log^{7}(pn)}{n},$$

appearing in Chernozhukov et al. (2015). □

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Remark 4.2 (Other types of bootstrap). Although we focus specifically on the multiplier bootstrap in this paper, results similar to those in Theorems 4.2 and 4.3 can be obtained for the nonparametric bootstrap as well, where we would replace the weights e_i in (27) and (30) by the weights $g_i - 1$, with the vector $g = (g_1, \ldots, g_n)'$ having a multinomial distribution with parameter n and probabilities $(1/n, \ldots, 1/n)$ and being independent of the vectors $Z_1, \ldots, Z_n, \widehat{Z}_1, \ldots, \widehat{Z}_n$. These results would follow from arguments similar to those used in the proofs of Theorems 4.2 and 4.3, with an application of Prop. 4.3 instead of Cor. 4.2 of Chernozhukov et al., 2017 at the end of the proof of Theorem 4.2. In turn, this would allow us to obtain results similar to those in Section 3 for the version of our test in Section 2 that uses the nonparametric bootstrap instead of the multiplier bootstrap. To the best of our knowledge, however, there is no theory that would allow us to say which type of the bootstrap provides better approximation of the test statistic T in Section 2. □

5. SIMULATIONS

In this section, we describe a simulation experiment which illustrates the finite sample performance of our test and compare it to other tests. The design is based on Delgado and Escanciano (2012). We simulate 1,000 Monte Carlo samples of sizes 100, 200, and 300 from the following six data generating processes:

N1: $Y = U$

N2: $Y = 0.1X + U$

A1: $Y = -0.1X + U$

A2: $Y = -0.1X^2 + U$

A3: $Y = -0.1 \exp(-250(X - 0.5)^2) + U$

A4: $Y = 0.2X - \beta \exp(-250(X - 0.5)^2) + U,$

where $\beta = 0.2$, X is uniformly distributed over $[0, 1]$, and U is drawn from the $N(0, 0.1^2)$ distribution, independently of X. Models $N1$ and $N2$ satisfy the null hypothesis, but models $A1$–$A4$ do not. In model $N1$, the random variables Y and X are independent, illustrating the case in which (1) is satisfied with equality for all x', x'', y. Model $N2$, on the other hand, satisfies (1) with strict inequality for all x', x'', y. Models $A1$ and $A2$ are models in the alternative hypothesis for which the null is violated at every pair of conditioning values x', x''. Models $A3$ and $A4$, on the other hand, are alternatives that deviate from the null only locally. Figure 1 shows the conditional mean functions $g(x) := \mathrm{E}[Y \mid X = x]$ for models $A3$ and $A4$. The null hypothesis is equivalent to the conditional mean function being nondecreasing.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

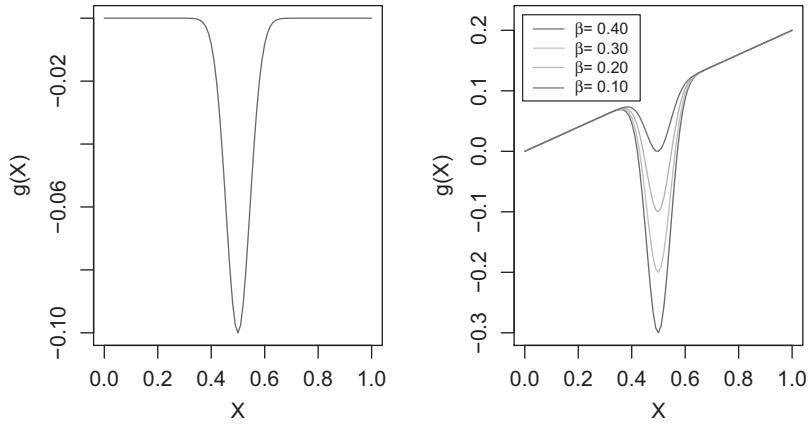


FIGURE 1. The conditional mean functions $g(X) := \mathrm{E}[Y \mid X]$ under A3 (left) and A4 (right).

For the implementation of our test, we choose the Epanechnikov kernel for K and construct the set of bandwidth values $\mathcal{H}$ using $u = 2/3$. The number of elements in $\mathcal{H}$ are 6, 7, and 7, 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 (7) with $b = n^{-2/3}$. 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. Seo (2018) tests are implemented with the refinement proposed in her Rem. 4.1 and with tuning parameters $\kappa_n = cn^{-1/2}\log\log n$, where $c = 0.23, 0.155, 0.22$ for her L^1-, L^2-, and L^∞-version, respectively. These parameters are chosen as suggested in footnote 12 of Seo (2018).

Table 1 shows the empirical rejection frequencies of various tests in each of the six 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 (2018)'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 et al. (2009) using the bandwidth values 0.5, 0.6, and 0.7, respectively.

All tests control size well in the two models N1 and N2 and all sample sizes, except that Seo's test over-rejects when $n = 300$. For model N2, all tests except ours have rejection frequencies close to zero rather than the nominal level.

All tests detect the alternatives A1 and A2 in the sense that their rejection frequencies are far above the nominal level of the test. CWK's and LLW's rejection frequencies are around 0.5 for the smallest sample size, but increase with the sample size to 0.9 or higher. Seo's tests are substantially more powerful against these alternatives with rejection frequencies close to one even for the smallest sample size.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Table 1. Rejection probabilities of our test ("CWK") and the tests of Seo (2018, "S"), Delgado and Escanciano (2012, "DE"), and Lee et al. (2009, "LLW")

Model	n	CWK	S-L^1	S-L^2	S-L^∞	DE	$LLW_{0.5}$	$LLW_{0.6}$	$LLW_{0.7}$
N1	100	0.054 (0.227)	0.057	0.064	0.063	0.046	0.034	0.035	0.036
	200	0.066 (0.171)	0.023	0.056	0.054	0.052	0.031	0.034	0.033
	300	0.062 (0.114)	0.088	0.105	0.168	0.042	0.036	0.039	0.039
N2	100	0.015 (0.156)	0.000	0.000	0.001	0.001	0.000	0.000	0.000
	200	0.018 (0.109)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	300	0.030 (0.076)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A1	100	0.471 (0.619)	1.000	0.998	0.799	0.634	0.408	0.542	0.612
	200	0.740 (0.758)	1.000	1.000	0.942	0.880	0.749	0.853	0.908
	300	0.890 (0.861)	1.000	1.000	1.000	0.980	0.911	0.964	0.980
A2	100	0.508 (0.604)	1.000	0.998	0.844	0.599	0.469	0.587	0.651
	200	0.770 (0.767)	1.000	1.000	0.983	0.906	0.805	0.892	0.925
	300	0.922 (0.862)	1.000	1.000	1.000	0.981	0.938	0.972	0.983
A3	100	0.236 (0.235)	0.065	0.106	0.172	0.135	0.064	0.067	0.052
	200	0.444 (0.208)	0.035	0.098	0.323	0.288	0.146	0.128	0.097
	300	0.602 (0.202)	0.786	0.822	0.798	0.447	0.273	0.218	0.188
A4	100	0.426 (0.162)	0.000	0.000	0.031	0.032	0.012	0.013	0.022
	200	0.791 (0.145)	0.000	0.001	0.087	0.157	0.014	0.009	0.014
	300	0.958 (0.143)	0.000	0.001	0.527	0.382	0.012	0.009	0.009

Note: The values in parentheses are the optimal bandwidths chosen by our test, averaged over the simulation samples.

The alternative A3 is substantially more difficult for the tests to detect in the sense that rejection frequencies are lower than for alternatives A1 and A2. LLW's rejection frequencies are only around 0.2 even for the largest sample size. CWK and Seo's rejection frequencies are fairly low for the small sample size, but increase to 0.6 and about 0.8 for the large sample size. The relative performance of the tests differs most starkly in model A4, in which CWK has high rejection frequencies from 0.4 to almost 1 depending on the sample size, but Seo's L^2- and L^2-tests and LLW have rejection frequencies close to zero for all sample sizes. Only the L^∞-test by Seo and DE have moderate power against A4 when the sample size is large.

To further investigate the power properties of our test, we generate 1,000 Monte Carlo samples of size $n = 100$ from $A4$ with different values of β. The right panel of Figure 1 shows how the magnitude of deviation from the null increases with β. Figure 2 shows the empirical rejection frequencies of our test, Seo's L^∞-test, and LLW with a range of different bandwidths as a function β. As β 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

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

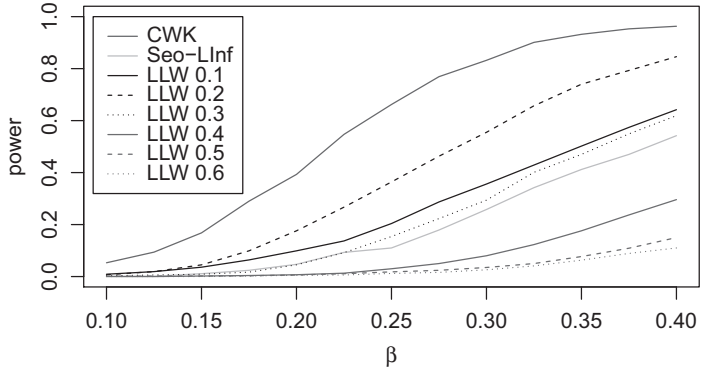


FIGURE 2. The power curves under various degrees of local deviation from the null.

bandwidth of 0.2, but it is significantly less powerful than ours for the bandwidth values considered here and for all values of β. Seo's test is also significantly less powerful for all values of β.

The simulation results illustrate the theoretical findings in Section 3. First, our test is less powerful than the other tests against alternatives for which the null is violated over a wide range of conditioning values x', x'' like A1 and A2. Second, our test is more powerful against smooth alternatives for which the deviations from the null occur only over a small range of conditioning values x', x'' like A3 and A4. Third, one can see that our test statistic is maximized at larger bandwidth values for alternatives A1 and A2, but at smaller bandwidth values for alternatives A3 and A4. This is a consequence of the adaptiveness of our test.

6. CONCLUSION

Monotone relationships between two variables play an important role in economics. Nonparametric tests of stochastic monotonicity are attractive because they allow the researcher to test for such monotone relationships without functional form assumptions on the joint distribution of the two variables. The disadvantage of nonparametric tests is that they require choices of tuning parameters such as bandwidths and whether or not the test rejects may be sensitive to the specific values that are chosen.

In this paper, we have proposed a new adaptive test that automatically chooses the bandwidth parameter so as to adapt to the unknown smoothness level of the conditional distribution of interest. The adaptiveness results in a test that is rate-optimal, simple to implement, and computationally attractive.

Among others, one extension of our results seems particularly desirable. In applied work, it is common to postulate stochastic monotonicity between two variables only after conditioning on a number of other control variables. If the number of such controls is moderately large, then the curse of dimensionality

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

prohibits the use of fully nonparametric estimators of the conditional distribution function as in (7). In this case, it might be desirable to use machine learning techniques that remain flexible yet practically feasible. Extending our results to allow for such estimators would therefore be useful.

References

Adler, R. & J. Taylor (2007) *Random Fields and Geometry*. Springer.

Andrews, D. & X. Shi (2013) Inference based on conditional moment inequalities. *Econometrica* 81, 609–666.

Armstrong, T. (2014) Weighted KS statistics for inference on conditional moment inequalities. *Journal of Econometrics* 181, 92–116.

Armstrong, T. (2015) Adaptive testing on a regression function at a point. *Annals of Statistics* 43, 2086–2101.

Armstrong, T. & H. Chan (2016) Multiscale adaptive inference on conditional moment inequalities. *Journal of Econometrics* 194, 24–43.

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

Belloni, A., V. Chernozhukov, D. Chetverikov, C. Hansen, & K. Kato (2018) High-Dimensional Econometrics and Regularized GMM. Discussion Paper.

Boudoukh, J., M. Richardson, T. Smith, & 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, & K. Kato (2013) Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics* 41(6), 2786–2819.

Chernozhukov, V., D. Chetverikov, & K. Kato (2014) Testing Many Moment Inequalities. Discussion Paper.

Chernozhukov, V., D. Chetverikov, & K. Kato (2015) Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields* 162, 47.

Chernozhukov, V., D. Chetverikov, & K. Kato (2017) Central limit theorems and bootstrap in high dimensions. *The Annals of Probability* 45, 1–44.

Chernozhukov, V., S. Lee, & A.M. Rosen (2013) Intersection bounds: Estimation and inference. *Econometrica* 81(2), 667–737.

Chetverikov, D. (2012) Testing Regression Monotonicity in Econometric Models. Discussion Paper.

Chetverikov, D. (2018) Adaptive tests of conditional moment inequalities. *Econometric Theory* 34, 186–227.

Chetverikov, D. & D. Wilhelm (2017) Nonparametric instrumental variable estimation under monotonicity. *Econometrica* 85(4), 1303–1320.

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

Delgado, M.A. & J.C. Escanciano (2013) Conditional stochastic dominance testing. *Journal of Business & Economic Statistics* 31(1), 16–28.

Dümbgen, L. & V. Spokoiny (2001) Multiscale testing of qualitative hypotheses. *Annals of Statistics* 29, 124–152.

Ellison, G. & 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. & A. Pakes (1995) Markov-perfect industry dynamics: A framework for empirical work. *The Review of Economic Studies* 62(1), 53–82.

Fan, J. (1996) Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association* 91(434), 674–688.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Fan, Y. & Q. Li (2000) Consistent model specification tests: Kernel-based tests versus Bierens' ICM tests. *Econometric Theory* 16(6), 1016–1041.

Ghosal, S., A. Sen, & A.W.v.d. Vaart (2000) Testing monotonicity of regression. *Annals of Statistics* 28(4), 1054–1082.

Hoderlein, S., H. Holzmann, M. Kasy, & A. Meister (2016) Erratum instrumental variables with unrestricted heterogeneity and continuous treatment. *The Review of Economic Studies* 84, rdw027.

Horowitz, J. & V. Spokoiny (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599–631.

Hsu, Y.-C., C.-A. Liu, & X. Shi (2019) Testing generalized regression monotonicity. *Econometric Theory* 35(6), 1146–1200.

Kasy, M. (2014) Instrumental variables with unrestricted heterogeneity and continuous treatment. *The Review of Economic Studies* 81(4), 1614.

Lee, S., O. Linton, & Y.-J. Whang (2009) Testing for stochastic monotonicity. *Econometrica* 77(2), 585–602.

Lee, S., K. Song, & Y.-J. Whang (2013) Testing functional inequalities. *Journal of Econometrics* 172(1), 14–32.

Lee, S., K. Song, & Y.-J. Whang (2018) Testing for a general class of functional inequalities. *Econometric Theory* 34(5), 1018–1064.

Lewbel, A. & 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 R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2523–2558. Elsevier Science B.V.

Olley, G.S. & A. Pakes (1996) The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64(6), 1263–1297.

Patton, A.J. & 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.

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

Seo, J. (2018) Tests of stochastic monotonicity with improved power. *Journal of Econometrics* 207(1), 53–70.

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

Talagrand, M. (2011) *Mean Field Models for Spin Glasses*. Springer.

van der Vaart, A. & J. Wellner (1996) *Weak Convergence and Empirical Processes, with Applications to Statistics*. Springer.

Wilhelm, D. (2019) Testing for the Presence of Measurement Error. Working Paper CWP48/19, CeMMAP.

A PROOFS FOR SECTION 3

Before proving the theorems from the main text, we prove three preliminary lemmas.

LEMMA A.1. *Let Assumption 3.2 be satisfied. Then there exists a set* $\mathcal{Y}_d \subset (0,1)$*, which is independent of the data, such that (i)* $|\mathcal{Y}_d| \leq n^3 + 2$*, (ii)* $\epsilon \leq F_{Y|X}(y|x) \leq 1-\epsilon$ *for some* $x \in (0,1)$ *and* $y \in \mathcal{Y}_d$*, and (iii) with probability at least* $1 - n^{-1}$*, both* T *and* $c(\alpha)$ *can be calculated with the maximum over* $y \in \mathcal{Y}$ *replaced by the maximum over* $y \in \mathcal{Y}_d$*.*

Proof. By Assumption 3.2, there exist $x^*, y^* \in (0,1)$ such that $\epsilon \leq F_{Y|X}(y^*|x^*) \leq 1-\epsilon$. For $N := n^3 + 2$, choose $y_1, \ldots, y_N \in [0,1]$ such that (i) $0 = y_1 < y_2 < \cdots < y_{N-1} < y_N = 1$,

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

(ii) $y^* = y_j$ for some $j = 2, \ldots, N-1$, and (iii) $\mathrm{P}(y_{j-1} < Y < y_j) \le n^{-3}$ for all $j = 2, \ldots, N$. Set $\mathcal{Y}_d := \{y_1, \ldots, y_N\}$. Then the probability that there is at most one Y_i, $i = 1, \ldots, n$, in each interval (y_{j-1}, y_j), $j = 2, \ldots, N$, is bounded from below by

$$\left(1 - \frac{n}{n^3}\right)^n = \left(1 - \frac{1}{n^2}\right)^n \ge 1 - \frac{1}{n}.$$

In turn, when there is at most one Y_i, $i = 1, \ldots, n$, in each interval (y_{j-1}, y_j), $j = 2, \ldots, N$, we have

$$T = \max_{(x,y,h) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{H}} \frac{\sum_{i=1}^n k_{i,h}(x) 1\{Y_i \le y\}}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}} = \max_{(x,y,h) \in \mathcal{X} \times \mathcal{Y}_d \times \mathcal{H}} \frac{\sum_{i=1}^n k_{i,h}(x) 1\{Y_i \le y\}}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}}$$

and, similarly,

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

The asserted claim follows. □

LEMMA A.2. *Let Assumptions 3.1 and 3.3 be satisfied. Also, assume that the bandwidth value b is such that $\log n \le nb$. Further, let $\mathcal{Y}_d$ be defined as in Lemma A.1. Then the estimator $\widehat{F}_{Y|X}$ in (7) satisfies*

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

where $C \in (0, \infty)$ is a constant depending only on C_X, C_L, β, and the kernel function K.

Proof. By Assumption 3.3, there exists a constant $C_P \in (0, \infty)$ depending only on C_L and β such that

$$|F_{Y|X}(y|x_2) - F_{Y|X}(y|x_1)| \le C_P |x_2 - x_1|, \quad \text{for all } x_1, x_2, y \in (0,1). \tag{A.1}$$

Further, since K is a continuous function with support $[-1, +1]$, there exists a constant $C_K \in (0, \infty)$ 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 $x \in [0,1]$,

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

by Assumption 3.1. Hence, by Bernstein's inequality,

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

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Thus, given that $\log n \leq nb$, there exists a constant $C_1 \in (0,\infty)$ depending only on C_X and C_K such that

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

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

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

almost surely, and so by the same argument,

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

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

$$\begin{aligned}
&\left|\mathrm{E}[1\{Y \leq y\}K_b(X-x)] - F_{Y|X}(y|x)\mathrm{E}[K_b(X-x)]\right| \\
&\quad = \left|\mathrm{E}[F_{Y|X}(y|X)K_b(X-x)] - F_{Y|X}(y|x)\mathrm{E}[K_b(X-x)]\right| \\
&\quad \leq \left|\mathrm{E}[(F_{Y|X}(y|X) - F_{Y|X}(y|x))K_b(X-x)]\right| \leq 2C_P b\mathrm{E}[K_b(X-x)] \leq 4C_P C_X C_K b
\end{aligned}$$

by the Lipschitz property of the functions $x \mapsto F_{Y|X}(y|x)$ in (A.13).

Now, given that $|\mathcal{X} \times \mathcal{Y}_d| \leq n(n^3+2)$, combining presented inequalities and using the union bound shows that for all $(x,y) \in \mathcal{X} \times \mathcal{Y}_d$,

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

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

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

where $C_2 \in (0,\infty)$ is a constant depending only C_X, C_P, and the kernel function K. Thus, given that $\mathrm{E}[K_b(X-x)] \geq c$ for all $x \in \mathcal{X}$ and some constant $c \in (0,\infty)$ depending only on C_X and the kernel function K, the asserted claim follows. □

LEMMA A.3. *Let Assumption 3.1 be satisfied. Then*

$$|k_{i,h}(x)| \leq Cn/h, \quad \textit{for all } i = 1,\dots,n,\ x \in (0,1), \textit{ and } h \in \mathcal{H}, \tag{A.2}$$

$$cn^3/h \leq \sum_{i=1}^{n} k_{i,h}(x)^2 \leq Cn^3/h, \quad \textit{for all } x \in (0,1) \textit{ and } h \in \mathcal{H}, \tag{A.3}$$

$$\sum_{i,j=1}^{n} |X_i - X_j|K_h(X_i - x)K_h(X_j - x) \geq cn^2 h, \quad \textit{for all } x \in (0,1) \textit{ and } h \in \mathcal{H} \tag{A.4}$$

with probability at least $1 - Cn^{-c}$, *where* $c, C \in (0,\infty)$ *are constants depending only on* C_X *and the kernel function* K.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Proof of Lemma A.3. In this proof, $c, C \in (0,\infty)$ are constants whose values may change at each appearance but can be chosen to depend only on C_X and the kernel function K.

Fix $x_1 \in (0,1)$ and $x_2 \in (x_1,1)$. Find $x_3 \in (x_2,1)$ such that

$$6C_X(1-x_3)\sup_{x\in[0,1]}K(x) \le \frac{x_2-x_1}{12C_X}\inf_{x\in[x_1,x_2]}K(x). \tag{A.5}$$

This is possible because the kernel function K satisfies

$$0 < \inf_{x\in[x_1,x_2]}K(x) \le \sup_{x\in[0,1]}K(x) < \infty.$$

Fix $x_4 \in (x_3,1)$. Further, let

$$K_1 := \inf_{x\in[x_1,x_2]}K(x),\quad K_2 := \inf_{x\in[x_3,x_4]}K(x),\quad K_3 := \sup_{x\in[x_3,1]}K(x),\quad K_4 := \sup_{x\in[-1,1]}K(x),$$

$$\widetilde{h} := \Big((x_2-x_1)\wedge(x_4-x_3)\Big)h_{\min}/2, \text{ and } \bar{h} := \sup\{h \le \widetilde{h}\colon [1/h] = 1/h\},$$

so that $L := 1/\bar{h}$ is an integer. Then, given that $h_{\min} = 1/\sqrt{n}$, it follows from Assumption 3.1, Bernstein's inequality, and the union bound that

$$\frac{n\bar{h}}{2C_X} \le \Big|\{i=1,\dots,n\colon \bar{h}(l-1) \le X_i \le \bar{h}l\}\Big| \le 2C_X n\bar{h}, \text{ for all } l=1,\dots,L, \tag{A.6}$$

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

Now, for all $x \in (0,1)$, let $l_x := \min\{l=1,\dots,L\colon \bar{h}l \ge x\}$. Then on the event in (A.6), for any $(\underline{x},\bar{x}) \in (0,1)^2$ such that $\bar{x}-\underline{x} \ge 2\bar{h}$,

$$\begin{aligned}\Big|\{i=1,\dots,n\colon \underline{x} \le X_i \le \bar{x}\}\Big| &\le \Big|\{i=1,\dots,n\colon \bar{h}(l_{\underline{x}}-1) \le X_i \le \bar{h}l_{\bar{x}}\}\Big| \\ &\le 2C_X n\bar{h}(l_{\bar{x}}-l_{\underline{x}}+1) \le 6C_X n(\bar{x}-\underline{x})\end{aligned}$$

and

$$\begin{aligned}\Big|\{i=1,\dots,n\colon \underline{x} \le X_i \le \bar{x}\}\Big| &\ge \Big|\{i=1,\dots,n\colon \bar{h}l_{\underline{x}} \le X_i \le \bar{h}(l_{\bar{x}}-1)\}\Big| \\ &\ge \frac{n\bar{h}(l_{\bar{x}}-1-l_{\underline{x}})}{2C_X} \ge \frac{n(\bar{x}-\underline{x})}{6C_X}.\end{aligned}$$

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

$$\frac{n(\bar{x}-\underline{x})}{6C_X} \le \Big|\{i=1,\dots,n\colon \underline{x} \le X_i \le \bar{x}\}\Big| \le 6C_X n(\bar{x}-\underline{x}), \quad \text{for all } (\underline{x},\bar{x}) \in \bar{\mathcal{X}}, \tag{A.7}$$

where $\bar{\mathcal{X}} := \{(\underline{x},\bar{x}) \in (0,1)^2\colon \bar{x}-\underline{x} \ge 2\bar{h}\}$. We will prove that inequalities in (A.2), (A.3), and (A.4), with $x \in (0,1)$ replaced by $x \in (0,1/2]$, hold on the event in (A.7). Since the same inequalities with $x \in (0,1)$ replaced by $x \in [1/2,1)$ follow from the same arguments, this will imply the asserted claim of the lemma.

So, for the rest of the proof, assume that the event in (A.7) holds. Fix $x \in (0,1/2]$ and $h \in \mathcal{H}$. To prove (A.2), we have for all $i=1,\dots,n$ that

$$\begin{aligned}h^2|k_{i,h}(x)| &\le 2h^2\sum_{j=1}^n K_h(X_i-x)K_h(X_j-x) \le 2K_4 h\sum_{j=1}^n K_h(X_j-x) \\ &\le 2K_4^2\Big|\{j\colon x-h \le X_j \le x+h\}\Big| \le 24K_4^2C_X nh = Cnh,\end{aligned}$$

where the fourth inequality follows from (A.7) since $h \ge \bar{h}$.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

To prove the right-hand side inequality in (A.3), we have

$$\begin{aligned}\frac{h^4}{4}\sum_{i=1}^n k_{i,h}(x)^2 &= h^2\sum_{i=1}^n K_h(X_i-x)^2\Big(h\sum_{j=1}^n \mathrm{sign}(X_i-X_j)K_h(X_j-x)\Big)^2\\ &\le h^2\sum_{i=1}^n K_h(X_i-x)^2\Big(h\sum_{j=1}^n K_h(X_j-x)\Big)^2\\ &\le K_4^4\Big|\{j\colon x-h\le X_j\le x+h\}\Big|^3 \le K_4^4(12C_X nh)^3 = C(nh)^3,\end{aligned}$$

where the last inequality again follows from (A.7) since $h\ge \bar{h}$.

To prove the left-hand side inequality in (A.3), we have for all $i=1,\dots,n$ such that $x+x_3h\le X_i\le x+x_4h$ that

$$\begin{aligned}&h\sum_{j=1}^n \mathrm{sign}(X_i-X_j)K_h(X_j-x)\\ &\quad\ge h\sum_{j\colon x+x_1h\le X_j\le x+x_2h} K_h(X_j-x) - h\sum_{j\colon x+x_3h\le X_j\le x+h} K_h(X_j-x)\\ &\quad\ge K_1\Big|\{j\colon x+x_1h\le X_j\le x+x_2h\}\Big| - K_3\Big|\{j\colon x+x_3h\le X_j\le x+h\}\Big|\\ &\quad\ge \frac{K_1 n(x_2-x_1)h}{6C_X} - 6K_3C_X n(1-x_3)h \ge \frac{K_1 n(x_2-x_1)h}{12C_X},\end{aligned}$$

where the third inequality follows from (A.7) since $((x_2-x_1)\wedge(1-x_3))h\ge 2\bar{h}$ and the fourth from (A.5). Hence,

$$\begin{aligned}\frac{h^4}{4}\sum_{i=1}^n k_{i,h}(x)^2 &= h^2\sum_{i=1}^n K_h(X_i-x)^2\Big(h\sum_{j=1}^n \mathrm{sign}(X_i-X_j)K_h(X_j-x)\Big)^2\\ &\ge h^2\sum_{i\colon x+x_3h\le X_i\le x+x_4h} K_h(X_i-x)^2\Big(h\sum_{j=1}^n \mathrm{sign}(X_i-X_j)K_h(X_j-x)\Big)^2\\ &\ge K_2^2\Big|\{x+x_3h\le X_i\le x+x_4h\}\Big|\times\Big(\frac{K_1 n(x_2-x_1)h}{12C_X}\Big)^2\\ &\ge \frac{K_2^2 n(x_4-x_3)h}{6C_X}\times\Big(\frac{K_1 n(x_2-x_1)h}{12C_X}\Big)^2 = c(nh)^3,\end{aligned}$$

where the last inequality follows from (A.7) since $(x_4-x_3)h\ge 2\bar{h}$.

Finally, to prove (A.4), we have

$$\begin{aligned}&h\sum_{i,j=1}^n |X_i-X_j|K_h(X_i-x)K_h(X_j-x)\\ &\quad\ge h\sum_{i\colon x+x_1h\le X_i\le x+x_2h}\ \sum_{j\colon x+x_3h\le X_j\le x+x_4h} |X_i-X_j|K_h(X_i-x)K_h(X_j-x)\\ &\quad\ge (x_3-x_2)h^2\sum_{i\colon x+x_1h\le X_i\le x+x_2h}\ \sum_{j\colon x+x_3h\le X_j\le x+x_4h} K_h(X_i-x)K_h(X_j-x)\end{aligned}$$

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

$$\geq (x_3 - x_2)K_1K_2\Big|\{i\colon x + x_1h \leq X_i \leq x + x_2h\}\Big| \times \Big|\{j\colon x + x_3h \leq X_j \leq x + x_4h\}\Big|$$
$$\geq \frac{(x_2 - x_1)(x_4 - x_3)(x_3 - x_2)K_1K_2n^2h^2}{36C_X^2} = c(nh)^2,$$

where the third inequality follows from (A.7) since $((x_2 - x_1) \wedge (x_4 - x_3))h \geq 2\bar{h}$. This completes the proof of the lemma. □

Proof of Theorem 3.1. In this proof, $c, C \in (0, \infty)$ are constants whose values may change at each appearance but can be chosen to depend only on C_X, C_L, ϵ, β, u, and the kernel function K. Also, let $\mathcal{Y}_d$ be defined as in Lemma A.1.

Denote

$$T_0 := \max_{(x,y,h)\in\mathcal{X}\times\mathcal{Y}_d\times\mathcal{H}} \frac{\sum_{i=1}^n k_{i,h}(x)(1\{Y_i \leq y\} - F_{Y|X}(y|X_i))}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}},$$

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

and

$c_0(\alpha) :=$ $(1-\alpha)$ conditional quantile of T_0^b given the data.

By Lemma A.1, under the null, $T \leq T_0$ and $c(\alpha) = c_0(\alpha)$ with probability at least $1 - n^{-1}$, and if Y is independent of X, then $T = T_0$ and $c(\alpha) = c_0(\alpha)$ with probability at least $1 - n^{-1}$. Therefore, to prove the asserted claims, we apply Theorem 4.3, conditional on $(X_i)_{i=1}^n$, with $S_n = T_0$, $c_n(\alpha) = c_0(\alpha)$,

$$Z_{ij} = \frac{\sqrt{n}k_{i,h_j}(x_j)(1\{Y_i \leq y_j\} - F_{Y|X}(y_j|X_i))}{\left(\sum_{l=1}^n k_{l,h_j}(x_j)^2\right)^{1/2}}, \quad \text{for all } i = 1, \ldots, n \text{ and } j = 1, \ldots, p,$$

and

$$\widehat{Z}_{ij} = \frac{\sqrt{n}k_{i,h_j}(x_j)(1\{Y_i \leq y_j\} - \widehat{F}_{Y|X}(y_j|X_i))}{\left(\sum_{l=1}^n k_{l,h_j}(x_j)^2\right)^{1/2}}, \quad \text{for all } i = 1, \ldots, n \text{ and } j = 1, \ldots, p,$$

where $p = |\mathcal{X} \times \mathcal{Y}_d \times \mathcal{H}|$ and $j \mapsto (x_j, y_j, h_j)$ is a one-to-one mapping from $\{1, \ldots, p\}$ to $\mathcal{X} \times \mathcal{Y}_d \times \mathcal{H}$. In particular, it suffices to show that (24), (25), (26), and (29), with expectation and probability signs replaced by conditional expectation and probability signs given $(X_i)_{i=1}^n$, hold with probability at least $1 - Cn^{-c}$ for some B_n, ζ_n, and κ_n such that

$$\left(\frac{B_n^2 \log^{10}(pn)}{n}\right)^{1/6} + \zeta_n \log^{3/2}(pn) + \kappa_n \leq Cn^{-c}. \tag{A.8}$$

Moreover, the constant 1 on the right-hand side of (24) can be replaced by c since then Theorem 4.3 can be applied with rescaled Z_{ij}'s and $\widehat{Z}_{ij}$'s. To be explicit, the conditions to be verified, in addition to (A.8), have the following form:

$$\mathrm{P}\left(\frac{1}{n}\sum_{i=1}^n \mathrm{E}[Z_{ij}^2 \mid (X_i)_{i=1}^n] \geq c \text{ for some } j = 1, \ldots, p\right) \geq 1 - Cn^{-c}, \tag{A.9}$$

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

$$\mathrm{P}\left(\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[|Z_{ij}|^{2+k} \mid (X_i)_{i=1}^n] \le B_n^k \text{ for all } j=1,\ldots,p \text{ and } k=1,2\right) \ge 1-Cn^{-c}, \tag{A.10}$$

$$\mathrm{P}\left(\mathrm{E}[\exp(|Z_{ij}|/B_n) \mid (X_i)_{i=1}^n] \le 2 \text{ for all } i=1,\ldots,n \text{ and } j=1,\ldots,p\right) \ge 1-Cn^{-c}, \tag{A.11}$$

$$\mathrm{P}\left(\mathrm{P}\left(\max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}(\widehat{Z}_{ij}-Z_{ij})^2 > \zeta_n^2 \mid (X_i)_{i=1}^n\right) < \kappa_n\right) \ge 1-Cn^{-c}. \tag{A.12}$$

To verify these conditions, we proceed in 4 steps.

Step 1. Here, we verify (A.9). To do so, note that by construction of $\mathcal{Y}_d$ in Lemma A.1, there exist $x^* \in (0,1)$ and $y^* \in \mathcal{Y}_d$ such that $\epsilon \le F_{Y|X}(y^*|x^*) \le 1-\epsilon$. Further, by Assumption 3.3, there exists a constant $C_P \in (0,\infty)$ depending only on C_L and β such that

$$|F_{Y|X}(y|x_2) - F_{Y|X}(y|x_1)| \le C_P|x_2-x_1|, \quad \text{for all } x_1,x_2,y \in (0,1). \tag{A.13}$$

Now, fix $j=1,\ldots,p$ such that $x^*-\epsilon/(2C_P) < x_j < x^*+\epsilon/(2C_P)$, $y_j = y^*$, and $h_j \le \epsilon/(4C_P)$, which exists with probability at least $1-Cn^{-c}$. For this j,

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[Z_{ij}^2 \mid (X_i)_{i=1}^n] = \frac{\sum_{i=1}^n k_{i,h_j}(x_j)^2 F_{Y|X}(y_j|X_i)(1-F_{Y|X}(y_j|X_i))}{\sum_{l=1}^n k_{l,h_j}(x_j)^2} \ge \frac{\epsilon^2}{16}.$$

This gives (A.9) and completes the first step.

Step 2. Here, we verify (A.10) and (A.11). To do so, note that

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

with probability at least $1-Cn^{-c}$ by Lemma A.3. Thus, with the same probability,

$$\frac{1}{n}\sum_{i=1}^{n}|Z_{ij}|^3 \le \frac{8\sqrt{n}\sum_{i=1}^n |k_{i,h_j}(x_j)|^3}{\left(\sum_{l=1}^n k_{l,h_j}(x_j)^2\right)^{3/2}} \le \frac{8\sqrt{n}\max_{1\le i\le n}\max_{1\le j\le p}|k_{i,h_j}(x_j)|}{\left(\sum_{l=1}^n k_{l,h_j}(x_j)^2\right)^{1/2}} \le 8\widetilde{B}_n$$

and

$$\frac{1}{n}\sum_{i=1}^{n}|Z_{ij}|^4 \le \frac{16n\sum_{i=1}^n |k_{i,h_j}(x_j)|^4}{\left(\sum_{l=1}^n k_{l,h_j}(x_j)^2\right)^2} \le \frac{16n\max_{1\le i\le n}\max_{1\le j\le p}|k_{i,h_j}(x_j)|^2}{\sum_{l=1}^n k_{l,h_j}(x_j)^2} \le 16\widetilde{B}_n,$$

for all $j=1,\ldots,p$ and

$$|Z_{ij}| \le \frac{2\sqrt{n}|k_{i,h_j}(x_j)|}{\left(\sum_{l=1}^n k_{l,h_j}(x_j)^2\right)^{1/2}} \le 2\widetilde{B}_n$$

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

for all $i = 1, \dots, n$ and $j = 1, \dots, p$. This shows that (A.10) and (A.11) hold with $B_n := 8\widetilde{B}_n$ and completes the second step.

Step 3. Here, we verify (A.12). To do so, note that

$$\max_{1\le j\le p} \frac{1}{n}\sum_{i=1}^{n} (\widehat{Z}_{ij} - Z_{ij})^2 \le \max_{1\le i\le n} \max_{1\le j\le p} |\widehat{F}_{Y|X}(y_j|X_i) - F_{Y|X}(y_j|X_i)|^2$$
$$= \max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_d} |\widehat{F}_{Y|X}(y|x) - F_{Y|X}(y|x)|^2$$

and that there exist constants $c_F, C_F \in (0,\infty)$ depending only on C_X, C_L, β, and the kernel function K such that

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

with probability at least $1 - C_F n^{-c_F}$ by Lemma A.2. Hence, by Markov's inequality,

$$\mathrm{P}\Big(\mathrm{P}\Big(\max_{1\le j\le p} \frac{1}{n}\sum_{i=1}^{n} (\widehat{Z}_{ij} - Z_{ij})^2 > (C_F n^{-c_F})^2 \mid (X_i)_{i=1}^n\Big) \ge \sqrt{C_F n^{-c_F}}\Big)$$
$$\le \frac{\mathrm{P}\Big(\max_{1\le j\le p} \frac{1}{n}\sum_{i=1}^{n} (\widehat{Z}_{ij} - Z_{ij})^2 > (C_F n^{-c_F})^2\Big)}{\sqrt{C_F n^{-c_F}}} \le \sqrt{C_F n^{-c_F}}.$$

This shows that (A.12) holds with $\zeta_n := C_F n^{-c_F}$ and $\kappa_n := (C_F n^{-c_F})^{1/2}$ and completes the third step.

Step 4. Here, we note that (A.8) holds with chosen B_n, ζ_n, and κ_n since $\log p \le C \log n$ by construction of the sets $\mathcal{X}$, $\mathcal{Y}_d$, and $\mathcal{H}$. Thus, we have verified conditions (A.8)–(A.12), and the asserted claims of the theorem now follow by applying Theorem (4.3) conditional on $(X_i)_{i=1}^n$. This completes the proof of the theorem. □

Proof of Theorem 3.2. In this proof, $c, C \in (0,\infty)$ are constants whose values may change at each appearance but can be chosen to depend only on C_X, u, and the kernel function K. Also, let $\mathcal{Y}_d$ be defined as in Lemma A.1 and T_0^b and $c_0(\alpha)$ as in the proof of Theorem 3.1.

We proceed in 3 steps. In the first step, we bound $c(\alpha)$ from above. In the second step, we bound T from below. In the third step, we combine the bounds on $c(\alpha)$ and T to complete the proof.

Step 1. Here, we show that

$$\mathrm{P}(c(\alpha) > C\sqrt{\log n}) \le n^{-1}. \tag{A.14}$$

To do so, note that by Lemma A.1, $c(\alpha) = c_0(\alpha)$ with probability at least $1 - n^{-1}$. Also, conditional on the data, the random variables

$$T_{x,y,h}^b := \frac{\sum_{i=1}^n e_i k_{i,h}(x)(1\{Y_i \le y\} - \widehat{F}_{Y|X}(y|X_i))}{\big(\sum_{i=1}^n k_{i,h}(x)^2\big)^{1/2}}, \quad (x,y,h) \in \mathcal{X}\times\mathcal{Y}_d\times\mathcal{H},$$

are zero-mean Gaussian with variance bounded from above by

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

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

for all $(x,y,h) \in \mathcal{X} \times \mathcal{Y}_d \times \mathcal{H}$ since $\widehat{F}_{Y|X}(y|x) \in [0,1]$ for all $(x,y) \in \mathcal{X} \times \mathcal{Y}_d$. Therefore, (A.14) follows from Lem. A.3.1 in Talagrand (2011) since $c_0(\alpha)$ is the $(1-\alpha)$ conditional quantile of T_0^b given the data, $T_0^b = \max_{(x,y,h)\in\mathcal{X}\times\mathcal{Y}_d\times\mathcal{H}} T_{x,y,h}^b$, and $p := |\mathcal{X} \times \mathcal{Y}_d \times \mathcal{H}|$, the number of elements in the set $\mathcal{X} \times \mathcal{Y}_d \times \mathcal{H}$, satisfies $\log p \leq C \log n$.

Step 2. Here, we show that there exist $n_0 \in \mathbb{N}$, and $h_*, \ell_* \in (0,1]$, all possibly depending on $F_{Y|X}$, such that for all $n \geq n_0$,

$$\mathrm{P}\left(T < ch_*\ell_*\sqrt{nh_*}\right) \leq \exp\left(-c(h_*\ell_*)^2 nh_*\right) + Cn^{-c}. \tag{A.15}$$

To do so, note that since the functions $x \mapsto F_{Y|X}(y|x)$ are continuously differentiable for all $y \in (0,1)$ and

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

it follows that there exist $y_* \in (0,1)$, $x_l, x_r \in (0,1)$ satisfying $x_l < x_r$, and $\ell_* \in (0,1]$ such that

$$\frac{\partial}{\partial x} F_{Y|X}(y_*|x) \geq \ell_*, \quad \text{for all } x \in (x_l, x_r).$$

Moreover, by the definition of $\mathcal{H}$, there exist $n_0 \in \mathbb{N}$ and $h_* \in (0,1]$ such that for all $n \geq n_0$, we have that $h_* \in \mathcal{H}$ and $h_* \leq (x_r - x_l)/3$.

Next, define x_{n*} using the following rule: if there exists at least one $i = 1, \ldots, n$ such that $X_i \in [2x_l/3 + x_r/3, x_l/3 + 2x_r/3]$, set $x_{n*} := X_i$ for any such i; otherwise set $x_{n*} := (x_l + x_r)/2$. Then $x_{n*} \in \mathcal{X}$ with probability at least $1 - Cn^{-c}$, and so

$$\mathrm{P}(T < T_{x_{n*}, y_*, h_*}) \leq Cn^{-c},$$

where

$$T_{x_{n*}, y_*, h_*} := \frac{\sum_{i=1}^n k_{i,h_*}(x_{n*}) 1\{Y_i \leq y_*\}}{\left(\sum_{i=1}^n k_{i,h_*}(x_{n*})^2\right)^{1/2}},$$

since

$$T = \max_{(x,y,h)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{H}} \frac{\sum_{i=1}^n k_{i,h}(x) 1\{Y_i \leq y\}}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}} = \max_{(x,y,h)\in\mathcal{X}\times(0,1)\times\mathcal{H}} \frac{\sum_{i=1}^n k_{i,h}(x) 1\{Y_i \leq y\}}{\left(\sum_{i=1}^n k_{i,h}(x)^2\right)^{1/2}}$$

To bound T_{x_{n*}, y_*, h_*} from below, we have

$$\begin{aligned}
&\mathrm{E}\Big[\sum_{i=1}^n k_{i,h_*}(x_{n*}) 1\{Y_i \leq y_*\} \mid (X_i)_{i=1}^n\Big] = \sum_{i=1}^n k_{i,h_*}(x_{n*}) F_{Y|X}(y_*|X_i) \\
&\quad = \sum_{i,j=1}^n (F_{Y|X}(y_*|X_i) - F_{Y|X}(y_*|X_j)) \mathrm{sign}(X_i - X_j) K_{h_*}(X_i - x_{n*}) K_{h_*}(X_j - x_{n*}) \\
&\quad \geq \ell_* \sum_{i,j=1}^n (X_i - X_j) \mathrm{sign}(X_i - X_j) K_{h_*}(X_i - x_{n*}) K_{h_*}(X_j - x_{n*}) \geq cn^2 h_* \ell_*
\end{aligned}$$

with probability at least $1 - Cn^{-c}$ by Lemma A.3. Thus,

$$\mathrm{E}\left[T_{x_{n*}, y_*, h_*} \mid (X_i)_{i=1}^n\right] = \mathrm{E}\left[\frac{\sum_{i=1}^n k_{i,h_*}(x_{n*}) 1\{Y_i \leq y_*\}}{\left(\sum_{l=1}^n k_{i,h_*}(x_{n*})^2\right)^{1/2}} \mid (X_i)_{i=1}^n\right] \geq ch_*\ell^*\sqrt{nh_*}$$

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

with probability at least $1 - Cn^{-c}$, again by Lemma A.3. Also,

$$\mathrm{Var}(T_{x_{n*},y_*,h_*} \mid (X_i)_{i=1}^n) \le 1$$

and

$$\max_{1\le i\le n} \frac{|k_{i,h_*}(x_{n*})|}{(\sum_{l=1}^n k_{l,h_*}(x_{n*})^2)^{1/2}} \le \frac{1}{\sqrt{cnh_*}}$$

with probability at least $1 - Cn^{-c}$, again by Lemma A.3. Hence, by Bernstein's inequality,

$$T_{x_{n*},y_*,h_*} \ge ch_*\ell_*\sqrt{nh_*}$$

with probability at least

$$1 - Cn^{-c} - \exp\left(-\frac{c(h_*\ell_*)^2 nh_*}{1+h_*\ell_*}\right) \ge 1 - Cn^{-c} - \exp\left(-c(h_*\ell_*)^2 nh_*\right),$$

since $h_*, \ell_* \le 1$, which gives the asserted claim of this step.

Step 3. Here, we complete the proof of the theorem. To do so, note that by Steps 1 and 2,

$$\mathrm{P}(T > c(\alpha)) \ge \mathrm{P}(ch_*\ell_*\sqrt{nh_*} > C\sqrt{\log n}) - \exp\left(-\frac{c(h_*\ell_*)^2 nh_*}{1+h_*\ell_*}\right) - Cn^{-c} \to 1$$

since h^* and ℓ^* are independent of n. This gives the asserted claim and completes the proof of the theorem. □

Proof of Theorem 3.3. In this proof, $c, C \in (0,\infty)$ are constants whose values may change at each appearance but can be chosen to depend only on $x_r - x_l$, C_X, u, and the kernel function K. Also, let c_I be the constant c on the left-hand side of (A.15) and let C_I be the constant C in (A.14). We will assume throughout the proof that (13) holds with $C = C_0$, where

$$C_0 = \frac{(C_I/c_I)^2}{(u(x_r - x_l)/3)^2}.$$

Next, let n_0 be the largest $n \in \mathbb{N}$ such that

$$\frac{1}{\sqrt{n}} > \frac{u(x_r - x_l)}{3}.$$

Throughout the proof, we will assume that $n > n_0$ since the asserted claim for $n \le n_0$ follows by choosing $c, C \in (0,\infty)$ in (14) such that $Cn_0^{-1} \ge 1$.

Now, since $n > n_0$, there exists $h_* \in \mathcal{H}$ such that

$$\frac{u(x_r - x_l)}{3} \le h_* < \frac{x_r - x_l}{3}.$$

Also, set

$$\ell_* := \sqrt{\frac{C_0 \log n}{n}}.$$

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Then, since

$$\frac{\partial}{\partial x}F_{Y|X}(y|x) \geq \ell_*, \quad \text{for some } y \in (0,1) \text{ and all } x \in (x_l, x_r),$$

by assumption, it follows from the same arguments as those in Step 2 of the proof of Theorem 3.2 that (A.15) holds here with chosen h_* and ℓ_*. Thus,

$$\mathrm{P}\left(T < c_I\left(\frac{u(x_r - x_l)}{3}\right)^{3/2}\sqrt{C_0 \log n}\right) \leq \exp\left(-c\left(\frac{u(x_r - x_l)}{3}\right)^3 \log n\right) + Cn^{-c}.$$

Also, by Step 1 in the proof of Theorem 3.2,

$$\mathrm{P}(c(\alpha) > C_I\sqrt{\log n}) \leq n^{-1}.$$

Combining these bounds and using the fact that

$$c_I\left(\frac{u(x_r - x_l)}{3}\right)^{3/2}\sqrt{C_0 \log n} = C_I\sqrt{\log n},$$

by construction, gives the asserted claim. □

Proof of Theorem 3.4. In this proof, $c, C \in (0, \infty)$ are constants whose values may change at each appearance but can be chosen to depend only on C_X, C_L, β, u, and the kernel function K. Also, define c_I and C_I as in the proof of Theorem 3.3. We will assume that (15) with $C = C_0$, where

$$C_0 = \left(\frac{2C_I/c_I}{(u/3)^{3/2}(2C_L)^{-3/(2\beta)}}\right)^{2+3/\beta} \vee 1.$$

Next, let n_0 be the largest $n \in \mathbb{N}$ such that

$$\frac{1}{\sqrt{n}} > \frac{u}{3}\left(\frac{1}{2C_L}\right)^{1/\beta}\left(\frac{\log n}{n}\right)^{1/(2\beta+3)}.$$

Throughout the proof, we will assume that $n > n_0$ since the asserted claim for $n \leq n_0$ follows by choosing $c, C \in (0, \infty)$ in (16) such that $Cn_0^{-1} \geq 1$.

Now, since $n > n_0$, there exists $h_* \in \mathcal{H}$ such that

$$\frac{u}{3}\left(\frac{1}{2C_L}\right)^{1/\beta}\left(\frac{\log n}{n}\right)^{1/(2\beta+3)} \leq h_* < \frac{1}{3}\left(\frac{1}{2C_L}\right)^{1/\beta}\left(\frac{\log n}{n}\right)^{1/(2\beta+3)}.$$

Also, set

$$\ell_* := \frac{1}{2}\left(\frac{C_0 \log n}{n}\right)^{\beta/(2\beta+3)}.$$

Then, since $C_0 \geq 1$ and

$$\frac{\partial}{\partial x}F_{Y|X}(y|x) \geq 2\ell_*, \quad \text{for some } x, y \in (0,1),$$

it follows from Assumption 3.3 that there exist $x_l, x_r \in (0,1)$ such that

$$x_r - x_l = \left(\frac{1}{2C_L}\right)^{1/\beta}\left(\frac{\log n}{n}\right)^{1/(2\beta+3)}$$

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

and

$$\frac{\partial}{\partial x}F_{Y|X}(y|x) \geq \ell_*, \quad \text{for some } y \in (0,1) \text{ and all } x \in (x_l, x_r).$$

Thus, it follows from the same arguments as those in Step 2 of the proof of Theorem 3.2 that (A.15) holds here with chosen h_* and ℓ_*. Hence,

$$\mathrm{P}\left(T < C_I\sqrt{\log n}\right) \leq \exp\left(-c\log n\right) + Cn^{-c}.$$

Also, by Step 1 in the proof of Theorem 3.2,

$$\mathrm{P}(c(\alpha) > C_I\sqrt{\log n}) \leq n^{-1}.$$

Combining these bounds gives the asserted claim. □

Proof of Theorem 3.5. Let $g_0 : \mathbb{R} \to \mathbb{R}$ be a function defined by $g_0(-1) := 0$ and

$$g_0'(x) := \begin{cases} 0, & \text{if } x \leq -1, \\ (1-|x|)^\beta, & \text{if } -1 < x \leq -1/2, \\ |x|^\beta, & \text{if } -1/2 < x \leq 0, \\ -x^\beta, & \text{if } 0 < x \leq 1/2, \\ -(1-x)^\beta, & \text{if } 1/2 < x \leq 1, \\ 0, & \text{if } x > 1 \end{cases}$$

and let $g : \mathbb{R} \to \mathbb{R}$ be a function defined by $g(x) := c_0 g_0(x)$ for all $x \in \mathbb{R}$, where $c_0 = 1/(4\sqrt{5})$. It is easy to check that the function g is such that

$$|g(x)| \leq 1, \quad \text{for all } x \in \mathbb{R}, \tag{A.16}$$

and

$$g(x) = 0, \quad \text{for all } x \in \mathbb{R} \setminus (-1, +1). \tag{A.17}$$

Also, let

$$J := \left[\left(\frac{n}{\log n}\right)^{1/(2\beta+3)}\right], \quad h := \frac{1}{2J}, \tag{A.18}$$

and

$$s_j := 2jh - h, \quad \text{for all } j = 1, \ldots, J.$$

Note that

$$J \geq n^{1/10} \tag{A.19}$$

for all $n \geq n_0$ and some universal $n_0 \in \mathbb{N}$ since $\beta \leq 1$. Throughout this proof, we will assume that $n \geq n_0$, so that (A.19) holds, since the asserted claim for $n < n_0$ follows by choosing constants $c, C \in (0, \infty)$ in (18) such that $Cn_0^{-c} \geq 1$. Also, by (A.19), $J \geq 1$, and so

$$\left(\frac{n}{\log n}\right)^{1/(2\beta+3)} \leq J + 1 \leq 2J = \frac{1}{h}.$$

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Hence,

$$h^{2\beta+3} \leq \frac{\log n}{n} \quad \text{and} \quad h \leq 1. \tag{A.20}$$

Further, let X be a random variable distributed uniformly on $[0,1]$ and let Z be a $N(0,1)$ random variable that is independent of X. Let $(X_i, Z_i)_{i=1}^n$ be a random sample from the distribution of the pair (X,Z). Moreover, let Φ and ϕ denote the cdf and the pdf of the $N(0,1)$ distribution, respectively.

Next, let M_0 be the distribution of $(X, \Phi(Z))$ and for all $j = 1, \ldots, J$, let M_j be the distribution of

$$\left(X, \Phi\left(Z + h^{1+\beta} g\left(\frac{X - s_j}{h}\right)\right)\right).$$

Then for all $x, y \in (0,1)$,

$$F_{Y|X}^{M_0}(y|x) = y \quad \text{and} \quad f_{Y|X}^{M_0}(y|x) := \frac{\partial}{\partial y} F_{Y|X}^{M_0}(y|x) = 1.$$

Moreover, for all $x, y \in (0,1)$ and $j = 1, \ldots, J$,

$$F_{Y|X}^{M_j}(y|x) = \Phi\left(\Phi^{-1}(y) - h^{1+\beta} g\left(\frac{x - s_j}{h}\right)\right)$$

and

$$f_{Y|X}^{M_j}(y|x) := \frac{\partial}{\partial y} F_{Y|X}^{M_j}(y|x) = \frac{\phi\left(\Phi^{-1}(y) - h^{1+\beta} g\left(\frac{x-s_j}{h}\right)\right)}{\phi(\Phi^{-1}(y))}.$$

Using (A.16), (A.18), and (A.20), this choice of the distributions M_j ensures that for all $j = 1, \ldots, J$,

$$\begin{aligned} \sup_{x \in (0,1)} \frac{\partial}{\partial x} F_{Y|X}^{M_j}(1/2|x) &\geq \frac{\partial}{\partial x} F_{Y|X}^{M_j}(1/2|x)\Big|_{x = s_j - h/2} \\ &\geq c_0 (h/2)^\beta \phi(h^{1+\beta}) \geq \frac{c_0 \phi(1)}{4} \left(\frac{\log n}{n}\right)^{\beta/(2\beta+3)}, \end{aligned}$$

so that (17) holds with $c = c_0 \phi(1)/4$ and $M = M_j$ for any $j = 1, \ldots, J$.

Now, fix any $\gamma \in (0,1)$ and any test φ such that $\sup_{M \in \mathcal{M}_{\beta,0}} \mathrm{E}_M[\varphi] \leq \gamma$. Since $M_0 \in \mathcal{M}_{\beta,0}$ by construction, it follows that $\mathrm{E}_{M_0}[\varphi] \leq \gamma$. Thus,

$$\min_{1 \leq j \leq J} \mathrm{E}_{M_j}[\varphi] - \gamma \leq \frac{1}{J} \sum_{j=1}^J \mathrm{E}_{M_j}[\varphi] - \mathrm{E}_{M_0}[\varphi] \leq \mathrm{E}\left[\left|\frac{1}{J} \sum_{j=1}^J \rho_j - 1\right|\right],$$

where

$$\begin{aligned} \rho_j &:= \prod_{i=1}^n \frac{f_{Y|X}^{M_j}(\Phi(Z_i)|X_i)}{f_{Y|X}^{M_0}(\Phi(Z_i)|X_i)} = \prod_{i=1}^n f_{Y|X}^{M_j}(\Phi(Z_i)|X_i) \\ &= \exp\left(\sum_{i=1}^n Z_i h^{1+\beta} g\left(\frac{X_i - s_j}{h}\right) - \frac{1}{2} \sum_{i=1}^n h^{2(1+\beta)} g\left(\frac{X_i - s_j}{h}\right)^2\right) \end{aligned}$$

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

for all $j = 1, \dots, J$. Here, given that $(Z_i)_{i=1}^n$ are i.i.d. $N(0,1)$ and are independent of $(X_i)_{i=1}^n$, we have that $\mathrm{E}[\rho_j \mid (X_i)_{i=1}^n] = 1$ for all $j = 1, \dots, J$ and that conditional on $(X_i)_{i=1}^n$, random variables $\rho_1, \dots, \rho_J$ are independent. Hence,

$$\left(\mathrm{E}\left[\left|\frac{1}{J}\sum_{j=1}^J \rho_j - 1\right|\right]\right)^2 \le \mathrm{E}\left[\left|\frac{1}{J}\sum_{j=1}^J \rho_j - 1\right|^2\right] \le \frac{1}{J^2}\sum_{j=1}^J \mathrm{E}[\rho_j^2].$$

Also, for all $j = 1, \dots, J$,

$$\begin{aligned}\mathrm{E}[\rho_j^2] &= \mathrm{E}\Big[\exp\Big(\sum_{i=1}^n h^{2(1+\beta)} g\Big(\frac{X_i - s_j}{h}\Big)^2\Big)\Big] = \prod_{i=1}^n \mathrm{E}\Big[\exp\Big(h^{2(1+\beta)} g\Big(\frac{X_i - s_j}{h}\Big)^2\Big)\Big] \\ &\le \prod_{i=1}^n \mathrm{E}\Big[1 + 2h^{2(1+\beta)} g\Big(\frac{X_i - s_j}{h}\Big)^2\Big] \le \Big(1 + 4c_0^2 h^{3+2\beta}\Big)^n,\end{aligned}$$

where the second line follows from (A.16), (A.17), (A.20), and the fact that $e^x \le 1 + 2x$ for all $x \in [0,1]$. Thus,

$$\begin{aligned}\frac{1}{J^2}\sum_{j=1}^J \mathrm{E}[\rho_j^2] &\le \exp\Big(-\log J + n\log(1 + 4c_0^2 h^{3+2\beta})\Big) \\ &\le \exp\Big(-\log J + 4c_0^2 n h^{3+2\beta}\Big) \\ &\le \exp\Big(-\frac{1}{10}\log n + 4c_0^2 \log n\Big) \le n^{-1/20}\end{aligned}$$

since $4c_0^2 = 1/20$ by the choice of c_0, where the second inequality follows from the fact that $\log(1+x) \le x$ for all $x > 0$, and the third from (A.19) and (A.20). Hence, (18) holds with $C = 1$, $c = 1/40$, and $M = M_j$ for at least one $j = 1, \dots, J$.

Thus, it remains to show that $M_j \in \mathcal{M}_\beta$ for all $j = 1, \dots, J$, which means that if (X, Y) is distributed according to M_j for any $j = 1, \dots, J$, then Assumptions 3.1, 3.2, and 3.3 are satisfied. So, fix $j = 1, \dots, J$ and assume that (X, Y) is distributed according to M_j. Then Assumption 3.1 holds trivially since X is distributed uniformly on $[0,1]$ and $C_X \ge 1$. Since $\epsilon \in (0, 1/2)$, Assumption 3.2 also holds trivially with $y = 1/2$ and any x such that $|x - s_j| > h$ because these values of x and y give $F_{Y|X}(y|x) = 1/2$.

To verify Assumption 3.3, note that for all $x \in \mathbb{R}$, $\phi(x) \le 1$ and $\phi'(x) \le 1$. Also, for all $x, y \in (0,1)$,

$$\frac{\partial}{\partial x} F_{Y|X}^{M_j}(y|x) = -h^\beta g'\left(\frac{x - s_j}{h}\right)\phi\left(\Phi^{-1}(y) - h^{1+\beta} g\left(\frac{x - s_j}{h}\right)\right).$$

Thus, for all $x_1, x_2, y \in (0,1)$,

$$\begin{aligned}&\left|\frac{\partial}{\partial x} F_{Y|X}^{M_j}(y|x_2) - \frac{\partial}{\partial x} F_{Y|X}^{M_j}(y|x_1)\right| \\ &\quad\le h^\beta \left|g'\Big(\frac{x_2 - s_j}{h}\Big) - g'\Big(\frac{x_1 - s_j}{h}\Big)\right| + h^{1+\beta}\left|g\Big(\frac{x_2 - s_j}{h}\Big) - g\Big(\frac{x_1 - s_j}{h}\Big)\right| \\ &\quad\le c_0\Big(|x_2 - x_1|^\beta + h^\beta |x_2 - x_1|\Big) \le 2c_0 |x_2 - x_1|^\beta \le |x_2 - x_1|^\beta\end{aligned}$$

since $c_0 \le 1/2$. Hence, given that $C_L \ge 1$, Assumption 3.3 is satisfied. This completes the proof of the theorem. □

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

B PROOFS FOR SECTION 4

In this section, we will use L_2 and ψ_1 norms of random variables. For any random variable Y, these norms are defined by $\|Y\|_{L_2} = (\mathrm{E}[Y^2])^{1/2}$ and $\|Y\|_{\psi_1} = \inf\{C > 0 \colon \mathrm{E}[\exp(|Y|/C)] \leq 2\}$, respectively.

Proof of Theorem 4.1. Fix $x_0 > 0$. Throughout the proof, we assume that

$$\frac{B_n^2 \log^{10}(pn)}{n} \leq (cx_0)^2, \tag{B.1}$$

where c is a sufficiently small universal constant (whose value will be clear from the proof), since otherwise the asserted claim follows trivially. Also, let K_1 and K_2 be universal constants from Lemma A.4. We assume that $K_2 \geq 1$ since otherwise we can replace K_2 by $K_2 \vee 1$.

Observe that

$$\begin{aligned}\Big\| \max_{1\leq i\leq n} \max_{1\leq j\leq p} |Z_{ij}| \Big\|_{L_2} &\leq 2 \Big\| \max_{1\leq i\leq n} \max_{1\leq j\leq p} |Z_{ij}| \Big\|_{\psi_1} \\ &\leq 8 \log(1+pn) \max_{1\leq i\leq n} \max_{1\leq j\leq p} \|Z_{ij}\|_{\psi_1} \\ &\leq 8 B_n \log(1+pn) \leq 16 B_n \log(pn), \end{aligned} \tag{B.2}$$

where the first and the second inequalities follow from the discussion on p. 95 and from Lem. 2.2.2 in van der Vaart and Wellner (1996), respectively, the third from (26), and the fourth from $p \geq 3$.

Next, let

$$J := \left\{ j = 1, \ldots, p \colon \frac{1}{n} \sum_{i=1}^n \mathrm{E}[Z_{ij}^2] \geq \frac{x_0^2}{(8K_2)^2 \log(pn)} \right\}$$

and $J^c := \{1, \ldots, p\} \setminus J$. Then, by Lemma A.4 and (B.2),

$$\begin{aligned}\mathrm{E}\Big[\max_{j\in J^c} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{ij} \Big| \Big] &\leq K_2 \Big(\frac{x_0 \sqrt{\log p}}{8K_2 \sqrt{\log(pn)}} + \frac{\log p}{\sqrt{n}} \Big\| \max_{1\leq i\leq n} \max_{1\leq j\leq p} |Z_{ij}| \Big\|_{L_2} \Big) \\ &\leq x_0/8 + \frac{16 K_2 B_n \log^2(pn)}{\sqrt{n}} \leq x_0/4, \end{aligned} \tag{B.3}$$

where the last inequality follows from (B.1) if c there is chosen small enough. Also, again by Lemma A.4 and (B.2), for any $t > 0$,

$$\begin{aligned}&\mathrm{P}\Big(\max_{j\in J^c} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{ij} \Big| > 2\mathrm{E}\Big[\max_{j\in J^c} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{ij} \Big| \Big] + t \Big) \\ &\quad \leq \exp\Big(- \frac{t^2 (8K_2)^2 \log(pn)}{3x_0^2} \Big) + 3 \exp\Big(- \frac{t\sqrt{n}}{8K_1 B_n \log(pn)} \Big). \end{aligned}$$

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Combining this inequality applied with $t = x_0/2$ with (B.3), using $K_2 \geq 1$ and (B.1), and assuming that c in (B.1) is small enough gives

$$\mathrm{P}\left(\max_{j\in J^c}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{ij}\right| > x_0\right) \leq 4\exp(-\log(pn)) \leq \left(\frac{4B_n^2\log^{10}(pn)}{n}\right)^{1/6} \tag{B.4}$$

since $p \geq 3$ by assumption and $B_n \geq 1$ by (24) and (25).

Further, for all $i = 1,\dots,n$ and $j = 1,\dots,p$,

$$\begin{aligned}\|Z_{ij}^g\|_{\psi_1} &\leq \sqrt{\log 2}\|Z_{ij}^g\|_{\psi_2} \leq \sqrt{8/3}\sqrt{\log 2}\|Z_{ij}^g\|_{L_2} \leq \sqrt{8/3}\sqrt{\log 2}\|Z_{ij}\|_{L_2}\\ &\leq 2\sqrt{8/3}\sqrt{\log 2}\|Z_{ij}\|_{\psi_1} \leq 2\sqrt{8/3}\sqrt{\log 2}B_n,\end{aligned}$$

where the first and the fourth inequalities follow from the discussion on p. 95 of van der Vaart and Wellner (1996), the second from exercise 1 on p. 105 of van der Vaart and Wellner (1996), the third from the construction of Z_{ij}^g, and the fifth from (26). Hence, using the same argument as above shows that

$$\mathrm{P}\left(\max_{j\in J^c}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{ij}^g\right| > x_0\right) \leq \left(\frac{128B_n^2\log^{10}(pn)}{n}\right)^{1/6}. \tag{B.5}$$

Now, combining (B.4) and (B.5), we conclude that the asserted claim will follow if we show that

$$\sup_{x\geq x_0}\left|\mathrm{P}\left(\max_{j\in J}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{ij} \leq x\right) - \mathrm{P}\left(\max_{j\in J}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{ij}^g \leq x\right)\right| \leq \left(\frac{CB_n^2\log^{10}(pn)}{n}\right)^{1/6},$$

where C is a constant depending only on x_0. In turn, this inequality follows by applying Prop. 2.1 in Chernozhukov et al. (2017) with

$$X_i = \frac{8K_2Z_i\sqrt{\log(pn)}}{x_0}, \quad i = 1,\dots,n, \tag{B.6}$$

and B_n there replaced by

$$B_n\left(1 \vee \frac{8K_2\sqrt{\log(pn)}}{x_0}\right)^3 \tag{B.7}$$

here. This completes the proof of the theorem. □

Proof of Theorem 4.2. Fix $x_0 > 0$. As in the proof of Theorem 4.1, we assume here that (B.1) holds with a sufficiently small universal constant c. Also, let K_3 and K_4 be universal constants from Lemma A.5. We assume that $K_4 \geq K_2 \vee 1$, where K_2 is a universal constant from Lemma A.4, since otherwise we can replace K_4 by $K_4 \vee K_2 \vee 1$. Moreover, let

$$J := \left\{j = 1,\dots,p\colon \frac{1}{n}\sum_{i=1}^n \mathrm{E}[Z_{ij}^2] \geq \frac{x_0^2}{(8K_4)^2\log(pn)}\right\}$$

and $J^c := \{1,\dots,p\}\backslash J$.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Now, observe that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i Z_{ij} \mid (Z_i)_{i=1}^n \sim N\left(0, \frac{1}{n}\sum_{i=1}^n Z_{ij}^2\right), \quad j = 1, \ldots, p.$$

Therefore, by Lem. A.3.1 in Talagrand (2011),

$$\mathrm{E}\left[\max_{j\in J_c}\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i Z_{ij} \mid (Z_i)_{i=1}^n\right] \le \sqrt{2\log p}\left(\max_{j\in J_c}\frac{1}{n}\sum_{i=1}^n Z_{ij}^2\right)^{1/2}, \tag{B.8}$$

and by Thm. 2.1.1 in Adler and Taylor (2007), for any $t > 0$,

$$\begin{aligned}&\mathrm{P}\Big(\max_{j\in J_c}\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i Z_{ij} > \mathrm{E}\Big[\max_{j\in J_c}\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i Z_{ij} \mid (Z_i)_{i=1}^n\Big] + t \mid (Z_i)_{i=1}^n\Big)\\ &\quad\le \exp\Big(-\frac{t^2}{2\max_{j\in J^c}\frac{1}{n}\sum_{i=1}^n Z_{ij}^2}\Big).\end{aligned} \tag{B.9}$$

Applying (B.9) with

$$t = \sqrt{2\log n}\left(\max_{j\in J_c}\frac{1}{n}\sum_{i=1}^n Z_{ij}^2\right)^{1/2}$$

and using (B.8) gives

$$\begin{aligned}&\mathrm{P}\Big(\max_{j\in J_c}\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i Z_{ij} > 2\sqrt{2\log(pn)}\Big(\max_{j\in J^c}\frac{1}{n}\sum_{i=1}^n Z_{ij}^2\Big)^{1/2} \mid (Z_i)_{i=1}^n\Big)\\ &\quad\le \frac{1}{n} \le \Big(\frac{B_n^2\log^{10}(pn)}{n}\Big)^{1/6}\end{aligned} \tag{B.10}$$

since $p \ge 3$ by assumption and $B_n \ge 1$ by (24) and (25). Next, by Lemma A.5,

$$\begin{aligned}\mathrm{E}\Big[\max_{j\in J^c}\frac{1}{n}\sum_{i=1}^n Z_{ij}^2\Big] &\le K_4\Big(\frac{x_0^2}{(8K_4)^2\log(pn)} + \frac{\log p}{n}\Big\|\max_{1\le i\le n}\max_{1\le j\le p}|Z_{ij}|\Big\|_{L_2}^2\Big)\\ &\le \frac{x_0^2}{64\log(pn)} + \frac{256K_4B_n^2\log^3(pn)}{n} \le \frac{x_0^2}{32\log(pn)},\end{aligned} \tag{B.11}$$

where the second inequality follows from (B.2) in the proof of Theorem 4.1 and $K_4 \ge 1$, and the third from (B.1) with c being chosen small enough. Also, again by Lemma A.5, for any $t > 0$,

$$\begin{aligned}&\mathrm{P}\Big(\max_{j\in J^c}\frac{1}{n}\sum_{i=1}^n Z_{ij}^2 > 2\mathrm{E}\Big[\max_{j\in J^c}\frac{1}{n}\sum_{i=1}^n Z_{ij}^2\Big] + t\Big)\\ &\quad\le 3\exp\Big(-\frac{\sqrt{nt}}{K_3\|\max_{1\le i\le n}\max_{1\le j\le p}|Z_{ij}|\|_{\psi_1}}\Big) \le 3\exp\Big(-\frac{\sqrt{nt}}{8K_3B_n\log(pn)}\Big),\end{aligned}$$

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

where the second inequality follows from (B.2) in the proof of Theorem 4.1. Applying this inequality with

$$t = \frac{x_0^2}{16\log(pn)}$$

and using (B.11) shows that

$$\mathrm{P}\left(\max_{j\in J^c}\frac{1}{n}\sum_{i=1}^n Z_{ij}^2 > \frac{x_0^2}{8\log(pn)}\right) \le 3\exp(-\log(pn)) \le 3/(pn) \tag{B.12}$$

since (B.1) holds with a sufficiently small constant c. Now, combining (B.10) and (B.12) gives

$$\mathrm{P}\left(\mathrm{P}\left(\max_{j\in J^c}\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i Z_{ij} > x_0 \mid (Z_i)_{i=1}^n\right) \le \left(\frac{B_n^2\log^{10}(pn)}{n}\right)^{1/6}\right) \ge 1 - \frac{3}{pn}. \tag{B.13}$$

Further, applying Cor. 4.2 in Chernozhukov et al. (2017) with X_i's defined in (B.6) and B_n there replaced by the expression in (B.7), with K_4 instead of K_2, here yields

$$\begin{aligned}&\mathrm{P}\Big(\sup_{x\in\mathbb{R}}\Big|\mathrm{P}\Big(\max_{j\in J}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{ij}^g \le x\Big) - \mathrm{P}\Big(\max_{j\in J}\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i Z_{ij} \le x \mid (Z_i)_{i=1}^n\Big)\Big| \\ &\quad\le \Big(\frac{CB_n^2\log^{10}(pn)}{n}\Big)^{1/6}\Big) \ge 1 - \frac{1}{pn},\end{aligned} \tag{B.14}$$

where C is a constant depending only on x_0. Combining (B.13) and (B.14) with (B.5) in the proof of Theorem 4.1 gives the asserted claim and completes the proof of the theorem. □

Proof of Theorem 4.3. For $\alpha \in (0,1)$, let $c_n^g(\alpha)$ denote the $(1-\alpha)$th quantile of S_n^g and $c_n^e(\alpha)$ denote the $(1-\alpha)$th quantile of the conditional distribution of S_n^e given $(Z_i)_{i=1}^n$. Also, for $\alpha \le 0$, define $c_n^g(\alpha) := \infty$ and $c_n^e(\alpha) := \infty$. Fix $\alpha_0 \in (0, 1/2)$ and denote $x_0 := c_n^g(\alpha_0) > 0$. Throughout the proof, we will use C to denote a constant that depends only on α_0 but whose value can change from place to place. Finally, we will assume, without loss of generality, that $(e_i)_{i=1}^n$ is independent of $(Z_i, \widehat{Z}_i)_{i=1}^n$.

The asserted claim can be equivalently written as the following two inequalities:

$$\mathrm{P}(S_n > c_n(\alpha)) \ge \alpha - C\left(\left(\frac{B_n^2\log^{10}(pn)}{n}\right)^{1/6} + \zeta_n\log^{3/2}(pn)\right) - \kappa_n, \quad \alpha \in (0, \alpha_0), \tag{B.15}$$

$$\mathrm{P}(S_n > c_n(\alpha)) \le \alpha + C\left(\left(\frac{B_n^2\log^{10}(pn)}{n}\right)^{1/6} + \zeta_n\log^{3/2}(pn)\right) + \kappa_n, \quad \alpha \in (0, \alpha_0). \tag{B.16}$$

Both inequalities follow from the same arguments. Therefore, we only prove (B.15). We proceed in four steps.

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Step 1. Here, we prove that for all $\alpha \in (0, \alpha_0)$ and $\epsilon > 0$,

$$c_n^g(\alpha) + \epsilon \leq c_n^g\left(\alpha - \left(\frac{CB_n^2 \log^{10}(pn)}{n}\right)^{1/6} - C\epsilon \log(pn)\right).$$

To do so, define J and J^c as in the proof of Theorem 4.1. Then for any $x \geq x_0$,

$$\begin{aligned}
&\mathrm{P}(S_n^g \leq x+\epsilon) - \mathrm{P}(S_n^g \leq x) = \mathrm{P}(x < S_n^g \leq x+\epsilon) \\
&\quad \leq \mathrm{P}\Big(\max_{j \in J^c} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{ij}^g > x_0\Big) + \mathrm{P}\Big(x < \max_{j \in J} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{ij}^g \leq x+\epsilon\Big) \\
&\quad \leq \Big(\frac{CB_n^2 \log^{10}(pn)}{n}\Big)^{1/6} + C\epsilon \log(pn),
\end{aligned}$$

where the second line follows from the union bound, and the third from (B.5) in the proof of Theorem 4.1 and Lemma A.6. Applying this inequality with $x = c_n^g(\alpha)$ gives the asserted claim of this step.

Step 2. Here, we prove that with probability at least $1 - 4/(pn)$, for all $\alpha \in (0, \alpha_0)$,

$$c_n^e(\alpha) \leq c_n^g\left(\alpha - \left(\frac{CB_n^2 \log^{10}(pn)}{n}\right)^{1/6}\right).$$

To do so, note that by Theorem 4.2, the event in (28) holds with probability at least $1 - 4/(pn)$. On this event, for any $\phi > 0$,

$$\begin{aligned}
\mathrm{P}(S_n^e \leq c_n^g(\alpha - \phi) \mid (Z_i)_{i=1}^n) &\geq \mathrm{P}(S_n^g \leq c_n^g(\alpha - \phi)) - \Big(\frac{CB_n^2 \log^{10}(pn)}{n}\Big)^{1/6} \\
&= 1 - \alpha + \phi - \Big(\frac{CB_n^2 \log^{10}(pn)}{n}\Big)^{1/6}.
\end{aligned}$$

Applying this inequality with

$$\phi = \Big(\frac{CB_n^2 \log^{10}(pn)}{n}\Big)^{1/6}$$

gives the asserted claim of this step.

Step 3. Here, we prove that with probability at least $1 - \kappa_n$, for all $\alpha \in (0, \alpha_0)$,

$$c_n(\alpha) \leq c_n^e(\alpha - 2/n) + 2\zeta_n \sqrt{2 \log(pn)}.$$

To do so, note that by (29),

$$\max_{1 \leq j \leq p} \frac{1}{n} \sum_{i=1}^n (\widehat{Z}_{ij} - Z_{ij})^2 \leq \zeta_n^2 \tag{B.17}$$

with probability at least $1 - \kappa_n$. Also,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n e_i(\widehat{Z}_{ij} - Z_{ij}) \mid (Z_i, \widehat{Z}_i)_{i=1}^n \sim N\left(0, \frac{1}{n} \sum_{i=1}^n (\widehat{Z}_{ij} - Z_{ij})^2\right), \quad j = 1, \ldots, p.$$

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

Therefore, on the event in (B.17), by the same arguments as those used in the proof of Theorem 4.2,

$$\mathrm{P}\left(\max_{1\le j\le p}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i(\widehat{Z}_{ij}-Z_{ij})>2\zeta_n\sqrt{2\log(pn)}\mid (Z_i,\widehat{Z}_i)_{i=1}^n\right)\le\frac{1}{n}$$

and, similarly,

$$\mathrm{P}\left(\max_{1\le j\le p}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i(Z_{ij}-\widehat{Z}_{ij})>2\zeta_n\sqrt{2\log(pn)}\mid (Z_i,\widehat{Z}_i)_{i=1}^n\right)\le\frac{1}{n}.$$

Thus, by the union bound, with probability at least $1-\kappa_n$,

$$\mathrm{P}\left(\max_{1\le j\le p}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i(\widehat{Z}_{ij}-Z_{ij})\right|>2\zeta_n\sqrt{2\log(pn)}\mid (Z_i,\widehat{Z}_i)_{i=1}^n\right)\le\frac{2}{n}.$$

Combining this bound with the inequality

$$|\widehat{S}_n^e-S_n^e|\le\max_{1\le j\le p}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i(\widehat{Z}_{ij}-Z_{ij})\right|,$$

gives the asserted claim of this step.

Step 4. Here, we complete the proof of the theorem. To do so, fix $\alpha\in(0,\alpha_0)$ and note that as in (B.10) of the proof of Theorem 4.2, we have $1/n\le(B_n^2\log^{10}(pn)/n)^{1/6}$. Therefore, by Steps 1, 2, and 3, with probability at least $1-\kappa_n$,

$$c_n(\alpha)\le c_n^g\left(\alpha-\left(\frac{CB_n^2\log^{10}(pn)}{n}\right)^{1/6}-C\zeta_n\log^{3/2}(pn)\right).$$

Hence,

$$\mathrm{P}(S_n>c_n(\alpha))\ge\mathrm{P}\left(S_n>c_n^g\left(\alpha-\left(\frac{CB_n^2\log^{10}(pn)}{n}\right)^{1/6}-C\zeta_n\log^{3/2}(pn)\right)\right)-\kappa_n.$$

In turn, by Theorem 4.1, the probability on the right-hand side of this inequality is bounded from below by

$$\begin{aligned}&\mathrm{P}\left(S_n^g>c_n^g\left(\alpha-\left(\frac{CB_n^2\log^{10}(pn)}{n}\right)^{1/6}-C\zeta_n\log^{3/2}(pn)\right)\right)-\left(\frac{CB_n^2\log^{10}(pn)}{n}\right)^{1/6}\\&\quad=\alpha-\left(\frac{CB_n^2\log^{10}(pn)}{n}\right)^{1/6}-C\zeta_n\log^{3/2}(pn).\end{aligned}$$

Combining the last two inequalities gives (B.15) and completes the proof of the theorem. □

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press

C TECHNICAL LEMMAS

LEMMA C.1. *Let $X_1,\dots,X_n$ be independent random vectors in $\mathbb{R}^p$ with $p \geq 2$ such that $\mathrm{E}[X_{ij}] = 0$ for all $i = 1,\dots,n$ and $j = 1,\dots,p$. Define $M := \max_{1\leq i\leq n}\max_{1\leq j\leq p}|X_{ij}|$ and $\sigma^2 := \max_{1\leq j\leq p}\sum_{i=1}^n \mathrm{E}[Z_{ij}^2]$. Then for any $t > 0$,*

$$\mathrm{P}\left(\max_{1\leq j\leq p}\left|\sum_{i=1}^n X_{ij}\right| \geq 2\mathrm{E}\left[\max_{1\leq j\leq p}\left|\sum_{i=1}^n X_{ij}\right|\right] + t\right) \tag{C.1}$$

$$\leq \exp\left(-\frac{t^2}{3\sigma^2}\right) + 3\exp\left(-\frac{t}{K_1\|M\|_{\psi_1}}\right), \tag{C.2}$$

where K_1 is a universal constant and $\|M\|_{\psi_1} := \inf\{C > 0\colon \mathrm{E}[\exp(M/C)] \leq 2\}$. In addition,

$$\mathrm{E}\left[\max_{1\leq j\leq p}\left|\sum_{i=1}^n X_{ij}\right|\right] \leq K_2\left(\sigma\sqrt{\log p} + \sqrt{\mathrm{E}[M^2]}\log p\right), \tag{C.3}$$

where K_2 is a universal constant.

Proof of Lemma C.1. See Lem. E.1 and E.2 in Chernozhukov et al. (2017). □

LEMMA A.5. *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 $t > 0$,*

$$\mathrm{P}\left(\max_{1\leq j\leq p}\sum_{i=1}^n X_{ij} \geq 2\mathrm{E}\Big[\max_{1\leq j\leq p}\sum_{i=1}^n X_{ij}\Big] + t\right) \leq 3\exp\left(-\frac{\sqrt{t}}{K_3\|\sqrt{M}\|_{\psi_1}}\right) \tag{C.4}$$

where K_3 is a universal constant and $\|\sqrt{M}\|_{\psi_1} := \inf\{C > 0\colon \mathrm{E}[\exp(\sqrt{M}/C)] \leq 2\}$. In addition,

$$\mathrm{E}\Big[\max_{1\leq j\leq p}\sum_{i=1}^n X_{ij}\Big] \leq K_4\Big(\max_{1\leq j\leq p}\mathrm{E}\Big[\sum_{i=1}^n X_{ij}\Big] + \mathrm{E}[M]\log p\Big), \tag{C.5}$$

where K_4 is a universal constant.

Proof of Lemma C.2. See Lem. E.4 in Chernozhukov et al. (2017) and Lemma 9 in Chernozhukov et al. (2015) for the proof of (C.4) and (C.5), respectively. □

LEMMA A.6. *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*

$$\mathrm{P}(Z \leq x + \epsilon) - \mathrm{P}(Z \leq x) \leq \frac{\epsilon}{\underline{\sigma}}(\sqrt{2\log p} + 2), \tag{C.6}$$

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

Proof of Lemma C.3. See Lem. A.1 in Chernozhukov et al. (2017). □

https://doi.org/10.1017/S0266466620000225 Published online by Cambridge University Press