ADAPTIVE TESTING ON A REGRESSION FUNCTION AT A POINT

By

Timothy B. Armstrong

August 2014
Revised February 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 1957R

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Adaptive Testing on a Regression Function at a Point

Timothy B. Armstrong*
Yale University

February 25, 2015

Abstract

We consider the problem of inference on a regression function at a point when the entire function satisfies a sign or shape restriction under the null. We propose a test that achieves the optimal minimax rate adaptively over a range of Hölder classes, up to a $\log\log n$ term, which we show to be necessary for adaptation. We apply the results to adaptive one-sided tests for the regression discontinuity parameter under a monotonicity restriction, the value of a monotone regression function at the boundary, and the proportion of true null hypotheses in a multiple testing problem.

1 Introduction

We consider a Gaussian regression model with random design. We observe $\{(X_i, Y_i)\}_{i=1}^n$ where X_i and Y_i are real valued random variables with (X_i, Y_i) iid and

$$Y_i = g(X_i) + \varepsilon_i, \quad \varepsilon_i | X_i \sim N(0, \sigma^2(X_i)), \quad X_i \sim F_X, \tag{1}$$

where F_X denotes the cdf of X_i. We are interested in hypothesis tests about the regression function g at a point, which we normalize to be zero. We impose regularity conditions on the conditional variance of Y_i and the distribution of X_i near this point: for some $\eta > 0$,

$$\eta t \leq |F_X(t) - F_X(-t)| \leq t/\eta, \;\; \eta \leq \sigma^2(x) \leq 1/\eta \text{ for } |x| < \eta,\, 0 < t < \eta. \tag{2}$$

*email: timothy.armstrong@yale.edu. Thanks to Alexandre Tsybakov, Matias Cattaneo, Yuichi Kitamura and Tony Cai for helpful discussions, and to the Associate Editor and two anonymous referees for helpful comments.

Note that this allows (but does not impose) that our point of interest, 0 may be on the boundary of the support of X_i.

We consider the null hypotheses

$$H_0 : \{g|g(x) = 0 \text{ all } x \in \text{supp}(X_1)\} \tag{3}$$

$$H_0 : \{g|g(x) \leq 0 \text{ all } x \in \text{supp}(X_1)\}, \tag{4}$$

where $\text{supp}(X_1)$ denotes the support of the distribution F_X, and the alternative $H_1 : \{g|g(0) \geq b, g \in \mathcal{F}\}$, where $\mathcal{F}$ imposes smoothness conditions on g. In particular, we consider Hölder classes of functions with exponent $\beta \leq 1$:

$$\mathcal{F} = \Sigma(\beta, L) \equiv \{g||g(x) - g(x')| \leq L|x - x'|^\beta \text{ all } x,\, x'\}$$

where $L > 0$ and $0 \leq \beta \leq 1$, so that the alternative is given by

$$H_1 : g \in \mathcal{G}(b, L, \beta) \equiv \{g|g(0) \geq b \text{ and } g \in \Sigma(L, \beta)\}.$$

The focus on $g(0)$ is a normalization in the sense that the results apply to inference on $g(x_0)$ for any point x_0 by redefining X_i to be $X_i - x_0$, so long as the point of interest x_0 is known. We also consider cases where certain shape restrictions are imposed under the null and alternative.

For simplicity, we treat the distribution F_X of X_i and the conditional variance function σ^2 as fixed and known under the null and alternative. Thus, we index probability statements with the function g, which determines the joint distribution of $\{(X_i, Y_i)\}_{i=1}^n$. We note, however, that the tests considered here can be extended to achieve the same rates without knowledge of these functions, so long as an upper bound for $\sup_x \sigma^2(x)$ is known or can be estimated.

It is known (see Lepski and Tsybakov, 2000) that the optimal rate for testing the null hypothesis (3) or (4) against the alternative H_1 when g is known to be in the Hölder class $\Sigma(L, \beta)$ is $n^{-\beta/(2\beta+1)}$. That is, for any $\varepsilon > 0$, there exists a constant C_* such that, for any $\alpha \in (0, 1)$ and sequence of tests ϕ_n with level α under the null hypothesis (3),

$$\limsup_n \inf_{g \in \mathcal{G}(C_* n^{-\beta/(2\beta+1)}, L, \beta)} E_g \phi_n \leq \alpha + \varepsilon.$$

Furthermore, using knowledge of β, one can construct a sequence of tests ϕ_n^* that are level

α for the null hypothesis (4) (and, therefore, also level α for the null hypothesis (3)) such that, for any $\varepsilon > 0$, there exists a C^* such that

$$\liminf_n \inf_{g \in \mathcal{G}(C^* n^{-\beta/(2\beta+1)}, L, \beta)} E_g \phi_n^* \geq 1 - \varepsilon. \tag{5}$$

We ask whether a single test ϕ_n can achieve the rate in (5) simultaneously for all $\beta \leq 1$. Such a test would be called adaptive with respect to β. We find that the answer is no, but that adaptivity can be obtained when the rate is modified by a $\log\log n$ term, which we show is the necessary rate for adaptation. In particular, we show that, for C_* small enough, any sequence ϕ_n of level α tests of (3) must have asymptotically trivial power for some β in the class $\mathcal{G}(C_*(n/\log\log n)^{-\beta/(2\beta+1)}, L, \beta)$ in the sense that, for any $\underline{\beta} < \overline{\beta} \leq 1$,

$$\limsup_n \inf_{\beta \in [\underline{\beta}, \overline{\beta}]} \inf_{\mathcal{G}(C_*(n/\log\log n)^{-\beta/(2\beta+1)}, L, \beta)} E_g \phi_n \leq \alpha.$$

Furthermore, we exhibit a sequence of tests ϕ_n^* that achieve asymptotic power 1 adaptively over the classes $\mathcal{G}(C^*(n/\log\log n)^{-\beta/(2\beta+1)}), L, \beta)$ for C^* large enough, while being level α for the null hypothesis (4): for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \inf_{\beta \in [\varepsilon, 1]} \inf_{\mathcal{G}(C^*(n/\log\log n)^{-\beta/(2\beta+1)}, L, \beta)} E_g \phi_n^* = 1.$$

Our interest in testing at a point stems from several problems in statistics and econometrics in which a parameter is given by the value of a regression or density function at the boundary, and where the function can plausibly be assumed to satisfy a monotonicity restriction. This setup includes the regression discontinuity model and inference on parameters that are "identified at infinity," both of which have received considerable attention in the econometrics literature (see, among others, Chamberlain, 1986; Heckman, 1990; Andrews and Schafgans, 1998; Hahn, Todd, and Van der Klaauw, 2001). In the closely related problem where g is a density rather than a regression function, our setup covers the problem of inference on the proportion of null hypotheses when testing many hypotheses (see Storey, 2002). We discuss these applications in Section 3. The results in this paper can be used to obtain adaptive one-sided confidence intervals for these parameters, and to show that they achieve the minimax adaptive rate.

The literature on asymptotic minimax bounds in nonparametric testing has considered many problems closely related to the ones considered here, and our results draw heavily from this literature. Here, we name only a few, and refer to Ingster and Suslina (2003),

for a more thorough exposition of the literature. Typically, the goal in this literature is to derive bounds in problems similar to the one considered here, but with the alternative given by $\{\varphi(g) \geq b\} \cap \mathcal{F}$, where $\varphi(g)$ is some function measuring distance from the null and $\mathcal{F}$ a class of functions imposing smoothness on g. Our problem corresponds to the case where $\varphi(g) = g(0)$ and $\mathcal{F} = \Sigma(L, \beta)$, where we focus on adaptivity with respect to $\beta \leq 1$. Lepski and Tsybakov (2000) consider this problem for fixed (L, β), and also consider the case where $\varphi(g)$ is the ℓ_∞ norm. Pouet (2000) considers $\varphi(g) = g(0)$ with $\mathcal{F}$ given by a class of analytic functions satisfying certain restrictions. Dümbgen and Spokoiny (2001) consider the ℓ_∞ norm and adaptivity over Hölder classes with respect to (L, β) and find, in contrast to our case, that adaptivity can be achieved without a loss in the minimax rate (or, for adaptivity over L, even the constant). In these papers, the optimal constants C^* and C_* are also derived in some cases. Spokoiny (1996) considers adaptivity to Besov classes under the ℓ_2 norm and shows that, as we derive in our case, the minimax rate can be obtained adaptively only up to an additional $\log \log n$ term. It should also be noted that the tests we use to achieve the minimax adaptive rate bear a close resemblence to tests used in other adaptive testing problems (see, e.g., Fan, 1996; Donoho and Jin, 2004, as well as some of the papers cited above).

Our results can be used to obtain one-sided confidence intervals for a monotone function at the boundary of its support, which complements results in the literature on adaptive confidence intervals for shape restricted densities. Low (1997) shows that adaptive confidence intervals cannot be obtained without shape restrictions on the function. Cai and Low (2004) develop a general theory of adaptive confidence intervals under shape restrictions. Cai, Low, and Xia (2013) consider adaptive confidence intervals for points on the interior of the support of a shape restricted density and show that, in contrast to our case, the adaptive rate can be achieved with no additional $\log \log n$ term. Dümbgen (2003) considers the related problem of adaptive confidence bands for the entire function. Our interest in points on the boundary stems from the specific applications considered in Section 3.

2 Results

We first state the lower bound for minimax adaptation. All proofs are in Section 4. For the purposes of some of the applications, we prove a slightly stronger result in which g may be known to be nonincreasing in $|x|$. Let $\mathcal{G}_{|x|\downarrow}$ be the class of functions that are nondecreasing on $(-\infty, 0]$ and nonincreasing on $[0, \infty)$.

Theorem 2.1. *Let $0 < \underline{\beta} < \overline{\beta} \leq 1$ be given. There exists a constant C_* depending only on $\underline{\beta}, \overline{\beta}$, L and the bounds on F_X and σ such that the following holds. Let ϕ_n be any sequence of tests taking the data $\{(X_i, Y_i)\}_{i=1}^n$ to a rejection probability in $[0,1]$ with asymptotic level α for the null hypothesis (3): $\limsup_n E_0\phi_n \leq \alpha$. Then*

$$\limsup_n \inf_{\beta \in [\underline{\beta}, \overline{\beta}]} \inf_{\mathcal{G}(C_*(n/\log\log n)^{-\beta/(2\beta+1)}, L, \beta) \cap \mathcal{G}_{|x|\downarrow}} E_g\phi_n \leq \alpha.$$

Note that the results of the theorem imply the same results when the requirement that $g \in \mathcal{G}_{|x|\downarrow}$ is removed from the alternative, or when the null is replaced by (4) with the possible requirement $g \in \mathcal{G}_{|x|\downarrow}$.

We now construct a test that achieves the $(n/\log\log n)^{\beta/(2\beta+1)}$ rate. For $k \in \{1, \ldots, n\}$, let $\hat{g}_k$ be the k-nearest neighbor estimator of $g(0)$, given by

$$\hat{g}_k = \frac{1}{k} \sum_{|X_j| \leq |X_{(k)}|} Y_j \qquad \text{where } |X_{(k)}| \text{ is the } k\text{th least value of } |X_i| \tag{6}$$

for $|X_{(k)}| < \eta$, and $\hat{g}_k = 0$ otherwise, where η is given in (2). Let

$$T_n = \max_{1 \leq k \leq n} \sqrt{k}\hat{g}_k$$

and let $c_{\alpha,n}$ be the $1-\alpha$ quantile of T_n under $g(x) = 0$ all x. Note that, by the law of the iterated logarithm (applied to the $N(0,1)$ variables $Y_i/\sigma(X_i)$ conditional on the X_i's), $\limsup_n c_{\alpha,n}/\sqrt{\log\log n} \leq \sqrt{2}\sup_x \sigma(x)$. Let ϕ_n^* be the test that rejects when $T_n > c_{\alpha,n}$.

Theorem 2.2. *The test ϕ_n^* given above has level α for the null hypothesis (4). Furthermore, there exists a constant C^* such that, for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} \inf_{\beta \in [\varepsilon, 1]} \inf_{\mathcal{G}(C^*(n/\log\log n)^{-\beta/(2\beta+1)}, L, \beta)} E_g\phi_n^* = 1.$$

An interesting question is the derivation of the sharp constant in the adaptive rate. While we leave this question for future research, we briefly discuss some conjectures. We conjecture that, under additional regularity conditions on the conditional variance $\sigma(x)$ and distribution of the covariate X_i, a sharp constant $C(\beta, L)$ exists such that, for arbitrary $\delta > 0$, Theorem 2.1 holds with C_* replaced by $(1-\delta)C(\beta, L)$, and Theorem 2.2 holds with C^* replaced by $(1+\delta)C(\beta, L)$ and ϕ_n^* replaced by a different test. A reasonable candidate for a test statistic to achieve the optimal constant would be a supremum over β of normalized

estimates based on the optimal kernel given in Example 1 of Lepski and Tsybakov (2000), with the bandwidth calibrated appropriately for each β. The conjectured behavior where minimax adaptive power goes to α or one on either side of a constant, where the constant does not depend on the size α of the test, would be an instance of asymptotic degeneracy related to the phenomenon observed for the ℓ_∞ case by Lepski and Tsybakov (2000) (in the nonadaptive setting) and Dümbgen and Spokoiny (2001) (for adaptivity with respect to L), and our conjecture is based partly on the fact that the tests and approximately least favorable distributions over alternatives used in our results have a similar structure to those used in the above papers.

3 Applications and Extensions

3.1 Inference on a Monotone Function at the Boundary

We note that, in the case where 0 is on the boundary of the support of X_i, the results in the previous section give the optimal rate for a one sided test concerning $g(0)$ under a monotonicity restriction on g. This can be used to obtain adaptive (up to a $\log\log n$ term) one-sided confidence intervals for a regression function at the boundary, where the $\log\log n$ term is necessary for adaptation. This can be contrasted to the construction of adaptive confidence regions for a monotone function on the interior of its support, in which case the $\log\log n$ term is not needed (cf. Cai, Low, and Xia, 2013).

To form a confidence interval based on our test, we define $T_n(\theta_0) = \max_{1\le k\le n} \sqrt{k}\,(\hat{g}_k - \theta_0)$, and form our confidence interval by inverting tests of $H_0 : g(0) \le \theta_0$ based on $T_n(\theta_0)$ with critical value $c_{\alpha,n}$ given above (the $1-\alpha$ quantile under $g = 0$ and $\theta_0 = 0$). The confidence interval is then given by $[\hat{c}^*, \infty)$ where $\hat{c}^* = \max_{1\le k\le \overline{k}}[\hat{g}_k - c_{\alpha,n}/\sqrt{k}]$, with $\overline{k}$ the largest value of k such that $|X_{(k)}| < \eta$. The following corollary to Theorems 2.1 and 2.2 shows that this CI achieves the adaptive rate.

Corollary 3.1. *Let $0 < \underline{\beta} < \overline{\beta} \le 1$ be given. There exists a constant C_* depending only on $\underline{\beta}, \overline{\beta}$, L and the bounds on F_X and σ such that the following holds. Let $[\hat{c}, \infty)$ be any sequence of one sided CIs with asymptotic coverage $1-\alpha$ for $g(0)$ when $g \in \mathcal{G}_{|x|\downarrow}$: $\liminf_n \inf_{g\in\mathcal{G}_{|x|\downarrow}} P_g(g(0) \in [\hat{c}, \infty)) \ge 1-\alpha$. Then*

$$\limsup_n \inf_{\beta\in[\underline{\beta},\overline{\beta}]} \inf_{g\in\Sigma(\beta,L)\cap\mathcal{G}_{|x|\downarrow}} P_g\left(\hat{c} > g(0) - C_*(n/\log\log n)^{-\beta/(2\beta+1)}\right) \le \alpha.$$

Furthermore, the CI $[\hat{c}^*, \infty)$ *given above has coverage at least* $1-\alpha$ *for* $g \in \mathcal{G}_{|x|\downarrow}$*, and there exists a* C^* *such that, for all* $\varepsilon > 0$*,*

$$\lim_{n\to\infty} \inf_{\beta\in[\varepsilon,1]} \inf_{g\in\Sigma(\beta,L)\cap\mathcal{G}_{|x|\downarrow}} P_g\left(\hat{c}^* > g(0) - C^*(n/\log\log n)^{-\beta/(2\beta+1)}\right) = 1.$$

The problem of inference on a regression function at the boundary has received considerable attention in the econometrics literature, where the problem is often termed "identification at infinity" (see, among others, Chamberlain, 1986; Heckman, 1990; Andrews and Schafgans, 1998; Khan and Tamer, 2010). In such cases, it may not be plausible to assume that the density of X_i is bounded away from zero or infinity near its boundary, and the boundary may not be finite (in which case we are interested in, e.g. $\lim_{x\to-\infty} g(x)$). Such cases require relaxing the conditions on F_X in (2), which can be done by placing conditions on the behavior of $u \mapsto g(F_X^{-1}(u))$. In the interest of space, however, we do not pursue this extension.

3.2 Regression Discontinuity

Consider the regression discontinuity model

$$Y_i = m(X_i) + \tau I(X_i > 0) + \varepsilon_i, \quad \varepsilon_i | X_i \sim N(0, \sigma^2(X_i)), \quad X_i \sim F_X.$$

Here, we strengthen (2) by requiring that there exists some $\eta > 0$ such that, for all $|x| < \eta$ and $0 < t < \eta$, the inequalities $\eta t \leq F_X(t) - F_X(0) \leq t/\eta$, $\eta t \leq F_X(0) - F(-t) \leq t/\eta$, and $\eta \leq \sigma^2(x) \leq 1/\eta$ are satisfied. The regression discontinuity model has been used in a large number of studies in empirical economics in the last decade, and has received considerable attention in the econometrics literature (see Imbens and Lemieux, 2008 for a review of some of this literature).

We are interested in inference on the parameter τ. Of course, τ is not identified without constraints on $m(X_i)$. We impose a monotonicity constraint on m and ask whether a one sided test for τ can be constructed that is adaptive to the Hölder exponent β of the unknown class $\Sigma(L, \beta)$ containing m. In particular, we fix τ_0 and consider the null hypothesis

$$H_0 : \tau \leq \tau_0 \text{ and } m \text{ nonincreasing} \tag{7}$$

and the alternative

$$H_1 : (m, \tau) \in \mathcal{G}^{\mathrm{rd}}(b, L, \beta) \equiv \{(m, \tau) | \tau \geq \tau_0 + b \text{ and } m \in \Sigma(L, \beta) \text{ nonincreasing}\}.$$

We extend the test of Section 2 to a test that is level α under H_0 and consistent against H_1 when $b = b_n$ is given by a $\log\log n$ term times the fastest possible rate simultaneously over $\beta \in [\varepsilon, 1]$, and we show that the $\log\log n$ term is necessary for adaptation.

To describe the test, let $\{(X_{i,1}, Y_{i,1})\}_{i=1}^{n_1}$ be the observations with $X_i \leq 0$ and let $\{(X_{i,2}, Y_{i,2})\}_{i=1}^{n_2}$ be the observations with $X_i > 0$. Let $\hat{g}_{1,k}$ be the k-nearest neighbor estimator given in (6) applied to the sample with $X_i \leq 0$ and let $\hat{g}_{2,k}$ be defined analogously for the sample with $X_i > 0$. Let

$$T_n^{\mathrm{rd}}(\tau) = \max_{1 \leq k \leq n} \sqrt{k}(\hat{g}_{2,k} - \hat{g}_{1,k} - \tau).$$

Let $c_{n,\alpha}^{\mathrm{rd}}$ be the $1-\alpha$ quantile of $T_n^{\mathrm{rd}}(0)$ when $m(x) = 0$ all x and $\tau = 0$. The test $\phi_{n,\tau_0}^{\mathrm{rd}}$ rejects when $T_n^{\mathrm{rd}}(\tau_0) > c_{n,\alpha}^{\mathrm{rd}}$.

The following corollary to Theorems 2.1 and 2.2 gives the optimal rate for adaptive testing in the regression discontinuity problem, and shows that the test $\phi_{n,\tau_0}^{\mathrm{rd}}$ achieves it. Let $E_{m,\tau}$ denote expectation under (m, τ).

Corollary 3.2. *Let $0 < \underline{\beta} < \overline{\beta} \leq 1$ be given. There exists a constant C_* depending only on $\underline{\beta}, \overline{\beta}$, L and the bounds on F_X and σ such that the following holds. Let ϕ_n be any sequence of tests taking the data $\{(X_i, Y_i)\}_{i=1}^n$ to a rejection probability in $[0, 1]$ with asymptotic level α for the null hypothesis (7): $\limsup_n E_0\phi_n \leq \alpha$. Then*

$$\limsup_n \inf_{\beta \in [\underline{\beta}, \overline{\beta}]} \inf_{(m,\tau) \in \mathcal{G}^{rd}(C_*(n/\log\log n)^{-\beta/(2\beta+1)}, L, \beta)} E_{m,\tau}\phi_n \leq \alpha.$$

Furthermore, the test ϕ_{n,τ_0}^{rd} given above has level α for the null hypothesis (4), and there exists a constant C^* such that, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \inf_{\beta \in [\varepsilon, 1]} \inf_{(m,\tau) \in \mathcal{G}^{rd}(C^*(n/\log\log n)^{-\beta/(2\beta+1)}, L, \beta)} E_{m,\tau}\phi_{n,\tau_0}^{rd} = 1.$$

3.3 Inference on the Proportion of True Null Hypotheses

Motivated by an application to large scale multiple testing, we now consider a related setting in which we are interested in nonparametric testing about a density, rather than a regression

function. We observe p-values $\{\hat{p}_i\}_{i=1}^n$ from n independent experiments. The p-values follow the mixture distribution

$$\hat{p}_i \sim f_p(x) = \pi \cdot I(x \in [0,1]) + (1-\pi) \cdot f_1(x) \tag{8}$$

where f_1 is an unknown density on $[0,1]$ and π is the proportion of true null hypotheses. We are interested in tests and confidence regions for π, following a large literature on estimation and inference on π in this setting (see, among others, Storey, 2002, Donoho and Jin, 2004, Meinshausen and Rice, 2006, Cai, Jin, and Low, 2007 and additional references in Efron, 2012).

Given observations from the density $f_p(x)$ with $f_1(x)$ completely unspecified, the best bounds that can be obtained for π in the population are $\pi \in [0, \overline{\pi}]$, where $\overline{\pi} = \overline{\pi}(f_p) \equiv \inf_{x\in[0,1]} f_p(x)$. If the infimum is known to be taken at a particular location x_0, we can test the null hypothesis that $\overline{\pi} \geq \pi_0$ against the alternative $\overline{\pi} < \pi_0$ by testing the null

$$H_0 : f_p(x) \geq \pi_0 \text{ all } x \tag{9}$$

against the alternative $f_p(x_0) < \pi_0$. In other words, we are interested in a version of the problem considered in Section 2, with the regression function g replaced by a density function f_p. Inverting these tests over π_0, we can obtain an upper confidence interval for $\bar{\pi}$. Note that, since the null hypothesis $\overline{\pi}(f_p) \geq \pi_0$ is equivalent to the statement that there exists a $\pi \geq \pi_0$ such that f_p follows the model (8) for some f_1, this can also be considered a test of the null $\pi \geq \pi_0$, and the CI can be considered a CI for π.

Assuming the p-values tend to be smaller when taken from the alternative hypothesis, we can expect that $f_1(x)$ is minimized at $x = 1$ so that $f_p(x)$ will also be minimized at 1. Following this logic, Storey (2002) proposes a uniform kernel density estimator of $f_p(1)$, which can be considered an estimator of $\overline{\pi}$ or of π itself (in the latter case, the estimator provides an asymptotic upper bound, but is not, in general, consistent). We now consider the related hypothesis testing problem with the null given in (9) and with the alternative

$$H_1 : f_p \in \mathcal{G}^{\pi_0}(b, L, \beta) \equiv \{f | f_p(1) \leq \pi_0 - b \text{ and } f_p \in \Sigma(L, \beta)\}$$

which allows for an upper confidence interval for $\overline{\pi}$ (and π itself). Under the maintained hypothesis that the infimum is taken at 1, the rate at which $b = b_n$ can approach 0 with H_1 and H_0 being distinguished gives the minimax rate for inference on $\overline{\pi}$ when the density

under the alternative is constrained to the Hölder class $\Sigma(L, \beta)$.

To extend the approach of the previous sections to this model, let $\hat{\pi}(\lambda) = \frac{1}{n(1-\lambda)} \sum_{i=1}^{n} I(\hat{p}_i > \lambda)$ be the estimate of π used by Storey (2002) for a given tuning parameter λ. We form our test by searching over the tuning parameter λ after an appropriate normalization:

$$T_n(\pi_0) = \max_{0 \leq \lambda < 1} \sqrt{n(1-\lambda)}[\pi_0 - \hat{\pi}(\lambda)]$$

where we write max since the maximum is obtained. We define our test $\phi_n(\pi_0)$ of (9) to reject when $T_n(\pi_0)$ is greater than the critical value $c_{n,\alpha}(\pi_0)$, given by the $1-\alpha$ quantile of $T_n(\pi_0)$ under the distribution $\pi_0 \cdot \text{unif}(0,1) + (1-\pi_0) \cdot \delta_0$, where δ_0 is a unit mass at 0 and $\text{unif}(0,1)$ denotes the uniform distribution on $(0,1)$.

We note that $T_n(\pi_0)$ is related to the test statistics used by Donoho and Jin (2004) and Meinshausen and Rice (2006), and can be considered a version of their approach that searches over the larger, rather than smaller, p-values. Donoho and Jin (2004) set $\pi_0 = 1$, and consider alternatives where π is close to one and the remaining p-values come from a normal location model with the mean slightly perturbed, achieving a certain form of adaptivity with respect to the amount of deviation of π and the normal location under the alternative. Meinshausen and Rice (2006) consider estimation of π in related settings with π close to one (see also Cai, Jin, and Low, 2007, for additional results in this setting). In contrast, $T_n(\pi_0)$ looks at the larger ordered p-values in order to achieve adaptivity to the smoothness of the distribution of p-values under the alternative in a setting where π may not be close to 1.

We now state the result giving the adaptive rate for the test $\phi_n(\pi_0)$.

Theorem 3.1. *The test $\phi_n(\pi_0)$ is level α for (9). Furthermore, there exists a constant C^* such that, for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} \inf_{\beta \in [\varepsilon, 1]} \inf_{\mathcal{G}^{\pi_0}(C^*(n/\log\log n)^{-\beta/(2\beta+1)}, L, \beta)} E_{f_p} \phi_n = 1.$$

Given the close relation between nonparametric inference on densities and conditional means (cf. Brown and Low, 1996; Nussbaum, 1996), a lower bound for this problem analogous to the one given in Theorem 2.1 for the regression problem seems likely. However, in the interest of space, we do not pursue such an extension.

4 Proofs

Proof of Theorem 2.1

The following gives a bound on average power over certain alternatives, and will be used to obtain a bound on minimax power over certain alternatives conditional on $X_1, \ldots, X_n$. Note that the bound goes to zero as $M \to \infty$ for $C < 1$.

Lemma 4.1. *Let $W_1, \ldots, W_N$ be independent under measures P_0 and $P_1, \ldots, P_N$, with $W_i \sim N(0, s_i^2)$ under P_0 and $W_i \sim N(m_{i,k}, s_i^2)$ under P_k. Let $\underline{M}$ and $\overline{M}$ be integers with $1 \leq 2^{\underline{M}} < 2^{\overline{M}} \leq N$, and let $M = \overline{M} - \underline{M} + 1$. Let ϕ be a test statistic that takes the data to a rejection probability in $[0, 1]$. Suppose that, for some C,*

$$|m_{i,k}/s_i| \leq C\sqrt{\log M}/\sqrt{k} \qquad \textit{all } i, k$$

and that $m_{i,k} = 0$ for $i > k$. Then

$$\frac{1}{M}\sum_{j=\underline{M}}^{\overline{M}} E_{P_{2^j}}\phi - E_{P_0}\phi \leq \sqrt{\frac{1}{M}\left(M^{C^2} - 1\right) + \frac{2}{M(\sqrt{2}-1)}C^2(\log M)M^{C^2/\sqrt{2}}}$$
$$\equiv B(C, M).$$

Proof. We express the average power as a sample mean of likelihood ratios under the null, following arguments used in, e.g., Lepski and Tsybakov (2000):

$$\frac{1}{M}\sum_{j=\underline{M}}^{\overline{M}} E_{P_{2^j}}\phi - E_{P_0}\phi = \frac{1}{M}\sum_{j=\underline{M}}^{\overline{M}} E_{P_0}\frac{dP_{2^j}}{dP_0}\phi - E_{P_0}\phi$$
$$= E_{P_0}\left\{\frac{1}{M}\sum_{j=\underline{M}}^{\overline{M}}\left[\exp\left(\sum_{i=1}^{N}(\mu_{i,j}Z_i - \mu_{i,j}^2/2)\right) - 1\right]\phi\right\}$$

where $\mu_{i,j} = m_{i,2^j}/s_i$ and $Z_i \equiv W_i/s_i$ are independent $N(0, 1)$ under P_0. By Cauchy-Schwarz,

the above display is bounded by the square root of

$$\frac{1}{M^2}\sum_{j=\underline{M}}^{\overline{M}}\sum_{\ell=\underline{M}}^{\overline{M}} E_{P_0}\left[\exp\left(\sum_{i=1}^{N}(\mu_{i,j}Z_i-\mu_{i,j}^2/2)\right)-1\right]\left[\exp\left(\sum_{i=1}^{N}(\mu_{i,\ell}Z_i-\mu_{i,\ell}^2/2)\right)-1\right]$$
$$=\frac{1}{M^2}\sum_{j=\underline{M}}^{\overline{M}}\sum_{\ell=\underline{M}}^{\overline{M}}\left[\exp\left(\sum_{i=1}^{N}\mu_{i,j}\mu_{i,\ell}\right)-1\right]$$
$$\leq \frac{1}{M^2}\sum_{j=\underline{M}}^{\overline{M}}[\exp\left(C^2\log M\right)-1]+\frac{2}{M^2}\sum_{j=\underline{M}}^{\overline{M}}\sum_{\ell=\underline{M}}^{j-1}[\exp\left(C^2(\log M)2^{-|j-\ell|/2}\right)-1] \tag{10}$$

where the equality follows from using properties of the normal distribution to evaluate the expectation, and the last step follows by plugging in the bound $C[\sqrt{\log M}/\sqrt{2^k}]I(i\leq 2^k)$ for $\mu_{i,k}=m_{i,2^k}/s_{2^k}$. Using the fact that $\exp(x)-1\leq x\cdot\exp(x)$, the inner sum of the second term can be bounded by

$$\sum_{\ell=\underline{M}}^{j-1} C^2(\log M)2^{-|j-\ell|/2}\exp\left(C^2(\log M)2^{-|j-\ell|/2}\right)$$
$$\leq C^2(\log M)\exp\left(C^2(\log M)/\sqrt{2}\right)\sum_{k=1}^{\infty}2^{-k/2}=C^2(\log M)M^{C^2/\sqrt{2}}\frac{1}{\sqrt{2}-1}.$$

Plugging this into (10) and taking the square root gives the claimed bound. □

Before proceeding, we recall a result regarding uniform convergence of empirical cdfs.

Lemma 4.2. *Let $Z_1,\ldots,Z_n$ be iid real valued random variables with cdf F_Z. Then, for any sequence a_n with $a_n n\to\infty$,*

$$\sup_{F(z)\geq a_n}\left|\frac{\frac{1}{n}\sum_{i=1}^{n}I(Z_i\leq z)-F_Z(z)}{F_Z(z)}\right|\xrightarrow{p}0.$$

Proof. See Wellner (1978), Theorem 0. □

Let P_X denote the product measure on the X_i's common to all distributions in the model, and let A_n be the event that

$$\eta t/2\leq\frac{1}{n}\sum_{i=1}^{n}I\left(|X_i|\leq t\right)\leq 2t/\eta\quad\text{for all }(\log n)/n<t<\eta. \tag{11}$$

We will use the fact that $P_X(A_n) \to 1$, which follows by plugging the condition (2) into the conclusion of Lemma 4.2 for $Z_i = |X_i|$.

We now construct a function in $\mathcal{G}(b, L, \beta)$ for each $\beta \in [\underline{\beta}, \overline{\beta}]$ that, along with Lemma 4.1, can be used to prove the theorem.

Lemma 4.3. *For a given L, β, n and c, define*

$$g_{\beta,n,c}(x) = \max\{c[(\log\log n)/n]^{\beta/(2\beta+1)} - L|x|^\beta, 0\}.$$

Let $0 < \underline{\beta} < \overline{\beta}$ be given. For small enough c, we have the following. For any sequence of tests ϕ_n taking the data into a $[0,1]$ rejection probability,

$$\lim_{n\to\infty} \inf_{\beta\in[\underline{\beta},\overline{\beta}]} \left[E_{g_{\beta,n,c}}\phi_n - E_0\phi_n\right] = 0.$$

Proof. Let $\hat{N}(\beta) = \hat{N}(\beta, X_1, \ldots, X_n) = \sum_{i=1}^n I(L|X_i|^\beta \le c[(\log\log n)/n]^{\beta/(2\beta+1)}) = \sum_{i=1}^n I(|X_i| \le (c/L)^{1/\beta}[(\log\log n)/n]^{1/(2\beta+1)})$. Let $\eta > 0$ satisfy condition (2). Letting $N(\beta) = \eta^{-1}\cdot n\cdot[(\log\log n)/n]^{1/(2\beta+1)}$, we have, for $(c/L) \le 1$, $\hat{N}(\beta) \le 2N(\beta)$ for all $\beta \in [\underline{\beta}, \overline{\beta}]$ on the event A_n defined in (11). Note that $N(\beta)/\log\log n = \eta^{-1}(n/\log\log n)^{2\beta/(2\beta+1)}$ and $g_{\beta,n,c}(x) \le c[(\log\log n)/n]^{\beta/(2\beta+1)}$ for all x, so that

$$g_{\beta,n,c}(x) \le c[(\log\log n)/n]^{\beta/(2\beta+1)} = c\eta^{-1/2}\left[N(\beta)/\log\log n\right]^{-1/2} \tag{12}$$

for all x.

Let $\underline{M}_n = \lceil \log_2[2N(\underline{\beta})]\rceil$ and $\overline{M}_n = \lfloor \log_2[2N(\overline{\beta})]\rfloor$, and let $\beta_{k,n}$ be such that $k = 2N(\beta_{k,n})$ (so that $\underline{\beta} \le \beta_{k,n} \le \overline{\beta}$ for $2^{\underline{M}_n} \le k \le 2^{\overline{M}_n}$). Let $M_n = \overline{M}_n - \underline{M}_n - 1$ and note that $M_n \ge (\log n)/K$ for a constant K that depends only on $\underline{\beta}$ and $\overline{\beta}$. Plugging these into the bound in (12) yields the bound

$$\frac{g_{\beta_{k,n},n,c}(x)}{\sigma(x)} \le \frac{c\eta^{-1/2}\sqrt{2}k^{-1/2}[\log(KM_n)]^{1/2}}{\inf_{|x|<\eta}\sigma(x)} \le 2c\eta^{-1}k^{-1/2}[\log M_n]^{1/2} \tag{13}$$

where the last inequality holds for large enough n (the last equality uses the fact that $\inf_{|x|<\eta}\sigma(x) \ge \eta^{1/2}$ for η satisfying condition (2)).

Since $\hat{N}(\beta) \le 2N(\beta)$ for all $\underline{\beta} \le \beta \le \overline{\beta}$ on the event A_n, we have, on this event, letting $X_{(i)}$ be the observation X_i corresponding to the ith least value of $|X_i|$, $|g_{\beta_{n,k},n,c}(X_{(i)})| = 0$ for $i > 2N(\beta_{n,k}) = k$ for all $\underline{\beta} \le \beta_{n,k} \le \overline{\beta}$. Using this and the bound in (13), we can apply

Lemma 4.1 conditional on $X_1, \ldots, X_n$ to obtain, for any test ϕ,

$$\frac{1}{M_n} \sum_{j=\underline{M}_n}^{\overline{M}_n} E_{g_{\beta_{n,2^j},n,c}}(\phi|X_1, \ldots, X_n) - E_0(\phi|X_1, \ldots, X_n) \leq B(2c\eta^{-1}, M_n)$$

on the event A_n for large enough n. Thus,

$$\begin{aligned}
&\lim_{n\to\infty} \inf_{\beta\in[\underline{\beta},\overline{\beta}]} E_{g_{\beta,n,c}}\phi_n - E_0\phi_n \leq \lim_{n\to\infty} \frac{1}{M_n} \sum_{j=\underline{M}_n}^{\overline{M}_n} E_{g_{\beta_{n,2^j},n,c}}\phi_n - E_0\phi_n \\
&\leq \lim_{n\to\infty} E_{P_X} \frac{1}{M_n} \sum_{j=\underline{M}_n}^{\overline{M}_n} \left[E_{g_{\beta_{n,2^j},n,c}}(\phi|X_1, \ldots, X_n) - E_0(\phi|X_1, \ldots, X_n)\right] I(A_n) \\
&+ [1 - P_X(A_n)] \\
&\leq \lim_{n\to\infty} B(2c\eta^{-1}, M_n) + [1 - P_X(A_n)].
\end{aligned}$$

This converges to zero for small enough c. □

Theorem 2.1 now follows from Lemma 4.3, since $g_{\beta,n,c} \in \mathcal{G}(c[(\log\log n)/n]^{\beta/(2\beta+1)}, L, \beta)$.

4.1 Proof of Theorem 2.2

For the test ϕ_n^*, we have, for $(b/L)^{1/\beta} < \eta$,

$$\begin{aligned}
&\inf_{g\in\mathcal{G}(b,L,\beta)} E_g(\phi_n^*|X_1, \ldots, X_n) \\
&\geq \inf_{g\in\mathcal{G}(b,L,\beta)} P_g \left\{ \frac{\sum_{|X_i|\leq(b/L)^{1/\beta}} Y_i}{\sqrt{\sum_{i=1}^n I(|X_i| \leq (b/L)^{1/\beta})}} > c_{\alpha,n} | X_1, \ldots, X_n \right\}.
\end{aligned}$$

Under P_g, the random variable $\frac{\sum_{|X_i|\leq(b/L)^{1/\beta}} Y_i}{\sqrt{\sum_{i=1}^n I(|X_i|\leq(b/L)^{1/\beta})}}$ in the conditional probability statement above is, conditional on $X_1, \ldots, X_n$, distributed as a normal variable with mean

$$\frac{\sum_{|X_i|\leq(b/L)^{1/\beta}} g(X_i)}{\sqrt{\sum_{i=1}^n I(|X_i| \leq (b/L)^{1/\beta})}} \geq \frac{b}{2} \frac{\sum_{i=1}^n I(|X_i| \leq (b/(2L))^{1/\beta})}{\sqrt{\sum_{i=1}^n I(|X_i| \leq (b/L)^{1/\beta})}} \tag{14}$$

and variance

$$\frac{\sum_{|X_i|\leq(b/L)^{1/\beta}} \sigma^2(X_i)}{\sum_{i=1}^n I(|X_i| \leq (b/L)^{1/\beta})} \leq \sup_x \sigma^2(x), \tag{15}$$

where the lower bound on the mean holds for $g \in \mathcal{G}(b, L, \beta)$ by noting that, for $g \in \mathcal{G}(b, L, \beta)$, $g(x) \geq b - L|x|^\beta$, so $g(x) \geq 0$ for $|x| \leq (b/L)^{1/\beta}$, and, for $|x| \leq [b/(2L)]^{1/\beta}$, $g(x) \geq b - L|[b/(2L)]^{1/\beta}|^\beta = b/2$. Let $K = 2\eta^{-1}$. On the event A_n defined in (11) (which holds with probability approaching one), for b in the appropriate range, the right hand side of (14) is bounded from below by

$$\frac{b}{2} \cdot \frac{\frac{1}{K} \cdot n \cdot (b/(2L))^{1/\beta}}{\sqrt{K \cdot n \cdot (b/L)^{1/\beta}}} = \frac{1}{2K\sqrt{K}} \cdot 2^{-1/\beta} \cdot L^{-1/(2\beta)} \sqrt{n} b^{1+1/(2\beta)}$$

For $b = c(n/\log\log n)^{-\beta/(2\beta+1)}$, this is

$$\frac{1}{2K\sqrt{K}} \cdot 2^{-1/\beta} \cdot L^{-1/(2\beta)} c^{1+1/(2\beta)} \sqrt{\log\log n}$$

and, for large enough n, this choice of b is in the range where the bounds in (11), (14) and (15) can be applied for all $\beta \in [\varepsilon, 1]$. Thus, on the event A_n, we have, for large enough c,

$$\begin{aligned} &\inf_{\beta \in [\varepsilon,1]} \inf_{g \in \mathcal{G}(c(n/\log\log n)^{\beta/(2\beta+1)}, L, \beta)} E_g(\phi_n^* | X_1, \ldots, X_n) \\ &\geq \inf_{\beta \in [\varepsilon,1]} 1 - \Phi\left(\frac{c_{\alpha,n} - \frac{1}{2K\sqrt{K}} \cdot 2^{-1/\beta} \cdot L^{-1/(2\beta)} c^{1+1/(2\beta)} \sqrt{\log\log n}}{\sup_x \sigma(x)} \right). \end{aligned}$$

By the law of the iterated logarithm applied to the iid $N(0,1)$ sequence $\{Y_i/\sigma(X_i)\}_{i=1}^n$, we have $c_{\alpha,n} \leq C\sqrt{\log\log n}$ for large enough n for any $C > \sqrt{2}\sup_x \sigma(x)$. For $c > \sup_{\beta \in [\varepsilon,1]} \left(\sqrt{2}\sup_x \sigma(x) 2K^{3/2} 2^{1/\beta} L^{1/(2\beta)}\right)^{2\beta/(2\beta+1)}$, it follows that the above display converges to 0 as $n \to \infty$. Since this bound holds on an event with probability approaching one, the result follows.

4.2 Proof of Corollary 3.1

The first display follows by Theorem 2.1 since $\phi_n = I(\hat{c} > 0)$ is level α for $H_0 : g = 0$, and the display is bounded by

$$\limsup_n \inf_{\beta \in [\underline{\beta}, \overline{\beta}]} \inf_{g \in \mathcal{G}(C_*(n/\log\log n)^{-\beta/(2\beta+1)}, L, \beta)} E_g I\left(\hat{c} > 0\right).$$

For the second display, note that, for any constant a, the distribution of $T_n(a)$ under g is the same as the distribution of $T_n(0)$ under the function $g-a$ that takes t to $g(t)-a$. Thus,

$$P_g\left(\hat{c}^* > g(0)-b\right) = P_g\left(T_n(g(0)-b) > c_{\alpha,n}\right) = P_{g-g(0)+b}\left(T_n(0)\right) > c_{\alpha,n}\right).$$

Since $g-g(0)+b$ is in $\mathcal{G}(b,L,\beta)$ for any $g \in \Sigma(L,\beta)$, the result follows from Theorem 2.2.

4.3 Proof of Corollary 3.2

The proof of the second part of the corollary (the extension of Theorem 2.2) is similar to the original proof and is omitted. To prove the first part of the corollary (the extension of Theorem 2.1), assume, without loss of generality, that $\tau_0 = 0$. Define $\mathrm{sgn}(X_i)$ to be -1 for $X_i \le 0$ and 1 for $X_i > 0$. Note that, for any function $g \in \mathcal{G}(b, L/2, \beta) \cap G_{|x|\downarrow}$, the function $m_g(x) = g(x)\cdot \mathrm{sgn}(X_i) - 2g(0)I(X_i > 0)$ is in $\Sigma(L,\beta)$ and is nonincreasing (to verify Hölder continuity, note that, for x, x' with $\mathrm{sgn}(x) = \mathrm{sgn}(x')$, $|m_g(x) - m_g(x')| \le |g(x) - g(x')|$ and, for x, x' with $\mathrm{sgn}(x) \ne \mathrm{sgn}(x')$, $|m_g(x) - m_g(x')| = |g(x) - g(0)| + |g(x') - g(0)| \le (L/2)|x|^\beta + (L/2)|x'|^\beta \le L|x-x'|^\beta$, where the last step follows since $|x-x'| \ge x \vee x'$).

Note that, under $m = m_g$, $\tau = 2g(0)$, the regression function is $x \mapsto m_g(x) + 2g(0)I(x_i > 0) = g(x)\cdot \mathrm{sgn}(X_i)$ so that $\{Y_i \cdot \mathrm{sgn}(X_i), X_i\}_{i=1}^n$ are distributed according to the original regression model (1) with the given function g. Of course, for $m(x) = 0$ all x and $\tau = 0$, the regression function is 0 for all x. Thus, for any level α test ϕ_n of $(m,\tau) = (0,0)$, we can construct a test ϕ_n^* of (3) in the original model (1) that has identical power at g to the power in the regression discontinuity model at $(m_g, 2g(0))$ for any g with $g \in \mathcal{G}(b, L/2, \beta) \cap \mathcal{G}_{|x|\downarrow}$ for some b, L and β. Since $(m_g, 2g(0)) \in \mathcal{G}^{\mathrm{rd}}(2b, L, \beta)$ whenever $g \in \mathcal{G}(b, L/2, \beta) \cap \mathcal{G}_{|x|\downarrow}$ by the argument above, it follows that $\inf_{\beta\in[\underline{\beta},\overline{\beta}]} \inf_{(m,\tau)\in\mathcal{G}^{\mathrm{rd}}(2c(n/\log\log n)^{-\beta/(2\beta+1)},L,\beta)} E_{(m,\tau)}\phi_n \le \inf_{\beta\in[\underline{\beta},\overline{\beta}]} \inf_{g\in\mathcal{G}(c(n/\log\log n)^{-\beta/(2\beta+1)},L/2,\beta)\cap\mathcal{G}_{|x|\downarrow}} E_g\phi_n^*$, which converges to zero for c small enough by Theorem 2.1.

4.4 Proof of Theorem 3.1

We first show that the distribution used to obtain the critical value is least favorable for this test statistic, so that the test does in fact have level α.

Lemma 4.4. *The distribution* $\underline{f}_{\pi_0} = \pi_0 \cdot \mathit{unif}(0,1) + (1-\pi_0)\delta_0$, *where* δ_0 *is a unit mass at*

0, *is least favorable for* $T_n(\pi_0)$ *under the null* $\pi \geq \pi_0$*:*

$$P_{f_p}(T_n(\pi_0) > c) \leq P_{\underline{f}_{\pi_0}}(T_n(\pi_0) > c) \quad \textit{for } f_p \textit{ defined by (8) with } \pi \geq \pi_0.$$

Proof. For $\hat{p}_1, \ldots, \hat{p}_n$ drawn from $f_p = \pi_0 \cdot \text{unif}(0,1) + (1-\pi_0) f_1$, let $q_1, \ldots, q_n$ be obtained from $\hat{p}_1, \ldots, \hat{p}_n$ by setting all $\hat{p}_i$'s drawn from the alternative f_1 to 0. Then $T_n(\pi_0)$ weakly increases when evaluated at the q_i's instead of the $\hat{p}_i$'s, and the distribution under f_p of $T_n(\pi_0)$ evaluated with the q_i's is equal to the distribution under $\underline{f}_{\pi_0}$ of $T_n(\pi_0)$ evaluated with the $\hat{p}_i$'s. □

The result now follows from similar arguments to the proof of Theorem 2.2 after noting that $c_{n,\alpha}(\pi_0)/\sqrt{\log\log n}$ is bounded as $n \to \infty$ (cf. Shorack and Wellner, 2009, Chapter 16).

References

Andrews, D. W. K., and M. M. A. Schafgans (1998): "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, 65(3), 497–517.

Brown, L. D., and M. G. Low (1996): "Asymptotic equivalence of nonparametric regression and white noise," *The Annals of Statistics*, 24(6), 2384–2398.

Cai, T. T., J. Jin, and M. G. Low (2007): "Estimation and confidence sets for sparse normal mixtures," *The Annals of Statistics*, 35(6), 2421–2449.

Cai, T. T., and M. G. Low (2004): "An Adaptation Theory for Nonparametric Confidence Intervals," *The Annals of Statistics*, 32(5), 1805–1840.

Cai, T. T., M. G. Low, and Y. Xia (2013): "Adaptive confidence intervals for regression functions under shape constraints," *The Annals of Statistics*, 41(2), 722–750.

Chamberlain, G. (1986): "Asymptotic efficiency in semi-parametric models with censoring," *Journal of Econometrics*, 32(2), 189–218.

Donoho, D., and J. Jin (2004): "Higher criticism for detecting sparse heterogeneous mixtures," *The Annals of Statistics*, 32(3), 962–994.

Dümbgen, L. (2003): "Optimal confidence bands for shape-restricted curves," *Bernoulli*, 9(3), 423–449.

Dümbgen, L., and V. G. Spokoiny (2001): "Multiscale Testing of Qualitative Hypotheses," *The Annals of Statistics*, 29(1), 124–152.

Efron, B. (2012): *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Cambridge University Press, Cambridge.

Fan, J. (1996): "Test of Significance Based on Wavelet Thresholding and Neyman's Truncation," *Journal of the American Statistical Association*, 91(434), 674–688.

Hahn, J., P. Todd, and W. Van der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69(1), 201–209.

Heckman, J. (1990): "Varieties of Selection Bias," *The American Economic Review*, 80(2), 313–318.

Imbens, G. W., and T. Lemieux (2008): "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142(2), 615–635.

Ingster, Y., and I. A. Suslina (2003): *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. Springer.

Khan, S., and E. Tamer (2010): "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 78(6), 2021–2042.

Lepski, O., and A. Tsybakov (2000): "Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point," *Probability Theory and Related Fields*, 117(1), 17–48.

Low, M. G. (1997): "On nonparametric confidence intervals," *The Annals of Statistics*, 25(6), 2547–2554.

Meinshausen, N., and J. Rice (2006): "Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses," *The Annals of Statistics*, 34(1), 373–393.

Nussbaum, M. (1996): "Asymptotic equivalence of density estimation and Gaussian white noise," *The Annals of Statistics*, 24(6), 2399–2430.

Pouet, C. (2000): "On Testing Nonparametric Hypotheses for Analytic Regression Functions in Gaussian Noise," *Mathematical Methods of Statistics*, 8(4), 536–549.

Shorack, G. R., and J. A. Wellner (2009): *Empirical Processes with Applications to Statistics*. SIAM.

Spokoiny, V. G. (1996): "Adaptive hypothesis testing using wavelets," *The Annals of Statistics*, 24(6), 2477–2498.

Storey, J. D. (2002): "A direct approach to false discovery rates," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3), 479–498.

Wellner, J. A. (1978): "Limit theorems for the ratio of the empirical distribution function to the true distribution function," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 45(1), 73–88.