

Simple Local Polynomial Density Estimators*

Matias D. Cattaneo[†]

Michael Jansson[‡]

Xinwei Ma[§]

June 11, 2019

Abstract

This paper introduces an intuitive and easy-to-implement nonparametric density estimator based on local polynomial techniques. The estimator is fully boundary adaptive and automatic, but does not require pre-binning or any other transformation of the data. We study the main asymptotic properties of the estimator, and use these results to provide principled estimation, inference, and bandwidth selection methods. As a substantive application of our results, we develop a novel discontinuity in density testing procedure, an important problem in regression discontinuity designs and other program evaluation settings. An illustrative empirical application is given. Two companion `Stata` and `R` software packages are provided.

Keywords: density estimation, local polynomial methods, regression discontinuity, manipulation test.

*A preliminary version of this paper circulated under the title “Simple Local Regression Distribution Estimators with an Application to Manipulation Testing”. We thank Sebastian Calonico, Toru Kitagawa, Zhuan Pei, Rocio Titiunik, Gonzalo Vazquez-Bare, the co-editor, Hongyu Zhao, an associate editor, and two reviewers for useful comments that improved our work and software implementations. Cattaneo gratefully acknowledges financial support from the National Science Foundation (SES 1357561 and SES 1459931). Jansson gratefully acknowledges financial support from the National Science Foundation (SES 1459967) and the research support of CREATES (funded by the Danish National Research Foundation under grant no. DNRF78).

[†]Department of Operations Research and Financial Engineering, Princeton University.

[‡]Department of Economics, UC Berkeley and *CREATES*.

[§]Department of Economics, UC San Diego.

1 Introduction

Flexible (nonparametric) estimation of a probability density function features prominently in empirical work in statistics, economics, and many other disciplines. Sometimes the density function is the main object of interest, while in other cases it is a useful ingredient in forming two-step nonparametric or semiparametric procedures. In program evaluation and causal inference settings, for example, nonparametric density estimators are used for manipulation testing, distributional treatment effect and counterfactual analysis, instrumental variables treatment effect specification and heterogeneity analysis, and common support/overlap testing. See [Imbens and Rubin \(2015\)](#) and [Abadie and Cattaneo \(2018\)](#) for reviews and further references.

A common problem faced when implementing density estimators in empirical work is the presence of evaluation points that lie on the boundary of the support of the variable of interest: whenever the density estimator is constructed at or near boundary points, which may or may not be known by the researcher, the finite- and large-sample properties of the estimator are affected. Standard kernel density estimators are invalid at or near boundary points, while other methods may remain valid but usually require choosing additional tuning parameters, transforming the data, a priori knowledge of the boundary point location, or some other boundary-related specific information or modification. Furthermore, it is usually the case that one type of density estimator is used for evaluation points at or near the boundary, while a different type is used for interior points.

We introduce a novel nonparametric estimator of a density function constructed using local polynomial techniques ([Fan and Gijbels, 1996](#)). The estimator is intuitive, easy to implement, does not require pre-binning of the data, and enjoys all the desirable features associated with local polynomial regression estimation. In particular, the estimator automatically adapts to the boundaries of the support of the density without requiring specific data modification or additional tuning parameter choices, a feature that is unavailable for most other density estimators in the literature: see [Karunamuni and Alberts \(2005\)](#) for a review on this topic. The most closely related approaches currently available in the literature are the local polynomial density estimators of [Cheng, Fan and Marron \(1997\)](#) and [Zhang and Karunamuni \(1998\)](#), which require knowledge of the boundary location and pre-binning of the data (or, more generally, pre-estimation of the density near the boundary), and hence introduce additional tuning parameters that need to be chosen.

The heuristic idea underlying our estimator, and differentiating it from other existing ones, is

simple to explain: whereas other nonparametric density estimators are constructed by smoothing out a histogram-type estimator of the density, our estimator is constructed by smoothing out the empirical distribution function using local polynomial techniques. Accordingly, our density estimator is constructed using a preliminary tuning-parameter-free and \sqrt{n} -consistent distribution function estimator (where n denotes the sample size), implying in particular that the only tuning parameter required by our approach is the bandwidth associated with the local polynomial fit at each evaluation point. For the resulting density estimator, we provide (i) asymptotic expansions of the leading bias and variance, (ii) asymptotic Gaussian distributional approximation and valid statistical inference, (iii) consistent standard error estimators, and (iv) consistent data-driven bandwidth selection based on an asymptotic mean squared error (MSE) expansion. All these results apply to both interior and boundary points in a fully automatic and data-driven way, without requiring boundary-specific transformations of the estimator or of the data, and without employing additional tuning parameters (beyond the main bandwidth present in any kernel-based nonparametric method).

As a substantive methodological application of our proposed density estimator, we develop a novel discontinuity in density testing procedure. In a seminal paper, [McCrary \(2008\)](#) proposed the idea of manipulation testing via discontinuity in density testing for regression discontinuity (RD) designs, and developed an implementation thereof using the density estimator of [Cheng, Fan and Marron \(1997\)](#), which requires pre-binning of the data and choosing two tuning parameters. On the other hand, the new proposed discontinuity in density test employing our density estimator only requires the choice of one tuning parameter, and enjoys other features associated with local polynomials methods. We also illustrate its performance with an empirical application employing the canonical Head Start data in the context of RD designs ([Ludwig and Miller, 2007](#)). For introductions to RD designs, and further references, see [Imbens and Lemieux \(2008\)](#), [Lee and Lemieux \(2010\)](#), and [Cattaneo, Titiunik and Vazquez-Bare \(2017\)](#). For recent papers on modern RD methodology see, for example, [Arai and Ichimura \(2018\)](#), [Ganong and Jäger \(2018\)](#), [Hyytinen, Meriläinen, Saarimaa, Toivanen and Tukiainen \(2018\)](#), [Dong, Lee and Gou \(2019\)](#), and references therein.

Finally, we provide two general purpose software packages, for `Stata` and `R`, implementing the main results discussed in the paper. [Cattaneo, Jansson and Ma \(2018\)](#) discusses the package

`rddensity`, which is specifically tailored to manipulation testing (i.e., two-sample discontinuity in density testing), while [Cattaneo, Jansson and Ma \(2019\)](#) discusses the package `lpdensity`, which provides generic density estimation over the support of the data.

The rest of the paper is organized as follows. Section 2 introduces the density estimator and Section 3 gives the main technical results. Section 4 applies these results to nonparametric discontinuity in density testing (i.e., manipulation testing), while Section 5 illustrates the new method with an empirical application. Section 6 discusses extensions and concludes. The supplemental appendix (SA hereafter) contains additional methodological and technical results and reports all theoretical proofs. In addition, to conserve space, we relegate to the SA and to our two companion software articles the presentation of simulation evidence highlighting the finite sample properties of our proposed density estimator.

2 Boundary Adaptive Density Estimation

Suppose $\{x_1, x_2, \dots, x_n\}$ is a random sample, where x_i is a continuous random variable with a smooth cumulative distribution function over its support $\mathcal{X} \subseteq \mathbb{R}$. The probability density function is $f(x) = \frac{\partial}{\partial x} \mathbb{P}[x_i \leq x]$, where the derivative is interpreted as a one-sided derivative at a boundary point of \mathcal{X} . Our results apply to bounded or unbounded support \mathcal{X} , which is an important feature in empirical applications employing density estimators.

Letting $\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(x_i \leq x)$ denote the classical empirical distribution function, our proposed local polynomial density estimator is

$$\hat{f}(x) = \mathbf{e}'_1 \hat{\beta}(x), \quad \hat{\beta}(x) = \underset{\mathbf{b} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^n \left[\hat{F}(x_i) - \mathbf{r}_p(x_i - x)' \mathbf{b} \right]^2 K \left(\frac{x_i - x}{h} \right),$$

where $\mathbf{e}_1 = (0, 1, 0, \dots, 0)'$ is the second $(p+1)$ -dimensional unit vector, $\mathbf{r}_p(u) = (1, u, u^2, \dots, u^p)'$ is a p -th order polynomial expansion, $K(\cdot)$ denotes a kernel function, h is a positive bandwidth, and $p \geq 1$. In other words, we take the empirical distribution function \hat{F} as the starting point, then construct a smooth local approximation to \hat{F} using a polynomial expansion, and finally obtain the density estimator \hat{f} as the slope coefficient in the local polynomial regression.

The idea behind the density estimator $\hat{f}(x)$ is explained graphically in [Figure 1](#). In this figure, we consider three distinct evaluation points on $\mathcal{X} = [-1, 1]$: a is near the lower boundary, b is

an interior point, and $c = 1$ is the upper boundary. The conventional kernel density estimator, $\hat{f}_{\text{KD}}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)$, is valid for interior points, but otherwise inconsistent. See, e.g., [Wand and Jones \(1995\)](#) for a classical reference. On the other hand, our density estimator $\hat{f}(x)$ is valid for all evaluation points $x \in \mathcal{X}$ and can be used directly, without any modifications to approximate the unknown density. Figure 1 is constructed using $n = 500$ observations. The top panel plots one realization of the empirical distribution function $\hat{F}(x)$ in dark gray, and the local polynomial fits for the three evaluation points $x \in \{a, b, c\}$ in red, the latter implemented with $p = 2$ (quadratic approximation) and bandwidth h (different value for each evaluation point considered). The vertical light gray areas highlight the localization region controlled by the bandwidth choice, that is, only observations falling in these regions are used to smooth out the empirical distribution function via local polynomial approximation, depending on the evaluation point. The estimator $\hat{f}(x)$ is the slope coefficient accompanying the first-order term in the local polynomial approximation, which is depicted in the bottom panel of Figure 1 as the solid line in red. The bottom panel also plots three other curves: dashed blue line corresponding to the population density function, dash-dotted green line corresponding to the average of our density estimate over simulations, and dashed black line corresponding to the average of the standard kernel density estimates $\hat{f}_{\text{KD}}(x)$.

Figure 1 illustrates how our proposed density estimator adapts to (near) boundary points automatically, showing graphically its good performance in repeated samples. Evaluation point b is an interior point and, consequently, a symmetric smoothing around that point is employed, just like the standard estimator $\hat{f}_{\text{KD}}(x)$ does. On the other hand, evaluation points a and c both exhibit boundary bias if the standard kernel density estimator is used: point a is near the boundary and hence employs asymmetric smoothing, while point c is at the upper boundary and hence employs one-sided smoothing. In contrast, our proposed density estimator $\hat{f}(x)$ automatically adapts to the boundary point, as the bottom panel in Figure 1 illustrates.

3 Main Technical Results

We summarize two main large sample results concerning the proposed density estimator: (i) an asymptotic distributional approximation with precise leading bias and variance characterizations, and (ii) a consistent standard error estimator, which is also data-driven and fully automatic. Both results are boundary adaptive and do not require prior knowledge of the shape of \mathcal{X} . We report pre-

liminary technical lemmas, additional theoretical results, and detailed proofs in the SA to conserve space. Extensions and other applications of our methods are mentioned in Section 6.

Assumption 1 (DGP) $\{x_1, x_2, \dots, x_n\}$ is a random sample with distribution function F that is $p + 1$ times continuously differentiable for some $p \geq 1$ in a neighborhood of the evaluation point x , and the probability density function of x_i , denoted by f , is positive at x .

This assumption imposes basic regularity conditions on the data generating process, ensuring that $f(x)$ is well-defined and possesses enough smoothness.

Assumption 2 (Kernel) The kernel function $K(\cdot)$ is nonnegative, symmetric, and continuous on its support $[-1, 1]$.

This assumption is standard in nonparametric estimation, and is satisfied for common kernel functions. We exclude kernels with unbounded support (e.g., Gaussian kernel) for simplicity, since such kernels will always hit boundaries. Our results, however, can be extended to accommodate kernel functions with unbounded support, albeit more cumbersome notation would be needed.

The following theorem gives a characterization of the asymptotic bias and variance of $\hat{f}(x)$, as well as a valid distributional approximation. All limits are taken as $n \rightarrow \infty$ (and $h \rightarrow 0$) unless explicitly stated otherwise, \rightsquigarrow denotes weak convergence, and $F^{(s)}(x) = \partial^s F(x)/\partial x^s$ denotes the derivative, or one-sided derivative if at a boundary point, of $F(x)$.

Theorem 1 (Distributional Approximation) Suppose Assumption 1 and 2 hold. If $nh^2 \rightarrow \infty$ and $nh^{2p+1} = O(1)$, then

$$\frac{\hat{f}(x) - f(x) - h^p \mathcal{B}(x)}{\sqrt{\frac{1}{nh} \mathcal{V}(x)}} \rightsquigarrow \mathcal{N}(0, 1),$$

where, defining

$$\begin{aligned} \mathbf{A}(x) &= f(x) \int_{h^{-1}(\mathcal{X}-x)} \mathbf{r}_p(u) \mathbf{r}_p(u)' K(u) du, & \mathbf{a}(x) &= f(x) \frac{F^{(p+1)}(x)}{(p+1)!} \int_{h^{-1}(\mathcal{X}-x)} u^{p+1} \mathbf{r}_p(u) K(u) du, \\ \mathbf{B}(x) &= f(x)^3 \iint_{h^{-1}(\mathcal{X}-x)} \min\{u, v\} \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) dudv, \end{aligned}$$

the asymptotic bias and variance are $\mathcal{B}(x) = \mathbf{e}'_1 \mathbf{A}(x)^{-1} \mathbf{a}(x)$ and $\mathcal{V}(x) = \mathbf{e}'_1 \mathbf{A}(x)^{-1} \mathbf{B}(x) \mathbf{A}(x)^{-1} \mathbf{e}_1$, respectively.

In this theorem, the integration region reflects the effect of boundaries. Because $K(\cdot)$ is compactly supported, if x is an interior point, we have $h^{-1}(\mathcal{X} - x) \supset [-1, 1]$ for h small enough, thus ensuring the kernel function is not truncated and the local approximation is symmetric around x . On the other hand, for x near or at a boundary of \mathcal{X} (i.e., for h not small enough relative to the distance of x to the boundary), we have $h^{-1}(\mathcal{X} - x) \not\supset [-1, 1]$, and the local approximation is asymmetric (or one-sided). It follows that the density estimator $\hat{f}(x)$ is boundary adaptive and design adaptive, as in the case of local polynomial regression (Fan and Gijbels, 1996).

A simple and automatic variance estimator is $\hat{\mathcal{V}}(x) = \mathbf{e}'_1 \hat{\mathbf{A}}(x)^{-1} \hat{\mathbf{B}}(x) \hat{\mathbf{A}}(x)^{-1} \mathbf{e}_1$, where

$$\begin{aligned} \hat{\mathbf{A}}(x) &= \frac{1}{nh} \sum_{i=1}^n \mathbf{r}_p(\tilde{x}_i) \mathbf{r}_p(\tilde{x}_i)' K(\tilde{x}_i) \\ \hat{\mathbf{B}}(x) &= \frac{1}{n^3 h^3} \sum_{i,j,k=1}^n \mathbf{r}_p(\tilde{x}_j) \mathbf{r}_p(\tilde{x}_k)' K(\tilde{x}_j) K(\tilde{x}_k) \left[\mathbb{1}(x_i \leq x_j) - \hat{F}(x_j) \right] \left[\mathbb{1}(x_i \leq x_k) - \hat{F}(x_k) \right], \end{aligned}$$

with $\tilde{x}_i = h^{-1}(x_i - x)$ denoting the normalized observations to save notation. Let $\rightarrow_{\mathbb{P}}$ denote convergence in probability.

Theorem 2 (Variance Estimation) *If the conditions in Theorem 1 hold, then $\hat{\mathcal{V}}(x) \rightarrow_{\mathbb{P}} \mathcal{V}(x)$.*

As shown in this theorem, the variance estimator $\hat{\mathcal{V}}(x)$ does not require knowledge of the relative positioning of the evaluation point to boundaries of \mathcal{X} , that is, $\hat{\mathcal{V}}(x)$ is also boundary adaptive. A boundary adaptive bias estimator $\hat{\mathcal{B}}(x)$ can also be constructed easily, as shown in the SA.

Using the results above, and under mild regularity conditions, it follows that a pointwise approximate MSE-optimal bandwidth choice for our proposed density estimator $\hat{f}(x)$ is

$$h_{\text{MSE}}(x) = \left(\frac{\mathcal{V}(x)}{2p\mathcal{B}(x)^2} \right)^{1/(1+2p)} n^{-1/(1+2p)},$$

which can be easily implemented by replacing $\mathcal{B}(x)$ and $\mathcal{V}(x)$ with preliminary consistent estimators $\hat{\mathcal{B}}(x)$ and $\hat{\mathcal{V}}(x)$. The SA offers details on implementation and consistency of this MSE-optimal bandwidth selector, which can be used to establish its optimality in the sense of Li (1987), and also bandwidth selection for estimating higher-order density derivatives. We omit these results here due to space limitations.

Finally, we recommend implementing the density estimator $\hat{f}(x)$ with $p = 2$, which corresponds

to the minimal odd polynomial order choice (i.e., analogous to local linear regression). Higher-order local polynomials could be used, but they typically exhibit erratic behavior near boundary points, and lead to counter-intuitive weighting schemes. See [Fan and Gijbels \(1996, Chapter 3.3\)](#) for an automatic polynomial order selection methods that can be applied to our estimator as well.

4 Application to Manipulation Testing

Testing for manipulation is useful when units are assigned to two (or more) distinct groups using a hard-thresholding rule based on an observable variable, as it provides an intuitive and simple method to check empirically whether units are able to alter (i.e., manipulate) their assignment. Manipulation tests are used in empirical work both as falsification tests of regression discontinuity (RD) designs and as empirical tests with substantive implications in other program evaluation settings. Available methods from the RD literature include the original implementation of [McCrary \(2008\)](#) based on [Cheng, Fan and Marron \(1997\)](#), the empirical likelihood testing procedure of [Otsu, Xu and Matsushita \(2014\)](#) based on boundary-corrected kernels, and the finite sample binomial test presented in [Cattaneo, Titiunik and Vazquez-Bare \(2017\)](#) based on local randomization ideas.

In this section, we introduce a new manipulation testing procedure based on our proposed local polynomial density estimator. Our method requires choosing only one tuning parameter, avoids pre-binning the data, and permits the use of simple well-known weighting schemes (e.g., uniform or triangular kernel), thereby avoiding the need of choosing the length and positions of bins for pre-binning or employing more complicated boundary kernels. In addition, our method is intuitive, easy-to-implement, and fully data-driven: bandwidth selection methods are formally developed and implemented, along with valid inference methods based on robust bias correction.

To describe the manipulation testing setup, suppose units are assigned to one group (“control”) if $x_i < \bar{x}$ and to another group (“treatment”) if $x_i \geq \bar{x}$. For example, in the application discussed below, we employ the Head Start data, where x_i is a poverty index at the county level, $\bar{x} = 59.1984$ is a fixed cutoff determining eligibility to the program. The goal is to test formally whether the density $f(x)$ is continuous at \bar{x} , using the two subsamples $\{x_i : x_i < \bar{x}\}$ and $\{x_i : x_i \geq \bar{x}\}$, and thus the null and alternative hypotheses are:

$$H_0 : \lim_{x \uparrow \bar{x}} f(x) = \lim_{x \downarrow \bar{x}} f(x) \quad \text{vs} \quad H_1 : \lim_{x \uparrow \bar{x}} f(x) \neq \lim_{x \downarrow \bar{x}} f(x).$$

This hypothesis testing problem induces a nonparametric boundary point at $x = \bar{x}$ because two distinct densities need to be estimated, one from the left and the other from the right. Our proposed density estimator $\hat{f}(x)$ is readily applicable because it is boundary adaptive and fully automatic, and it can also be used to plot the density near the cutoff in an automatic way: see Figure 2 below for an example using the Head Start data.

Let \hat{F}_- and \hat{F}_+ be the empirical distribution functions constructed using only units with $x_i < \bar{x}$ and with $x_i \geq \bar{x}$, respectively. Then, \hat{f} can be applied twice, to the data below and above the cutoff, to obtain two estimators of the density at the boundary point \bar{x} , which we denote by $\hat{f}_-(\bar{x})$ and $\hat{f}_+(\bar{x})$, respectively. Thus, our proposed manipulation test statistic takes the form:

$$T_p(h) = \frac{\frac{n_+}{n} \hat{f}_+(\bar{x}) - \frac{n_-}{n} \hat{f}_-(\bar{x})}{\sqrt{\frac{n_+}{n} \frac{1}{nh_+} \hat{\mathcal{V}}_+(\bar{x}) + \frac{n_-}{n} \frac{1}{nh_-} \hat{\mathcal{V}}_-(\bar{x})}},$$

where $n_- = \sum_{i=1}^n \mathbf{1}(x_i < \bar{x})$ and $n = n_- + n_+$, $\hat{\mathcal{V}}_-(x)$ and $\hat{\mathcal{V}}_+(x)$ denote the variance estimators mentioned previously but now computed for the two subsamples $x_i < \bar{x}$ and $x_i \geq \bar{x}$, respectively, and h_- and h_+ denote the bandwidths used below and above \bar{x} . Employing our main theoretical results, we provide precise conditions so that the finite sample distribution of $T_p(h)$ can be approximated by the standard normal distribution, which leads to the following result: under the regularity conditions given above, and if $n \min\{h_-^2, h_+^2\} \rightarrow \infty$ and $n \max\{h_-^{1+2p}, h_+^{1+2p}\} \rightarrow 0$, then

$$\text{Under } \mathbf{H}_0 : \lim_{n \rightarrow \infty} \mathbb{P}[|T_p(h)| \geq \Phi_{1-\alpha/2}] = \alpha,$$

$$\text{Under } \mathbf{H}_1 : \lim_{n \rightarrow \infty} \mathbb{P}[|T_p(h)| \geq \Phi_{1-\alpha/2}] = 1,$$

where Φ_α denotes the α -quantile of the standard Gaussian distribution, $\alpha \in (0, 1)$. This establishes asymptotic validity and consistency of the α -level testing procedure that rejects \mathbf{H}_0 iff $|T(h)| \geq \Phi_{1-\alpha/2}$. The SA includes detailed proofs, and related implementation details.

A key implementation issue of our manipulation test is the choice of bandwidth h , a problem common to all nonparametric manipulation tests available in the literature. To select h in an automatic and data-driven way, we obtain an approximate MSE-optimal bandwidth choice for the point estimator $\hat{f}_+(\bar{x}) - \hat{f}_-(\bar{x})$, and then propose a consistent implementation thereof, which is denoted by \hat{h}_p . We give the details in the SA, where we also present alternative MSE-optimal bandwidth

selectors for each-side density estimator separately. Given the data-driven bandwidth choice \hat{h}_p , or its theoretical (infeasible) counterpart h_p , we propose a simple robust bias-corrected test statistic implementation following ideas in [Calonico, Cattaneo and Titiunik \(2014\)](#) and [Calonico, Cattaneo and Farrell \(2018\)](#); see the latter reference for theoretical results on higher-order refinements and the important role of pre-asymptotic variance estimation in the context of local polynomial regression estimation. Specifically, our proposed data-driven robust bias-corrected test statistic is $T_{p+1}(\hat{h}_p)$, which rejects H_0 iff $|T_{p+1}(\hat{h}_p)| \geq \Phi_{1-\alpha/2}$ for a nominal α -level test. This approach corresponds to a special case of manual bias-correction together with the corresponding adjustment of Studentization. A natural choice is $p = 2$, and this is the default in our companion **Stata** and **R** software implementations.

5 Empirical Illustration

We apply our manipulation test to the data of [Ludwig and Miller \(2007\)](#) on the original Head Start implementation in the U.S. In this empirical application, a discontinuity on access to program funds at the county level occurred in 1965 when the program was first implemented: the federal government provided grant writing assistance to the 300 poorest counties as measured by a poverty index, which was computed in 1965 using 1960 Census variables, thus creating a discontinuity in program eligibility. Using our notation, x_i denotes the poverty index for county i , and $\bar{x} = 59.1984$ is the cutoff point (i.e., the poverty index of the 300-th poorest municipality).

A manipulation test in this context amounts to testing whether there is a disproportional number of counties are situated above \bar{x} relative to those present below the cutoff. [Figure 2](#) presents the histogram of counties below and above the cutoff together with our local polynomial density estimate and associated pointwise robust bias-corrected confidence intervals over a grid of points near the cutoff \bar{x} , implemented using $p = 2$ and the MSE-optimal data-driven bandwidth estimate. [Table 1](#) presents the empirical results from our manipulation test. We consider two main approaches, both covered by our theoretical work and available in our software implementation: (i) using two distinct bandwidths on each side of the cutoff ($h_- \neq h_+$), and (ii) using a common bandwidth for each side of the cutoff ($h_- = h_+$), with h_- and h_+ denoting the bandwidth on the left and on the right, respectively. For each case, we consider three distinct implementations of our manipulation test, which varies the degree of polynomial approximation used to smooth

out the empirical distribution function: $T_q(h_p)$ denotes the test statistic constructed using a q -th order local polynomial density estimator, with bandwidth choice that is MSE-optimal for p -th order local polynomial density estimator. For example, our recommended choice is $T_3(h_2)$, with either common bandwidth or two different bandwidths, which amounts to first choosing MSE-optimal bandwidth(s) for a local quadratic fit, and then conducting inference using a cubic approximation. This approach is the simplest implementation of the robust bias correction inference: $T_p(h_p)$ does not lead to a valid inference approach because a first-order bias will make the test over-reject the null hypothesis. We also report the original implementation of the McCrary test for comparison.

Our empirical results show no evidence of manipulation. In fact, this finding is consistent with the underlying institutional knowledge of the program: the poverty index was constructed in 1965 at the federal level using county level information from the 1960 Census, which implies it is indeed highly implausible that individual counties could have manipulated their assigned poverty index. Our findings are robust to different bandwidth and local polynomial order specifications. Finally, we note two theory-based empirical findings: (i) our proposed manipulation test employs robust bias-corrected methods, and hence leads to asymmetric confidence intervals (not necessarily centered around the density point estimator); and (ii) the effective sample size of the original McCrary test is much smaller than our proposed manipulation test because of the pre-binning of the data, and hence can lead to important reduction in power of the test.

6 Conclusion

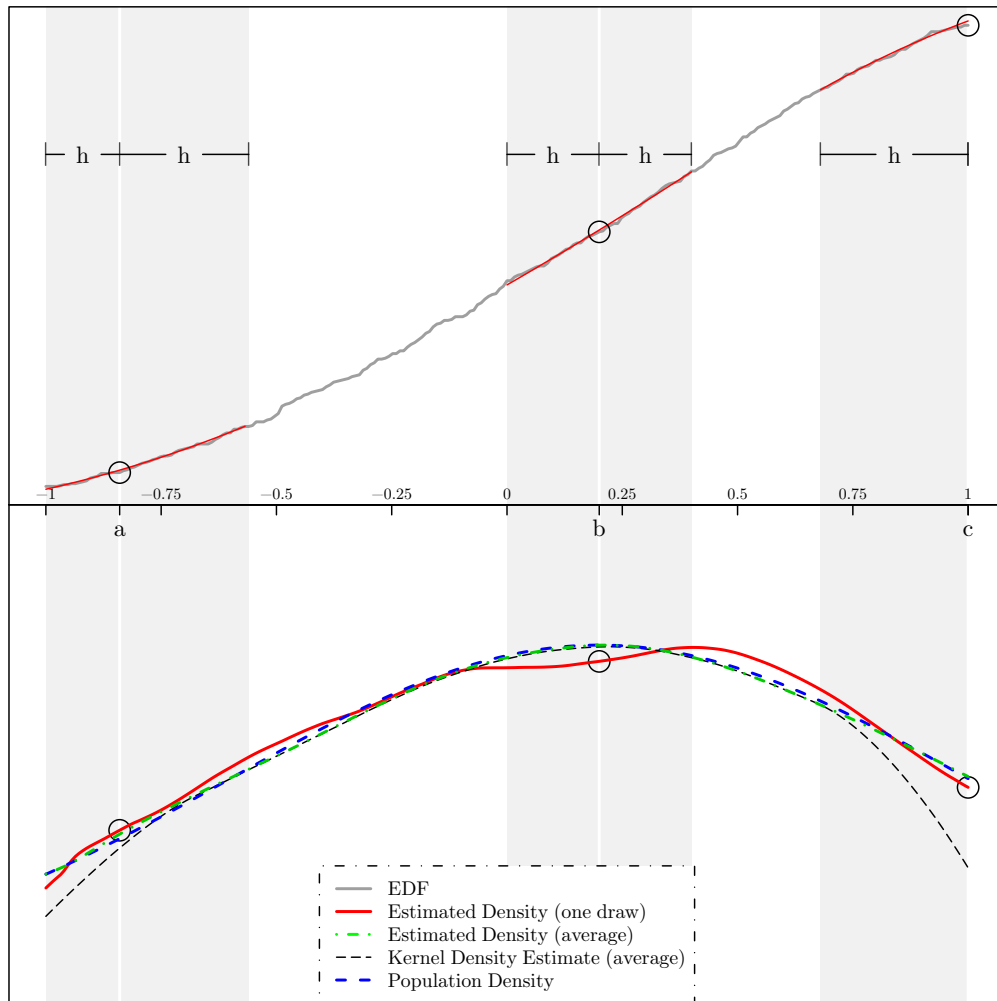
We introduced a boundary adaptive kernel-based density estimator employing local polynomial methods, which requires choosing only one tuning parameter and does not require boundary-specific data transformations (such as pre-binning). We studied the main asymptotic properties of the estimator, and used these results to developed a new manipulation test via discontinuity in density testing. Several extensions and generalizations of our results are underway in ongoing work, and two distinct general purpose software packages in `Stata` and `R` are readily available [Cattaneo et al. \(2018, 2019\)](#).

References

- Abadie, A., and Cattaneo, M. D. (2018), “Econometric Methods for Program Evaluation,” *Annual Review of Economics*, 10, 465–503.
- Arai, Y., and Ichimura, H. (2018), “Simultaneous Selection of Optimal Bandwidths for the Sharp Regression Discontinuity Estimator,” *Quantitative Economics*, 9, 441–482.
- Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018), “On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference,” *Journal of the American Statistical Association*, 113, 767–779.
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014), “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs,” *Econometrica*, 82, 2295–2326.
- Cattaneo, M. D., Jansson, M., and Ma, X. (2018), “Manipulation Testing based on Density Discontinuity,” *Stata Journal*, 18, 234–261.
- (2019), “`lpdensity`: Local Polynomial Density Estimation and Inference,” working paper.
- Cattaneo, M. D., Titiunik, R., and Vazquez-Bare, G. (2017), “Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality,” *Journal of Policy Analysis and Management*, 36, 643–681.
- Cheng, M.-Y., Fan, J., and Marron, J. S. (1997), “On Automatic Boundary Corrections,” *Annals of Statistics*, 25, 1691–1708.
- Dong, Y., Lee, Y.-Y., and Gou, M. (2019), “Regression Discontinuity Designs with a Continuous Treatment,” SSRN working paper No. 3167541.
- Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, New York: Chapman & Hall/CRC.
- Ganong, P., and Jäger, S. (2018), “A Permutation Test for the Regression Kink Design,” *Journal of the American Statistical Association*, 113, 494–504.
- Hyttinen, A., Meriläinen, J., Saarimaa, T., Toivanen, O., and Tukiainen, J. (2018), “When Does Regression Discontinuity Design Work? Evidence from Random Election Outcomes,” *Quantitative Economics*, 9, 1019–1051.
- Imbens, G., and Lemieux, T. (2008), “Regression Discontinuity Designs: A Guide to Practice,” *Journal of Econometrics*, 142, 615–635.
- Imbens, G. W., and Rubin, D. B. (2015), *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press.
- Karunamuni, R., and Alberts, T. (2005), “On Boundary Correction in Kernel Density Estimation,” *Statistical Methodology*, 2, 191–212.
- Lee, D. S., and Lemieux, T. (2010), “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, 48, 281–355.
- Li, K.-C. (1987), “Asymptotic Optimality for C_p , C_L , Cross-validation and Generalized Cross-validation: Discrete Index Set,” *Annals of Statistics*, 15, 958–975.

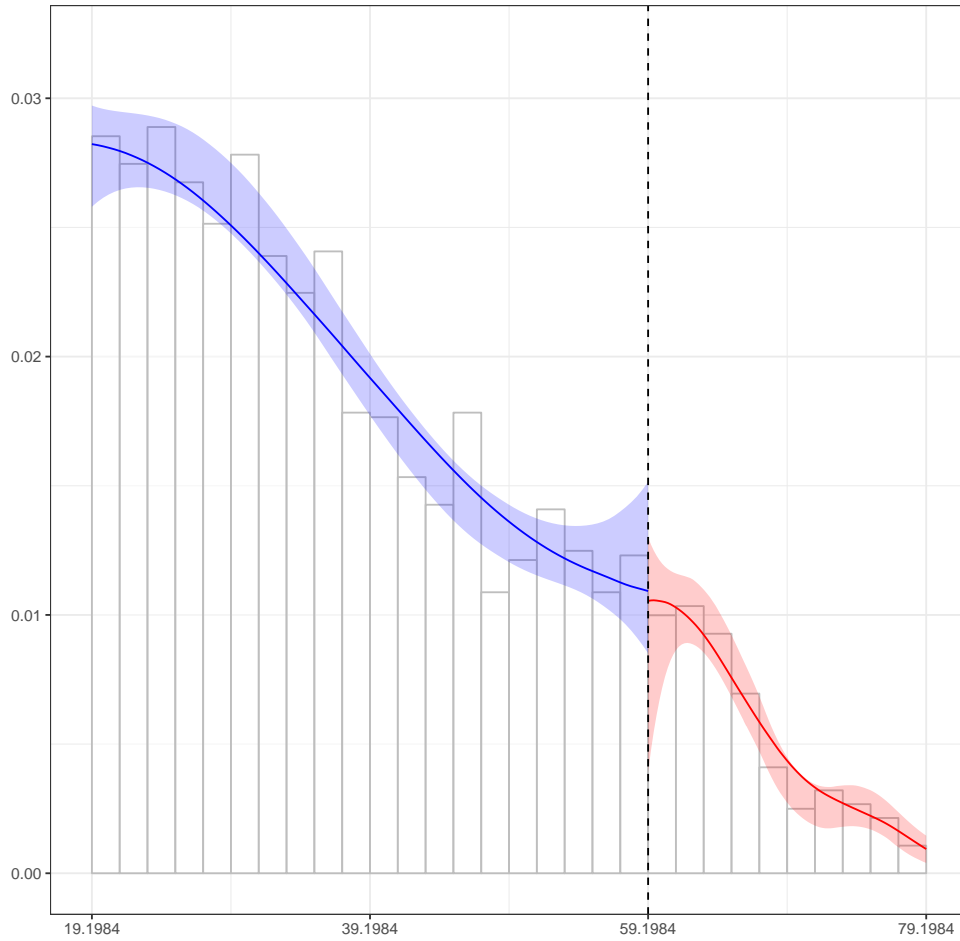
- Ludwig, J., and Miller, D. L. (2007), “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design,” *Quarterly Journal of Economics*, 122, 159–208.
- McCrary, J. (2008), “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test,” *Journal of Econometrics*, 142, 698–714.
- Otsu, T., Xu, K.-L., and Matsushita, Y. (2014), “Estimation and Inference of Discontinuity in Density,” *Journal of Business and Economic Statistics*, 31, 507–524.
- Wand, M., and Jones, M. (1995), *Kernel Smoothing*, Florida: Chapman & Hall/CRC.
- Zhang, S., and Karunamuni, R. J. (1998), “On Kernel Density Estimation Near Endpoints,” *Journal of Statistical Planning and Inference*, 70, 301–316.

Figure 1: Graphical Illustration of Density Estimator.



Notes: (i) Constructed using companion R (and Stata) package described in Cattaneo et al. (2019) with simulated data.

Figure 2: Manipulation Testing, Head Start Data.



Notes: (i) Histogram estimate (light grey in background) of the running variable (poverty index) computed with default values in **R**; (ii) local polynomial density estimate (solid blue and red) and robust bias corrected confidence intervals (shaded blue and red) computed using companion **R** (and **Stata**) package described in [Cattaneo et al. \(2018\)](#); and (iii) $n_- = 2,504$, $n_+ = 300$, and $\bar{x} = 59.1984$.

Table 1: Manipulation Testing, Head Start Data.

	Pre-binning		Bandwidths		Eff. n		Test	
	left	right	left	right	left	right	T	p -val
$h_- \neq h_+$								
$T_2(\hat{h}_1)$			15.771	2.326	581	65	0.024	0.981
$T_3(\hat{h}_2)$			19.776	8.296	762	210	-1.146	0.252
$T_4(\hat{h}_3)$			32.487	10.808	1598	232	-1.083	0.279
$h_- = h_+$								
$T_2(\hat{h}_1)$			3.274	3.274	99	95	-1.355	0.175
$T_3(\hat{h}_2)$			9.213	9.213	316	221	-0.515	0.607
$T_4(\hat{h}_3)$			12.270	12.270	419	243	-0.712	0.477
McCrary	76	60	13.950	13.950	24	24	0.142	0.887

Notes: (i) $T_p(h)$ denotes the manipulation test statistic using p -th order density estimators with bandwidth choice h (which could be common on both sides or different on either side of the cutoff), and \hat{h}_p denotes the estimated MSE-optimal bandwidths for p -th order density estimator or difference of estimators (depending on the case considered); (ii) Columns under “Bandwidths” report estimated MSE-optimal bandwidths, Columns under “Eff. n ” report effective sample size on either side of the cutoff, and Columns under “Test” report value of test statistic (T) and two-sided p -value (p -val); (iii) first three rows allow for different bandwidths on each side of the cutoff, while the next three rows employ a common bandwidth on both sides of the cutoff (chosen to be MSE-optimal for the difference of density estimates). All estimates are obtained using companion **R** (and **Stata**) package described in Cattaneo et al. (2018); and (iv) the last row, labeled “McCrary”, corresponds to the original implementation of McCrary (2008), and therefore columns under “Pre-binning” report the total number of bins used for pre-bining of the data and columns under “Eff. n ” report the number of bins used for local linear density estimation.

Simple Local Polynomial Density Estimators Supplemental Appendix

Matias D. Cattaneo*

Michael Jansson[†]

Xinwei Ma[‡]

June 11, 2019

Abstract

This Supplemental Appendix contains general theoretical results and their proofs, which encompass those discussed in the main paper, discusses additional methodological and technical results, and reports simulation evidence.

arXiv:1811.11512v2 [econ.EM] 7 Jun 2019

*Department of Operations Research and Financial Engineering, Princeton University.

[†]Department of Economics, UC Berkeley and *CREATES*.

[‡]Department of Economics, UC San Diego.

Contents

1	Setup	1
1.1	Overview of Main Results	2
1.2	Some Matrices.....	3
2	Large Sample Properties	3
2.1	Assumptions.....	3
2.2	Preliminary Lemmas	4
2.3	Main Results.....	6
3	Bandwidth Selection.....	7
3.1	For Density and Derivatives Estimates ($v \geq 1$).....	7
3.2	For CDF Estimate ($v = 0$).....	10
4	Application to Manipulation Testing	11
4.1	Unrestricted Model	11
4.2	Restricted Model.....	14
5	Other Standard Error Estimators	16
5.1	Plug-in Standard Error	16
5.2	Jackknife-based Standard Error.....	16
6	Simulation Study.....	17
6.1	DGP 1: Truncated Normal Distribution.....	17
6.2	DGP 2: Exponential Distribution.....	18
	References.....	19
7	Proof	20
7.1	Proof of Lemma 1	20
7.2	Proof of Lemma 2	20
7.3	Proof of Lemma 3	21
7.4	Proof of Lemma 4	22
7.5	Proof of Theorem 1.....	22
7.6	Proof of Theorem 2.....	23
7.7	Proof of Lemma 5	24
7.8	Proof of Lemma 6	24
7.9	Proof of Theorem 3.....	25
7.10	Proof of Lemma 7	25
7.11	Proof of Lemma 8	25
7.12	Proof of Lemma 9	25
7.13	Proof of Lemma 10	27
7.14	Proof of Corollary 1	27
7.15	Proof of Lemma 11	27
7.16	Proof of Lemma 12	27
7.17	Proof of Lemma 13	27
7.18	Proof of Lemma 14	30
7.19	Proof of Corollary 2.....	30

1 Setup

We repeat the setup in the main paper for completeness. Recall that $\{x_i\}_{1 \leq i \leq n}$ is a random sample from the cumulative distribution function (hereafter CDF) F , supported on $\mathcal{X} = [x_L, x_U]$. Note that it is possible to have $x_L = -\infty$ and/or $x_U = \infty$. We will assume both x_L and x_U are finite, to facilitate discussion on boundary estimation issues.

Define the empirical distribution function (hereafter EDF)

$$\tilde{F}(x) = \frac{1}{n} \sum_i \mathbf{1}[x_i \leq x].$$

Note that in the main paper, we use $\hat{F}(\cdot)$ to denote the above EDF. We avoid such notation in this Supplemental Appendix, and instead use $\tilde{F}(\cdot)$, because a (smoothed) CDF estimator can be obtained from our local polynomial approach.

Given $p \in \mathbb{N}$, our local polynomial distribution estimator is defined as

$$\hat{\beta}_p(x) = \arg \min_{\mathbf{b} \in \mathbb{R}^{p+1}} \sum_i \left(\tilde{F}(x_i) - \mathbf{r}_p(x_i - x)' \mathbf{b} \right)^2 K \left(\frac{x_i - x}{h} \right),$$

where $\mathbf{r}_p(u) = [1, u, u^2, \dots, u^p]$ is a (one-dimensional) polynomial expansion; K is a kernel function whose properties are to be specified later; $h = h_n$ is a bandwidth sequence. The estimator, $\hat{\beta}_p(x)$, is motivated as a local Taylor series expansion, hence the target parameter is (i.e., the population counterpart, assuming exists)

$$\beta_p(x) = \left[\frac{1}{0!} F(x), \frac{1}{1!} F^{(1)}(x), \dots, \frac{1}{p!} F^{(p)}(x) \right]'$$

Therefore, we also write

$$\hat{\beta}_p(x) = \left[\frac{1}{0!} \hat{F}_p(x), \frac{1}{1!} \hat{F}_p^{(1)}(x), \dots, \frac{1}{p!} \hat{F}_p^{(p)}(x) \right]'$$

or equivalently, $\hat{F}_p^{(v)} = v! \mathbf{e}'_v \hat{\beta}_p(x)$, provided that $v \leq p$, and \mathbf{e}_v is the $(v+1)$ -th unit vector of \mathbb{R}^{p+1} . (The subscript p is omitted in the main paper to economize notation.) We also use $f = F^{(1)}$ to denote the corresponding probability density function (hereafter PDF) for convenience.

The estimator has the following matrix form:

$$\hat{\beta}_p(x) = \mathbf{H}^{-1} \left(\frac{1}{n} \mathbf{X}'_h \mathbf{K}_h \mathbf{X}_h \right)^{-1} \left(\frac{1}{n} \mathbf{X}'_h \mathbf{K}_h \mathbf{Y} \right), \quad \mathbf{X}_h = \left[\left(\frac{x_i - x}{h} \right)^j \right]_{1 \leq i \leq n, 0 \leq j \leq p}$$

where \mathbf{K}_h is a diagonal matrix collecting $\{h^{-1} K((x_i - x)/h)\}_{1 \leq i \leq n}$, and \mathbf{Y} is a column vector collecting $\{\tilde{F}(x_i)\}_{1 \leq i \leq n}$. We also adopt the convention $K_h(u) = h^{-1} K(u/h)$.

In this Supplemental Appendix, we use n to denote sample size, and limits are taken with $n \rightarrow \infty$, unless otherwise specified. The standard Euclidean norm is denoted by $|\cdot|$, and other norms will be defined at their first appearances. Maximum and minimum of two real numbers a and b are denoted by $a \vee b$ and $a \wedge b$, respectively. For sequence of numbers (or random variables), $a_n \lesssim b_n$ implies $\limsup_n |a_n/b_n|$ is finite, and $a_n \asymp b_n$ implies both directions. The notation $a_n \lesssim_{\mathbb{P}} b_n$ is used

to denote that $|a_n/b_n|$ is asymptotically tight: $\limsup_{\varepsilon \uparrow \infty} \limsup_n \mathbb{P}[|a_n/b_n| \geq \varepsilon] = 0$. $a_n \asymp_{\mathbb{P}} b_n$ implies both $a_n \lesssim_{\mathbb{P}} b_n$ and $b_n \lesssim_{\mathbb{P}} a_n$. When b_n is a sequence of nonnegative numbers, $a_n = O(b_n)$ is sometimes used for $a_n \lesssim b_n$, so does $a_n = O_{\mathbb{P}}(b_n)$. For probabilistic convergence, we use $\rightarrow_{\mathbb{P}}$ for convergence in probability and \rightsquigarrow for weak convergence (convergence in distribution). Standard normal distribution is denoted as $\mathcal{N}(0, 1)$, with CDF Φ and PDF ϕ . Throughout, we use C to denote generic constants which do not depend on sample size. The exact value can change given the context.

1.1 Overview of Main Results

In this subsection, we give an overview of our results, including a (first order) mean squared error (hereafter MSE) expansion, and asymptotic normality. Fix some $v \geq 1$ and p , we have the following:

$$\left| \hat{F}_p^{(v)}(x) - F^{(v)}(x) \right| = O_{\mathbb{P}} \left(h^{p+1-v} \mathcal{B}_{p,v}(x) + h^{p+2-v} \tilde{\mathcal{B}}_{p,v}(x) + \sqrt{\frac{1}{nh^{2v-1}} \mathcal{V}_{p,v}(x)} \right).$$

The previous result gives MSE expansion for derivative estimators, $1 \leq v \leq p$, but not for $v = 0$. With $v = 0$, $\hat{F}_p(x)$ is essentially a smoothed EDF, which estimates the CDF $F(x)$. Since $F(x)$ is \sqrt{n} -estimable, one should be expected that the estimated distribution function will have very different properties compared to the estimated derivatives. Indeed, we have

$$\left| \hat{F}_p(x) - F(x) \right| = O_{\mathbb{P}} \left(h^{p+1} \mathcal{B}_{p,0}(x) + h^{p+2} \tilde{\mathcal{B}}_{p,0}(x) + \sqrt{\frac{1}{n} \mathcal{V}_{p,0}(x)} \right).$$

There is another complication, however, when x is in the boundary region. For a drifting sequence x in the boundary region, the EDF $\tilde{F}(x)$ is “super-consistent” in the sense that it converges at rate $\sqrt{h/n}$. The reason is that when x is near x_L or x_U , $\tilde{F}(x)$ is essentially estimating 0 or 1, and the variance, $F(x)(1 - F(x))$ vanishes asymptotically, giving rise to the additional factor \sqrt{h} . This is shared by our estimator: for $v = 0$ and x in the boundary region, the CDF estimator $\tilde{F}_p(x)$ is super-consistent, with $\mathcal{V}_{p,0}(x) \asymp h$.

Also note that for the MSE expansion, we provide not only the first order bias but also the second order bias. The second order bias will be used for bandwidth selection, since it is well-known that in some cases the first order bias can vanish. (More precisely, when x is an interior evaluation point and $p - v$ is even. See, for example, Fan and Gijbels 1996.)

The MSE expansion provides the rate of convergence of our estimator. The following shows that, under suitable regularity conditions, they are also asymptotically normal. Again first consider $v \geq 1$.

$$\sqrt{nh^{2v-1}} \left(\hat{F}_p^{(v)}(x) - F^{(v)}(x) - h^{p+1-v} \mathcal{B}_{p,v}(x) \right) \rightsquigarrow \mathcal{N}\left(0, \mathcal{V}_{p,v}(x)\right),$$

provided that the bandwidth is not too large, so that after scaling, the remaining bias does not feature in first-order asymptotics. For $v = 0$, i.e. the smoothed EDF, we have

$$\sqrt{\frac{n}{\mathcal{V}_{p,0}(x)}} \left(\hat{F}_p(x) - F(x) - h^{p+1} \mathcal{B}_{p,0}(x) \right) \rightsquigarrow \mathcal{N}\left(0, 1\right),$$

where we moved the variance $\mathcal{V}_{p,0}(x)$ as a scaling factor in the above display, to encompass the situation where x lies near boundaries.

1.2 Some Matrices

In this subsection we collect some matrices which will be used throughout this Supplemental Appendix. They show up in asymptotic results as components of the (leading) bias and variance. Note that x can be either a fixed point, or it can be a drifting sequence to capture the issue of estimation and inference in boundary regions. For the latter case, x takes the form $x = x_L + ch$ or $x = x_U - ch$ for some $c \in [0, 1]$.

Define

$$\begin{aligned} \mathbf{S}_{p,x} &= \int_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}} \mathbf{r}_p(u) \mathbf{r}_p(u)' K(u) du, & \mathbf{c}_{p,x} &= \int_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}} \mathbf{r}_p(u) u^{p+1} K(u) du, & \check{\mathbf{c}}_{p,x} &= \int_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}} \mathbf{r}_p(u) u^{p+2} K(u) du, \\ \mathbf{\Gamma}_{p,x} &= \iint_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}} (u \wedge v) \mathbf{r}_p(u) \mathbf{r}_p(v) K(u) K(v) dudv, & \mathbf{T}_{p,x} &= \int_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}} \mathbf{r}_p(u) \mathbf{r}_p(u)' K(u)^2 du. \end{aligned}$$

Later we will assume that the kernel function K is supported on $[-1, 1]$, hence with a shrinking bandwidth sequence $h \downarrow 0$, the region of integration in the above display can be replaced by

x	$(x_L - x)/h$	$(x_U - x)/h$
x interior	-1	+1
$x = x_L + ch$ in lower boundary	- c	+1
$x = x_U - ch$ in upper boundary	-1	+ c

Since we do not allow $x_L = x_U$, no drifting sequence x can be in both lower and upper boundary regions, at least in large samples.

2 Large Sample Properties

2.1 Assumptions

In this section we give assumptions, preliminary lemmas and our main results. Other assumptions specific to certain results will be given in corresponding sections.

Let \mathcal{O} be a connected subset of \mathbb{R} with nonempty interior, $\mathcal{C}^s(\mathcal{O})$ denotes functions that are at least s -times continuously differentiable in the interior of \mathcal{O} , and that the derivatives can be continuously extended to the boundary of \mathcal{O} .

Assumption 1 (DGP).

$\{x_i\}_{1 \leq i \leq n}$ is a random sample from distribution F , supported on $\mathcal{X} = [x_L, x_U]$. Further, $F \in \mathcal{C}^{\alpha_x}(\mathcal{X})$ for some $\alpha_x \geq 1$, and $f(x) = F^{(1)}(x) > 0$ for all $x \in \mathcal{X}$.

Assumption 2 (Kernel).

The kernel function $K(\cdot)$ is nonnegative, symmetric, and belongs to $\mathcal{C}^0([-1, 1])$. Further, it integrates to one: $\int_{\mathbb{R}} K(u) du = 1$.

2.2 Preliminary Lemmas

We first consider the object $\mathbf{X}'_h \mathbf{K}_h \mathbf{X}_h / n$

Lemma 1. *Assume Assumptions 1 and 2 hold, $h \rightarrow 0$ and $nh \rightarrow \infty$. Then*

$$\frac{1}{n} \mathbf{X}'_h \mathbf{K}_h \mathbf{X}_h = f(x) \mathbf{S}_{p,x} + o(1) + O_{\mathbb{P}}\left(1/\sqrt{nh}\right).$$

Lemma 1 shows that the matrix $\mathbf{X}'_h \mathbf{K}_h \mathbf{X}_h / n$ is asymptotically invertible. Also note that this result covers both interior and boundary evaluation point x , and depending on the nature of x , the exact form of $\mathbf{S}_{p,x}$ differs.

With simple algebra, one has

$$\hat{\beta}_p(x) - \beta_p(x) = \mathbf{H}^{-1} \left(\frac{1}{n} \mathbf{X}'_h \mathbf{K}_h \mathbf{X}_h \right)^{-1} \left(\frac{1}{n} \mathbf{X}'_h \mathbf{K}_h (\mathbf{Y} - \mathbf{X} \beta_p(x)) \right),$$

and the following gives a further decomposition of the ‘‘numerator.’’

$$\begin{aligned} \frac{1}{n} \mathbf{X}'_h \mathbf{K}_h (\mathbf{Y} - \mathbf{X} \beta_p(x)) &= \frac{1}{n} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\tilde{F}(x_i) - \mathbf{r}_p(x_i - x)' \beta_p(x) \right) K_h(x_i - x) \\ &= \frac{1}{n} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(F(x_i) - \mathbf{r}_p(x_i - x)' \beta_p(x) \right) K_h(x_i - x) \\ &\quad + \int_{\frac{x_l - x}{h}}^{\frac{x_u - x}{h}} \mathbf{r}_p(u) \left(\tilde{F}(x + hu) - F(x + hu) \right) K(u) f(x + hu) du \\ &\quad + \frac{1}{n} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\tilde{F}(x_i) - F(x_i) \right) K_h(x_i - x) - \int_{\frac{x_l - x}{h}}^{\frac{x_u - x}{h}} \mathbf{r}_p(u) \left(\tilde{F}(x + hu) - F(x + hu) \right) K(u) f(x + hu) du. \end{aligned}$$

The first part represents the smoothing bias, and the second part can be analyzed as a sample average. The real challenge comes from the third term, which can have a nonnegligible (first order) contribution. We further decompose it as

$$\begin{aligned} \frac{1}{n} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\tilde{F}(x_i) - F(x_i) \right) K_h(x_i - x) &= \frac{1}{n^2} \sum_{i,j} \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\mathbb{1}[x_j \leq x_i] - F(x_i) \right) K_h(x_i - x) \\ &= \frac{1}{n^2} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(1 - F(x_i) \right) K_h(x_i - x) + \frac{1}{n^2} \sum_{i,j;i \neq j} \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\mathbb{1}[x_j \leq x_i] - F(x_i) \right) K_h(x_i - x). \end{aligned}$$

As a result,

$$\begin{aligned} \frac{1}{n} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\tilde{F}(x_i) - \mathbf{r}_p(x_i - x)' \beta_p(x) \right) K_h(x_i - x) & \\ &= \frac{1}{n} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(F(x_i) - \mathbf{r}_p(x_i - x)' \beta_p(x) \right) K_h(x_i - x) \quad (\text{smoothing bias } \hat{\mathbf{B}}_S) \\ &\quad + \int_{\frac{x_l - x}{h}}^{\frac{x_u - x}{h}} \mathbf{r}_p(u) \left(\tilde{F}(x + hu) - F(x + hu) \right) K(u) f(x + hu) du \quad (\text{linear variance } \hat{\mathbf{L}}) \\ &\quad + \frac{1}{n^2} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(1 - F(x_i) \right) K_h(x_i - x) \quad (\text{leave-in bias } \hat{\mathbf{B}}_{LI}) \\ &\quad + \frac{1}{n^2} \sum_{i,j;i \neq j} \left\{ \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\mathbb{1}[x_j \leq x_i] - F(x_i) \right) K_h(x_i - x) \right. \\ &\quad \left. - \mathbb{E} \left[\mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\mathbb{1}[x_j \leq x_i] - F(x_i) \right) K_h(x_i - x) \middle| x_j \right] \right\}. \quad (\text{quadratic variance } \hat{\mathbf{R}}) \end{aligned}$$

To provide intuition for the above decomposition, the smoothing bias is a typical feature of non-parametric estimators; leave-in bias arises since each observation is used twice, in constructing the EDF \tilde{F} and as a design point (that is, \tilde{F} has to be evaluated at x_i); and a second order U-statistic shows up because the “dependent variable,” \mathbf{Y} , is estimated, which leads to double summation.

We first analyze the bias terms.

Lemma 2. *Assume Assumptions 1 and 2 hold with $\alpha_x \geq p + 1$, $h \rightarrow 0$ and $nh \rightarrow \infty$. Then*

$$\hat{\mathbf{B}}_s = h^{p+1} \frac{F^{(p+1)}(x)f(x)}{(p+1)!} \mathbf{c}_{p,x} + o_{\mathbb{P}}(h^{p+1}), \quad \hat{\mathbf{B}}_{\text{LI}} = O_{\mathbb{P}}(n^{-1}).$$

By imposing additional smoothness, it is also possible to characterize the next term in the smoothing bias, which has order h^{p+2} . We report the higher order bias in a later section as it is used for bandwidth selection.

Next we consider the “influence function” part, $\hat{\mathbf{L}}$. This term is crucial in the sense that (under suitable conditions so that $\hat{\mathbf{R}}$ becomes negligible) it determines the asymptotic variance of our estimator, and with correct scaling, it is asymptotically normally distributed.

Lemma 3. *Assume Assumptions 1 and 2 hold with $\alpha_x \geq 2$, $h \rightarrow 0$ and $nh \rightarrow \infty$. Define the scaling matrix*

$$\mathbf{N}_x = \begin{cases} \text{diag}\{1, h^{-1/2}, h^{-1/2}, \dots, h^{-1/2}\} & x \text{ interior,} \\ \text{diag}\{h^{-1/2}, h^{-1/2}, h^{-1/2}, \dots, h^{-1/2}\} & x \text{ boundary,} \end{cases}$$

then

$$\sqrt{n}\mathbf{N}_x [f(x)\mathbf{S}_{p,x}]^{-1} \hat{\mathbf{L}} \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{V}_{p,x}),$$

with

$$\mathbf{V}_{p,x} = \begin{cases} F(x)(1-F(x))\mathbf{e}_0\mathbf{e}'_0 + f(x)(\mathbf{I} - \mathbf{e}_0\mathbf{e}'_0)\mathbf{S}_{p,x}^{-1}\mathbf{\Gamma}_{p,x}\mathbf{S}_{p,x}^{-1}(\mathbf{I} - \mathbf{e}_0\mathbf{e}'_0) & x \text{ interior} \\ f(x) (\mathbf{S}_{p,x}^{-1}\mathbf{\Gamma}_{p,x}\mathbf{S}_{p,x}^{-1} + \mathbf{c}\mathbf{e}_0\mathbf{e}'_0) & x = x_L + ch \\ f(x) (\mathbf{S}_{p,x}^{-1}\mathbf{\Gamma}_{p,x}\mathbf{S}_{p,x}^{-1} + \mathbf{c}\mathbf{e}_0\mathbf{e}'_0 - (\mathbf{e}_1\mathbf{e}'_0 + \mathbf{e}_0\mathbf{e}'_1)) & x = x_U - ch. \end{cases}$$

The scaling matrix depends on whether the evaluation point is located in the interior or boundary, which is a unique feature of our estimator. To see the intuition, consider an interior point x , and recall that the first element of $\hat{\boldsymbol{\beta}}_p(x)$ is the smoothed EDF, which is \sqrt{n} -estimable. Therefore, the property of $\hat{F}_p(x)$ is very different from those of the estimated density and higher order derivatives.

When x is either in the lower or upper boundary region, $\hat{F}_p(x)$ essentially estimates 0 or 1, respectively, hence it is super-consistent in the sense that it converges even faster than $1/\sqrt{n}$. In this case, the leading $1/\sqrt{n}$ -variance vanishes, and higher order residual noise dominates, which makes $\hat{F}_p(x)$ no longer independent of the estimated density and derivatives, justifying the formula of boundary evaluation points.

Finally we consider the second order U-statistic component.

Lemma 4. *Assume Assumptions 1 and 2 hold, $h \rightarrow 0$ and $nh \rightarrow \infty$. Then*

$$\mathbb{V}[\hat{\mathbf{R}}] = \frac{2}{n^2 h} f(x)F(x)(1-F(x))\mathbf{T}_{p,x} + O(n^{-2}).$$

In particular, when x is in the boundary region, the above has order $O(n^{-2})$.

2.3 Main Results

In this section we provide two main results, one on asymptotic normality, and the other on standard error.

Theorem 1 (Asymptotic Normality). *Assume Assumptions 1 and 2 hold with $\alpha_x \geq p+1$ for some integer $p \geq 0$. Further $h \rightarrow 0$, $nh^2 \rightarrow \infty$ and $nh^{2p+1} = O(1)$. Then*

$$\begin{aligned} \sqrt{nh^{2v-1}} \left(\hat{F}_p^{(v)}(x) - F^{(v)}(x) - h^{p+1-v} \mathcal{B}_{p,v}(x) \right) &\rightsquigarrow \mathcal{N}\left(0, \mathcal{V}_{p,v}(x)\right), \quad 1 \leq v \leq p, \\ \sqrt{\frac{n}{\mathcal{V}_{p,0}(x)}} \left(\hat{F}_p(x) - F(x) - h^{p+1} \mathcal{B}_{p,0}(x) \right) &\rightsquigarrow \mathcal{N}\left(0, 1\right). \end{aligned}$$

The constants are

$$\mathcal{B}_{p,v}(x) = v! \frac{F^{(p+1)}(x)}{(p+1)!} \mathbf{e}'_v \mathbf{S}_{p,x}^{-1} \mathbf{c}_{p,x},$$

and

$$\mathcal{V}_{p,v}(x) = \begin{cases} (v!)^2 f(x) \mathbf{e}'_v \mathbf{S}_{p,x}^{-1} \mathbf{\Gamma}_{p,x} \mathbf{S}_{p,x}^{-1} \mathbf{e}_v & 1 \leq v \leq p \\ F(x)(1-F(x)) & v = 0, x \text{ interior} \\ hf(x) (\mathbf{e}'_0 \mathbf{S}_{p,x}^{-1} \mathbf{\Gamma}_{p,x} \mathbf{S}_{p,x}^{-1} \mathbf{e}_0 + c) & v = 0, x = x_L + ch \text{ or } x_U - ch. \end{cases}$$

Remark 1 (On $nh^{2p+1} = O(1)$). This condition ensures that higher order bias, after scaling, is asymptotically negligible. \parallel

Remark 2 (On $nh^2 \rightarrow \infty$). This condition ensures that the second order U-statistic, $\hat{\mathbf{R}}$, has smaller order compared to $\hat{\mathbf{L}}$. Note that this condition can be dropped for boundary x or when the parameter of interest is the CDF \hat{F}_p . \parallel

Now we provide a standard error, which is also boundary adaptive. Given the formula in Theorem 1, it is possible to estimate the asymptotic variance by plugging in unknown quantities regarding the data generating process. For example consider $\mathcal{V}_{p,1}(x)$ for the estimated density. Assume the researcher knows the location of the boundary x_L and x_U , the matrices $\mathbf{S}_{p,x}$ and $\mathbf{\Gamma}_{p,x}$ can be constructed with numerical integration, since they are related to features of the kernel function, not the data generating process. The unknown density $f(x)$ can also be replaced by its estimate, as long as $p \geq 1$.

Another approach is to utilize the decomposition of the estimator, in particular the $\hat{\mathbf{L}}$ term. To

introduce our variance estimator, we make the following definitions.

$$\begin{aligned}\hat{\mathbf{S}}_{p,x} &= \frac{1}{n} \mathbf{X}_h \mathbf{K}_h \mathbf{X}_n = \frac{1}{n} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \mathbf{r}_p \left(\frac{x_i - x}{h} \right)' K_h(x_i - x) \\ \hat{\mathbf{\Gamma}}_{p,x} &= \frac{1}{n^3} \sum_{i,j,k} \mathbf{r}_p \left(\frac{x_j - x}{h} \right) \mathbf{r}_p \left(\frac{x_k - x}{h} \right)' K_h(x_j - x) K_h(x_k - x) \left(\mathbf{1}[x_i \leq x_j] - \tilde{F}(x_j) \right) \left(\mathbf{1}[x_i \leq x_k] - \tilde{F}(x_k) \right).\end{aligned}$$

Following is the main result regarding variance estimation. It is automatic and fully-adaptive, in the sense that no knowledge about the boundary location is needed.

Theorem 2 (Variance Estimation).

Assume Assumptions 1 and 2 hold with $\alpha_x \geq p + 1$ for some integer $p \geq 0$. Further $h \rightarrow 0$, $nh^2 \rightarrow \infty$ and $nh^{2p+1} = O(1)$. Then

$$\hat{\mathcal{V}}_{p,v}(x) \equiv (v!)^2 \mathbf{e}'_v \mathbf{N}_x \hat{\mathbf{S}}_{p,x}^{-1} \hat{\mathbf{\Gamma}}_{p,x} \hat{\mathbf{S}}_{p,x}^{-1} \mathbf{N}_x \mathbf{e}_v \xrightarrow{\mathbb{P}} \mathcal{V}_{p,v}(x).$$

Define the standard error as

$$\hat{\sigma}_{p,v}(x) \equiv (v!) \sqrt{\frac{1}{nh^{2v}} \mathbf{e}'_v \hat{\mathbf{S}}_{p,x}^{-1} \hat{\mathbf{\Gamma}}_{p,x} \hat{\mathbf{S}}_{p,x}^{-1} \mathbf{e}_v},$$

then

$$\hat{\sigma}_{p,v}(x)^{-1} \left(\hat{F}_p^{(v)}(x) - F^{(v)}(x) - h^{p+1-v} \mathcal{B}_{p,v}(x) \right) \rightsquigarrow \mathcal{N}(0, 1).$$

3 Bandwidth Selection

In this section we consider the problem of constructing MSE-optimal bandwidth for our local polynomial regression-based distribution estimators. We focus exclusively on the case $v \geq 1$, hence the object of interest will be either the density function or derivatives thereof. Valid bandwidth choice for the distribution function $\hat{F}_p(x)$ is also an interesting topic, but difficulty arises since it is estimated at the parametric rate. We will briefly mention MSE expansion of the estimated CDF at the end.

3.1 For Density and Derivatives Estimates ($v \geq 1$)

Consider some $1 \leq v \leq p$, the following lemma gives finer characterization of the bias.

Lemma 5. Assume Assumptions 1 and 2 hold with $\alpha_x \geq p + 2$, $h \rightarrow 0$ and $nh^3 \rightarrow \infty$. Then the leading bias of $\hat{F}_p^{(v)}(x)$ is

$$h^{p+1-v} \mathcal{B}_{p,v}(x) = h^{p+1-v} \left\{ \frac{F^{(p+1)}(x)}{(p+1)!} v! \mathbf{e}'_v \mathbf{S}_{p,x}^{-1} \mathbf{c}_{p,x} + h \left(\frac{F^{(p+2)}(x)}{(p+2)!} + \frac{F^{(p+1)}(x)}{(p+1)!} \frac{F^{(2)}(x)}{f(x)} \right) v! \mathbf{e}'_v \mathbf{S}_{p,x}^{-1} \tilde{\mathbf{c}}_{p,x} \right\}.$$

The above lemma is a refinement of Lemma 1 and 2, and characterizes the higher-order bias. To see its necessity, we note that when $p - v$ is even and x is an interior evaluation point, the leading bias is zero. This is because $\mathbf{e}'_v \mathbf{S}_{p,x}^{-1} \mathbf{c}_{p,x}$ is zero, which is explained in Fan and Gijbels (1996). Except for rare cases such as $F^{(p+1)}(x) = 0$ or $F^{(p+2)}(x) = 0$, we have

Order of bias: $h^{p+1-v}\mathcal{B}_{p,v}(x) \asymp$

	$p - v$ odd	even
x interior	h^{p+1-v}	h^{p+2-v}
boundary	h^{p+1-v}	h^{p+1-v}

Note that for boundary evaluation points, the leading bias never vanishes.

The leading variance is also characterized by Theorem 1, and we reproduce it here:

$$\frac{1}{nh^{2v-1}}\mathcal{V}_{p,v}(x) = \frac{1}{nh^{2v-1}}(v!)^2 f(x)\mathbf{e}'_v \mathbf{S}_{p,x}^{-1} \mathbf{\Gamma}_{p,x} \mathbf{S}_{p,x}^{-1} \mathbf{e}_v.$$

The MSE-optimal bandwidth is defined as a minimizer of the following

$$h_{p,v}(x) = \arg \min_{h>0} \left[\frac{1}{nh^{2v-1}}\mathcal{V}_{p,v}(x) + h^{2p+2-2v}\mathcal{B}_{p,v}(x)^2 \right].$$

Given the discussion we had earlier on the bias, it is easy to see that the MSE-optimal bandwidth has the following asymptotic order:

Order of MSE-optimal bandwidth: $h_{p,v}(x) \asymp$

	$p - v$ odd	even
x interior	$n^{-\frac{1}{2p+1}}$	$n^{-\frac{1}{2p+3}}$
boundary	$n^{-\frac{1}{2p+1}}$	$n^{-\frac{1}{2p+1}}$

Again only the case where $p - v$ is even and x is interior needs special attention.

There are two notions of bandwidth consistency. Let h be some non-stochastic bandwidth sequence, and \hat{h} be an estimated bandwidth. Then \hat{h} is consistent *in rate* if $\hat{h} \asymp h$ (in most cases it is even true that $\hat{h}/h \rightarrow_{\mathbb{P}} C \in (0, \infty)$). And \hat{h} is consistent *in rate and constant* if $\hat{h}/h \rightarrow_{\mathbb{P}} 1$.

To construct consistent bandwidth, either rate consistent or consistent in both rate and constant, we need estimates of both the bias and variance. The variance part is relatively easy, as we have already demonstrated in Theorem 2:

$$n\ell^{2v-1} \frac{\hat{\sigma}_{p,v}(x)^2}{\mathcal{V}_{p,v}(x)} \rightarrow_{\mathbb{P}} 1,$$

where ℓ is some preliminary bandwidth used to construct $\hat{\sigma}_{p,v}(x)$.

To introduce our bias estimate, first assume there are consistent estimators for $F^{(p+1)}(x)$ and $F^{(p+2)}(x)$, denoted by $\hat{F}^{(p+1)}(x)$ and $\hat{F}^{(p+2)}(x)$. They can be obtained, for example, using our local polynomial regression-based approach, or can be constructed with some reference model (such as the normal distribution). The critical step is to obtain consistent estimators of the matrices, which are given in the following lemma.

Lemma 6. *Assume Assumptions 1 and 2 hold, $\ell \rightarrow 0$ and $n\ell \rightarrow \infty$. Then*

$$\widehat{\mathbf{S}_{p,x}^{-1} \mathbf{c}_{p,x}} = \left(\frac{1}{n} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{\ell} \right) \mathbf{r}_p \left(\frac{x_i - x}{\ell} \right)' K_{\ell}(x_i - x) \right)^{-1} \left(\frac{1}{n} \sum_i \left(\frac{x_i - x}{\ell} \right)^{p+1} \mathbf{r}_p \left(\frac{x_i - x}{\ell} \right) K_{\ell}(x_i - x) \right) \\ \rightarrow_{\mathbb{P}} \mathbf{S}_{p,x}^{-1} \mathbf{c}_{p,x},$$

and

$$\widehat{\mathbf{S}}_{p,x}^{-1} \tilde{\mathbf{c}}_{p,x} = \left(\frac{1}{n} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{\ell} \right) \mathbf{r}_p \left(\frac{x_i - x}{\ell} \right)' K_\ell(x_i - x) \right)^{-1} \left(\frac{1}{n} \sum_i \left(\frac{x_i - x}{\ell} \right)^{p+2} \mathbf{r}_p \left(\frac{x_i - x}{\ell} \right) K_\ell(x_i - x) \right) \\ \rightarrow_{\mathbb{P}} \mathbf{S}_{p,x}^{-1} \tilde{\mathbf{c}}_{p,x}.$$

Note that we used different notation, ℓ , as it corresponds to a preliminary bandwidth. Define

$$h^{p+1-v} \hat{\mathcal{B}}_{p,v}(x) = h^{p+1-v} \left\{ \frac{\hat{F}^{(p+1)}(x)}{(p+1)!} v! \mathbf{e}'_v \widehat{\mathbf{S}}_{p,x}^{-1} \mathbf{c}_{p,x} + h \frac{\hat{F}^{(p+2)}(x)}{(p+2)!} v! \mathbf{e}'_v \widehat{\mathbf{S}}_{p,x}^{-1} \tilde{\mathbf{c}}_{p,x} \right\},$$

and assume that $\hat{\sigma}_{p,v}(x)$ is constructed using the preliminary bandwidth ℓ . Then

$$\hat{h}_{p,v}(x) = \arg \min_{h>0} \left[\frac{\ell^{2v-1}}{h^{2v-1}} \hat{\sigma}_{p,v}(x)^2 + h^{2p+2-2v} \hat{\mathcal{B}}_{p,v}(x)^2 \right].$$

We make some remarks here.

Remark 3 (Preliminary bandwidth ℓ). The optimization argument h enters the RHS of the previous display in three places. First it is part of the variance component, by $1/h^{2v-1}$. Second it shows as a multiplicative factor of the bias component, $h^{2p-2v+2}$. Finally within the definition of $\hat{\mathcal{B}}_{p,v}(x)$, there is another multiplicative h , in front of the higher order bias.

The preliminary bandwidth ℓ , serves a different role. It is used to estimate the variance and bias components. Of course one can use different preliminary bandwidths for $\hat{\sigma}_{p,v}(x)$, $\widehat{\mathbf{S}}_{p,x}^{-1} \mathbf{c}_{p,x}$ and $\widehat{\mathbf{S}}_{p,x}^{-1} \tilde{\mathbf{c}}_{p,x}$. ||

Remark 4 (Consistent bias estimator). The bias estimator we proposed, $h^{p-v+1} \hat{\mathcal{B}}_{p,v}(x)$, is consistent in rate for the true leading bias, but not necessarily in constant. Compare $\hat{\mathcal{B}}_{p,v}(x)$ and $\mathcal{B}_{p,v}(x)$, it is easily seen that the term involving $F^{(p+1)}(x)F^{(2)}(x)/f(x)$ is not captured. To capture this term, we need one additional nonparametric estimator for $F^{(2)}(x)$. This is indeed feasible, and one can employ our local polynomial regression-based estimator for this purpose. ||

Theorem 3 (Consistent bandwidth). *Let $1 \leq v \leq p$. Assume the preliminary bandwidth ℓ is chosen such that $nh^{2v-1} \hat{\sigma}_{p,v}(x)^2 / \mathcal{V}_{p,v}(x) \rightarrow_{\mathbb{P}} 1$, $\widehat{\mathbf{S}}_{p,x}^{-1} \mathbf{c}_{p,x} \rightarrow_{\mathbb{P}} \mathbf{S}_{p,x}^{-1} \mathbf{c}_{p,x}$, and $\widehat{\mathbf{S}}_{p,x}^{-1} \tilde{\mathbf{c}}_{p,x} \rightarrow_{\mathbb{P}} \mathbf{S}_{p,x}^{-1} \tilde{\mathbf{c}}_{p,x}$. Under the conditions of Lemma 1 and Theorem 2:*

- *If either x is in boundary regions or $p - v$ is odd, let $\hat{F}^{(p+1)}(x)$ be consistent for $F^{(p+1)} \neq 0$. Then*

$$\frac{\hat{h}_{p,v}(x)}{h_{p,v}(x)} \rightarrow_{\mathbb{P}} 1.$$

- *If x is in interior and $p - v$ is even, let $\hat{F}^{(p+2)}(x)$ be consistent for $F^{(p+2)} \neq 0$. Further assume $nh^3 \rightarrow 0$ and $h_{p,v}(x)$ is well-defined. Then*

$$\frac{\hat{h}_{p,v}(x)}{h_{p,v}(x)} \rightarrow_{\mathbb{P}} C \in (0, \infty).$$

3.2 For CDF Estimate ($v = 0$)

In this subsection we mention briefly how to choose bandwidth for the CDF estimate, $\hat{F}_p^{(0)}(x) \equiv \hat{F}_p(x)$. We assume x is in interior. Previous discussions on bias remains to apply:

$$h^{p+1}\mathcal{B}_{p,0}(x) = h^{p+1} \left\{ \frac{F^{(p+1)}(x)}{(p+1)!} \mathbf{e}'_0 \mathbf{S}_{p,x}^{-1} \mathbf{c}_{p,x} + h \left(\frac{F^{(p+2)}(x)}{(p+2)!} + \frac{F^{(p+1)}(x)}{(p+1)!} \frac{F^{(2)}(x)}{f(x)} \right) \mathbf{e}'_0 \mathbf{S}_{p,x}^{-1} \tilde{\mathbf{c}}_{p,x} \right\},$$

which means the bias of $\hat{F}_p(x)$ has order h^{p+1} if either x is boundary or p is odd, and h^{p+2} otherwise. Difficulty arises since the CDF estimator has leading variance of order

$$\mathcal{V}_{p,0}(x) \asymp \frac{\mathbb{1}[x \text{ interior}] + h}{n},$$

which cannot be used for bandwidth selection, because the above is proportional to the bandwidth (i.e., there is no bias-variance trade-off).

The trick is to use a higher order variance term. Recall that the local polynomial regression-based estimator is essentially a second order U-statistic, which is then decomposed into two terms, a linear term $\hat{\mathbf{L}}$ and a quadratic term $\hat{\mathbf{R}}$, where the latter is a degenerate second-order U-statistic. The variance of the quadratic term $\hat{\mathbf{R}}$ has been ignored so far, as it is negligible compared to the variance of the linear term. For the CDF estimator, however, it is the variance of this quadratic term that leads to a bias-variance trade-off. The exact form of this variance is given in Lemma 4. With this additional variance term included, we have (with some abuse of notation)

$$\mathcal{V}_{p,0}(x) \asymp \frac{\mathbb{1}[x \text{ interior}] + h}{n} + \frac{\mathbb{1}[x \text{ interior}] + h}{n^2 h}.$$

Provided x is an interior point, the additional variance term increases as the bandwidth shrinks. As a result, a MSE-optimal bandwidth for $\hat{F}_p(x)$ is well-defined, and estimating this bandwidth is also straightforward.

Order of MSE-optimal bandwidth: $h_{p,0}(x) \asymp$

	$p - v$ odd	even
x interior	$n^{-\frac{2}{2p+3}}$	$n^{-\frac{2}{2p+5}}$
boundary	undefined	undefined

What if x is in a boundary region? Then the MSE-optimal bandwidth for $\hat{F}_p(x)$ is not well defined. The leading variance now takes the form $h/n + 1/n^2$, which is proportional to the bandwidth. (This is not surprising, since for boundary x the CDF is known, and a very small bandwidth gives a super-consistent estimator.). Although MSE-optimal bandwidth for $\hat{F}_p(x)$ is not well-defined for boundary x , it is still feasible to minimize the empirical MSE. To see how this works, one first estimate the bias term and variance term with some preliminary bandwidth ℓ , leading to $\hat{\mathcal{B}}_{p,0}(x)$ and $\hat{\mathcal{V}}_{p,0}(x)$. Then the MSE-optimal bandwidth can be constructed by minimizing the empirical MSE. Under regularity conditions, $\hat{\mathcal{B}}_{p,0}(x)$ will converge to some nonzero constant, while, if x is boundary, $\hat{\mathcal{V}}_{p,0}(x)$ has order ℓ , the same as the preliminary bandwidth. Then the MSE-optimal bandwidth constructed in this way will have the following order:

Order of estimated MSE-optimal bandwidth: $\hat{h}_{p,0}(x) \asymp$

	$p - v$ odd	even
x interior	$n^{-\frac{2}{2p+3}}$	$n^{-\frac{2}{2p+5}}$
boundary	$(n^2/\ell)^{-\frac{1}{2p+3}}$	$(n^2/\ell)^{-\frac{1}{2p+5}}$

Note that the preliminary bandwidth enters the rate of $\hat{h}_{p,0}(x)$ for boundary x , because it determines the rate at which the variance estimator $\hat{\mathcal{V}}_{p,0}(x)$ vanishes. Although this estimated bandwidth is not consistent for any well-defined object, it can be useful in practice, and it reflects the fact that for boundary x it is appropriate to use bandwidth shrinks fast when the object of interest is the CDF

4 Application to Manipulation Testing

We devote this section to density discontinuity (manipulation) tests in regression discontinuity designs. Assume there is a natural (and known) partition of the support $\mathcal{X} = [x_L, x_U] = [x_L, \bar{x}] \cup [\bar{x}, x_U] = \mathcal{X}_- \cup \mathcal{X}_+$, and the regularity conditions we imposed so far are satisfied on each of the partitions, \mathcal{X}_- and \mathcal{X}_+ . To be precise, assume the distribution F is continuously differentiable to a certain order on each of the partitions, but the derivatives are not necessarily continuous across the cutoff \bar{x} . In this case consistent estimates of the densities (and derivatives thereof) require fitting local polynomials separately on each sides of \bar{x} . Alternatively, one can use the joint estimation framework introduced below.

4.1 Unrestricted Model

By an unrestricted model with cutoff \bar{x} , we consider the following polynomial basis \mathbf{r}_p

$$\mathbf{r}_p(u) = \left[\mathbf{1}_{\{u < 0\}} \quad u \mathbf{1}_{\{u < 0\}} \quad \cdots \quad u^p \mathbf{1}_{\{u < 0\}} \quad \Big| \quad \mathbf{1}_{\{u \geq 0\}} \quad u \mathbf{1}_{\{u \geq 0\}} \quad \cdots \quad u^p \mathbf{1}_{\{u \geq 0\}} \right]' \in \mathbb{R}^{2p+2}.$$

The following two vectors will arise later, which we give the definition here:

$$\mathbf{r}_{-,p}(u) = \left[1 \quad u \quad \cdots \quad u^p \quad 0 \quad \cdots \quad 0 \right]', \quad \mathbf{r}_{+,p}(u) = \left[0 \quad 0 \quad \cdots \quad 0 \quad 1 \quad \cdots \quad u^p \right]'$$

Also we define the vectors to extract the corresponding derivatives

$$\mathbf{I}_{2p+2} = \left[\mathbf{e}_{0,-} \quad \mathbf{e}_{1,-} \quad \cdots \quad \mathbf{e}_{p,-} \quad \mathbf{e}_{0,+} \quad \mathbf{e}_{1,+} \quad \cdots \quad \mathbf{e}_{p,+} \right].$$

With the above definition, the estimator at the cutoff is¹

$$\hat{\boldsymbol{\beta}}_p(\bar{x}) = \arg \min_{\mathbf{b} \in \mathbb{R}^{2p+2}} \sum_i \left(\tilde{F}(x_i) - \mathbf{r}_p(x_i - \bar{x})' \mathbf{b} \right)^2 K_h(x_i - \bar{x}).$$

We assume the same bandwidth is used below and above the cutoff to avoid cumbersome notation. Generalizing to using different bandwidths is straightforward. Other notations (for example \mathbf{X} and

¹The EDF is defined with the whole sample as before: $\tilde{F}(u) = n^{-1} \sum_i \mathbf{1}[x_i \leq u]$.

\mathbf{X}_h) are redefined similarly, with the scaling matrix \mathbf{H} adjusted so that $\mathbf{H}^{-1}\mathbf{r}_p(u) = \mathbf{r}_p(h^{-1}u)$ is always true. we denote the estimates by

$$\hat{F}_p^{(v)}(\bar{x}-) = v! \mathbf{e}'_{v,-} \hat{\beta}_p(\bar{x}), \quad \hat{F}_p^{(v)}(\bar{x}+) = v! \mathbf{e}'_{v,+} \hat{\beta}_p(\bar{x}).$$

Remark 5 (Separate estimation). An alternative implementation is to apply our local polynomial-based estimator separately to the two samples, one with observations below the cutoff, and the other with observations above the cutoff. To be precise, let $\tilde{F}_-(\cdot)$ and $\tilde{F}_+(\cdot)$ be the empirical distribution functions constructed by the two samples. That is,

$$\tilde{F}_-(x) = \frac{1}{n_-} \sum_{i: x_i < \bar{x}} \mathbf{1}[x_i \leq x], \quad \tilde{F}_+(x) = \frac{1}{n_+} \sum_{i: x_i \geq \bar{x}} \mathbf{1}[x_i \leq x],$$

where n_- and n_+ denote the size of the two samples, respectively. The the local polynomial approach, applied to $\tilde{F}_-(\cdot)$ and $\tilde{F}_+(\cdot)$ separately, will yield two sets of estimates, which we denote by $\hat{F}_{p,-}^{(v)}(\bar{x})$ and $\hat{F}_{p,+}^{(v)}(\bar{x})$. To see the relation between joint and separate estimations, we note the following (which can be easily seen using least squares algebra)

$$\begin{aligned} v = 0 & \quad \hat{F}_{p,-}(\bar{x}) = \frac{n}{n_-} \hat{F}_p(\bar{x}-), \quad \hat{F}_{p,+}(\bar{x}) = \frac{n}{n_+} \hat{F}_p(\bar{x}+) - \frac{n_-}{n_+} \\ v \geq 1 & \quad \hat{F}_{p,-}^{(v)}(\bar{x}) = \frac{n}{n_-} \hat{F}_p^{(v)}(\bar{x}-), \quad \hat{F}_{p,+}^{(v)}(\bar{x}) = \frac{n}{n_+} \hat{F}_p^{(v)}(\bar{x}+). \end{aligned}$$

The difference comes from the fact that by separate estimation, one obtains estimates of the conditional CDF and the derivatives. \parallel

In the following lemmas, we will give asymptotic results for the joint estimation problem. Proofs are omitted.

Lemma 7. *Let Assumptions of Lemma 1 hold separately on \mathcal{X}_- and \mathcal{X}_+ , then*

$$\frac{1}{n} \mathbf{X}'_h \mathbf{K}_h \mathbf{X}_h = f(\bar{x}-) \mathbf{S}_{-,p} + f(\bar{x}+) \mathbf{S}_{+,p} + O(h) + O_{\mathbb{P}}(1/\sqrt{nh}),$$

where

$$\mathbf{S}_{-,p} = \int_{-1}^0 \mathbf{r}_{-,p}(u) \mathbf{r}_{-,p}(u)' K(u) du, \quad \mathbf{S}_{+,p} = \int_0^1 \mathbf{r}_{+,p}(u) \mathbf{r}_{+,p}(u)' K(u) du.$$

Again we decompose the estimator into four terms, namely $\hat{\mathbf{B}}_{\text{LI}}$, $\hat{\mathbf{B}}_{\text{S}}$, $\hat{\mathbf{L}}$ and $\hat{\mathbf{R}}$.

Lemma 8. *Let Assumptions of Lemma 2 hold separately on \mathcal{X}_- and \mathcal{X}_+ , then*

$$\hat{\mathbf{B}}_{\text{S}} = h^{p+1} \left\{ \frac{F^{(p+1)}(\bar{x}-) f(\bar{x}-)}{(p+1)!} \mathbf{c}_{-,p} + \frac{F^{(p+1)}(\bar{x}+) f(\bar{x}+)}{(p+1)!} \mathbf{c}_{+,p} \right\} + o_{\mathbb{P}}(h^{p+1}), \quad \hat{\mathbf{B}}_{\text{LI}} = O_{\mathbb{P}}\left(\frac{1}{n}\right),$$

where

$$\mathbf{c}_{-,p} = \int_{-1}^0 u^{p+1} \mathbf{r}_{-,p}(u) K(u) du, \quad \mathbf{c}_{+,p} = \int_0^1 u^{p+1} \mathbf{r}_{+,p}(u) K(u) du.$$

Lemma 9. *Let Assumptions of Lemma 3 hold separately on \mathcal{X}_- and \mathcal{X}_+ , then*

$$\begin{aligned} \mathbb{V} \left[\sqrt{\frac{n}{h}} (\mathbf{e}_{1,+} - \mathbf{e}_{1,-})' \left(f(\bar{x}+) \mathbf{S}_{+,p} + f(\bar{x}-) \mathbf{S}_{-,p} \right)^{-1} \hat{\mathbf{L}} \right] &= f(\bar{x}-) \mathbf{e}'_{1,-} \mathbf{S}_{-,p}^{-1} \mathbf{\Gamma}_{-,p} \mathbf{S}_{-,p}^{-1} \mathbf{e}_{1,-} \\ &\quad + f(\bar{x}+) \mathbf{e}'_{1,+} \mathbf{S}_{+,p}^{-1} \mathbf{\Gamma}_{+,p} \mathbf{S}_{+,p}^{-1} \mathbf{e}_{1,+} + O(h), \end{aligned}$$

where

$$\mathbf{\Gamma}_{-,p} = \iint_{[-1,0]^2} (u \wedge v) \mathbf{r}_{-,p}(u) \mathbf{r}_{-,p}(v)' K(u) K(v) \, du dv, \quad \mathbf{\Gamma}_{+,p} = \iint_{[0,1]^2} (u \wedge v) \mathbf{r}_{+,p}(u) \mathbf{r}_{+,p}(v)' K(u) K(v) \, du dv.$$

Note that the above gives the asymptotic variance of the difference $\hat{f}(\bar{x}+) - \hat{f}(\bar{x}-)$, and the variance takes an additive form. This is not surprising, since the two density estimates, $\hat{f}(\bar{x}+)$ and $\hat{f}(\bar{x}-)$, rely on distinctive subsamples, meaning that they are asymptotically independent.

Finally the order of $\hat{\mathbf{R}}$ can also be established.

Lemma 10. *Let Assumptions of Lemma 4 hold separately on \mathcal{X}_- and \mathcal{X}_+ , then*

$$\hat{\mathbf{R}} = O_{\mathbb{P}} \left(\sqrt{\frac{1}{n^2 h}} \right).$$

Now we state the main result concerning the manipulation testing. Let $\hat{\mathbf{S}}_{p,\bar{x}}$ and $\hat{\mathbf{\Gamma}}_{p,\bar{x}}$ be constructed as in Section 2.3, and

$$\hat{\mathcal{V}}_{p,1}(\bar{x}) = \frac{1}{h} (\mathbf{e}_{1,+} - \mathbf{e}_{1,-})' \hat{\mathbf{S}}_{p,\bar{x}} \hat{\mathbf{\Gamma}}_{p,\bar{x}} \hat{\mathbf{S}}_{p,\bar{x}} (\mathbf{e}_{1,+} - \mathbf{e}_{1,-}).$$

Corollary 1. *Assume Assumptions 1 and 2 hold separately on \mathcal{X}_- and \mathcal{X}_+ with $\alpha_x \geq p+1$ for some integer $p \geq 1$. Further, $n \cdot h^2 \rightarrow \infty$ and $n \cdot h^{2p+1} \rightarrow 0$. Then under the null hypothesis $\mathbf{H}_0 : f(\bar{x}+) = f(\bar{x}-)$,*

$$T_p(h) = \frac{\hat{f}_p(\bar{x}+) - \hat{f}_p(\bar{x}-)}{\sqrt{\frac{1}{nh} \hat{\mathcal{V}}_{p,1}(\bar{x})}} \rightsquigarrow \mathcal{N}(0, 1).$$

As a result, under the alternative hypothesis $\mathbf{H}_1 : f(\bar{x}+) \neq f(\bar{x}-)$,

$$\lim_{n \rightarrow \infty} \mathbb{P}[|T_p(h)| \geq \Phi_{1-\alpha/2}] = 1.$$

Here $\Phi_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -quantile of the standard normal distribution.

Remark 6 (Separate estimation). Recall that it is possible to implement our local polynomial estimator separately for the two subsamples, below and above the cutoff \bar{x} . Let $\hat{f}_{p,-}(\bar{x})$ and $\hat{f}_{p,+}(\bar{x})$ be the two density estimates, and $\hat{\mathcal{V}}_{p,1,-}(\bar{x})$ and $\hat{\mathcal{V}}_{p,1,+}(\bar{x})$ be the associated variance estimates. Then the test statistic is equivalently:

$$T_p(h) = \frac{\frac{n_+}{n} \hat{f}_{p,+}(\bar{x}) - \frac{n_-}{n} \hat{f}_{p,-}(\bar{x})}{\sqrt{\frac{1}{nh} \left(\frac{n_+}{n} \hat{\mathcal{V}}_{p,1,+}(\bar{x}) + \frac{n_-}{n} \hat{\mathcal{V}}_{p,1,-}(\bar{x}) \right)}}.$$

||

4.2 Restricted Model

In the previous subsection, we gave a test procedure on the discontinuity of the density by estimating on the two sides of the cutoff separately. This procedure is flexible and requires minimum assumptions. There are ways, however, to improve the power of the test when the densities are estimated with additional assumptions on the smoothness of the CDF

In a restricted model, the polynomial basis is re-defined as

$$\mathbf{r}_p(u) = \begin{bmatrix} 1 & \mathbf{1}(u < 0) & \mathbf{1}(u \geq 0) & u^2 & u^3 & \dots & u^p \end{bmatrix}' \in \mathbb{R}^{p+2},$$

and the estimator in the fully restricted model is

$$\hat{\boldsymbol{\beta}}_p(\bar{x}) = \left[\hat{F}_p(\bar{x}) \quad \hat{f}_p(\bar{x}-) \quad \hat{f}_p(\bar{x}+) \quad \frac{1}{2}\hat{F}_p^{(2)}(\bar{x}) \quad \dots \quad \frac{1}{p!}\hat{F}_p^{(p)}(\bar{x}) \right]' = \arg \max_{\mathbf{b} \in \mathbb{R}^{p+2}} \sum_i \left(\tilde{F}(x_i) - \mathbf{r}_p(x_i - \bar{x})' \mathbf{b} \right)^2 K_h(x_i - \bar{x}).$$

Again the notations (for example \mathbf{X} and \mathbf{X}_h) are redefined similarly, with the scaling matrix \mathbf{H} adjusted to ensure $\mathbf{H}^{-1} \mathbf{r}_p(u) = \mathbf{r}_p(h^{-1}u)$. Here $\hat{F}_p(\bar{x})$ is the estimated CDF and $\frac{1}{2}\hat{F}_p^{(2)}(\bar{x}), \dots, \frac{1}{p!}\hat{F}_p^{(p)}(\bar{x})$ are the estimated higher order derivatives, which we assume are all continuous at \bar{x} , while $\hat{f}_p(\bar{x}-)$ and $\hat{f}_p(\bar{x}+)$ are the estimated densities on the two sides of \bar{x} . Therefore we call the above model restricted, since it only allows discontinuity of the first derivative of F (i.e. the density) but not the other derivatives.

With the modification of the polynomial basis, all other matrices in the previous subsection are redefined similarly, and

$$\mathbf{I}_{p+2} = \begin{bmatrix} \mathbf{e}_0 & \mathbf{e}_{1,-} & \mathbf{e}_{1,+} & \mathbf{e}_2 & \dots & \mathbf{e}_p \end{bmatrix}_{(p+2) \times (p+2)}.$$

where the subscripts indicate the corresponding derivatives to extract. Moreover

$$\mathbf{r}_{-,p}(u) = \begin{bmatrix} 1 & u & 0 & u^2 & \dots & u^p \end{bmatrix}, \quad \mathbf{r}_{+,p}(u) = \begin{bmatrix} 1 & 0 & u & u^2 & \dots & u^p \end{bmatrix}.$$

Lemma 11. *Let Assumptions of Lemma 1 hold with the exception that f may be discontinuous across \bar{x} , then*

$$\frac{1}{n} \mathbf{X}'_h \mathbf{K}_h \mathbf{X}_h = \{f(\bar{x}-) \mathbf{S}_{-,p} + f(\bar{x}+) \mathbf{S}_{+,p}\} + O(h) + O_{\mathbb{P}}(1/\sqrt{nh}),$$

where

$$\mathbf{S}_{-,p} = \int_{-1}^0 \mathbf{r}_{-,p}(u) \mathbf{r}_{-,p}(u)' K(u) du, \quad \mathbf{S}_{+,p} = \int_0^1 \mathbf{r}_{+,p}(u) \mathbf{r}_{+,p}(u)' K(u) du.$$

Again we decompose the estimator into four terms, $\hat{\mathbf{B}}_{\text{LI}}$, $\hat{\mathbf{B}}_{\text{S}}$, $\hat{\mathbf{L}}$ and $\hat{\mathbf{R}}$, which correspond to leave-in bias, smoothing bias, linear variance and quadratic variance, respectively.

Lemma 12. *Let Assumptions of Lemma 2 hold with the exception that f may be discontinuous across \bar{x} , then*

$$\hat{\mathbf{B}}_{\text{S}} = h^{p+1} \left\{ \frac{F^{(p+1)}(\bar{x}-) f(\bar{x}-)}{(p+1)!} \mathbf{c}_{-,p} + \frac{F^{(p+1)}(\bar{x}+) f(\bar{x}+)}{(p+1)!} \mathbf{c}_{+,p} \right\} + o_{\mathbb{P}}(h^{p+1}), \quad \hat{\mathbf{B}}_{\text{LI}} = O_{\mathbb{P}}\left(\frac{1}{n}\right), \quad (1)$$

where

$$\mathbf{c}_{-,p} = \int_{-1}^0 u^{p+1} \mathbf{r}_{-,p}(u) K(u) du, \quad \mathbf{c}_{+,p} = \int_0^1 u^{p+1} \mathbf{r}_{+,p}(u) K(u) du.$$

Lemma 13. *Let Assumptions of Lemma 3 hold with the exception that f may be discontinuous across \bar{x} , then*

$$\begin{aligned} & \mathbb{V} \left[\sqrt{\frac{n}{h}} (\mathbf{e}_{1,+} - \mathbf{e}_{1,-})' \left(f(\bar{x}+) \mathbf{S}_{+,p} + f(\bar{x}-) \mathbf{S}_{-,p} \right)^{-1} \hat{\mathbf{L}} \right] \\ &= (\mathbf{e}_{1,+} - \mathbf{e}_{1,-})' \left(f(\bar{x}+) \mathbf{S}_{+,p} + f(\bar{x}-) \mathbf{S}_{-,p} \right)^{-1} \left(f(\bar{x}+)^3 \mathbf{\Gamma}_{+,p} \right. \\ &\quad \left. + f(\bar{x}-)^3 \mathbf{\Psi} \mathbf{\Gamma}_{+,p} \mathbf{\Psi} \right) \left(f(\bar{x}+) \mathbf{S}_{+,p} + f(\bar{x}-) \mathbf{S}_{-,p} \right)^{-1} (\mathbf{e}_{1,+} - \mathbf{e}_{1,-}) + O(h), \end{aligned}$$

where

$$\mathbf{\Gamma}_{-,p} = \iint_{[-1,0]^2} (u \wedge v) \mathbf{r}_{-,p}(u) \mathbf{r}_{-,p}(v)' K(u) K(v) dudv, \quad \mathbf{\Gamma}_{+,p} = \iint_{[0,1]^2} (u \wedge v) \mathbf{r}_{+,p}(u) \mathbf{r}_{+,p}(v)' K(u) K(v) dudv.$$

and

$$\mathbf{\Psi} = \begin{bmatrix} (-1)^0 & & & & & \\ & (-1)^1 & & & & \\ & (-1)^1 & & & & \\ & & (-1)^2 & & & \\ & & & (-1)^3 & & \\ & & & & \ddots & \\ & & & & & (-1)^p \end{bmatrix}.$$

Again we can show that the quadratic part is negligible.

Lemma 14. *Let Assumptions of Lemma 4 hold with the exception that f may not be continuous across \bar{x} , then*

$$\hat{\mathbf{R}} = O_{\mathbb{P}} \left(\sqrt{\frac{1}{n^2 h}} \right).$$

Now we state the main result concerning the manipulation testing. Let $\hat{\mathbf{S}}_{p,\bar{x}}$ and $\hat{\mathbf{\Gamma}}_{p,\bar{x}}$ be constructed as in Section 2.3, and

$$\hat{\mathcal{V}}_{p,1}(\bar{x}) = \frac{1}{h} (\mathbf{e}_{1,+} - \mathbf{e}_{1,-})' \hat{\mathbf{S}}_{p,\bar{x}} \hat{\mathbf{\Gamma}}_{p,\bar{x}} \hat{\mathbf{S}}_{p,\bar{x}} (\mathbf{e}_{1,+} - \mathbf{e}_{1,-}).$$

Corollary 2. *Assume Assumptions 1 and 2 hold separately on \mathcal{X}_- and \mathcal{X}_+ with $\alpha_x \geq p + 1$ for some integer $p \geq 1$. Further, $n \cdot h^2 \rightarrow \infty$ and $n \cdot h^{2p+1} \rightarrow 0$. Then under the null hypothesis $\mathbf{H}_0 : f(\bar{x}+) = f(\bar{x}-)$,*

$$T_p(h) = \frac{\hat{f}_p(\bar{x}+) - \hat{f}_p(\bar{x}-)}{\sqrt{\frac{1}{nh} \hat{\mathcal{V}}_{p,1}(\bar{x})}} \rightsquigarrow \mathcal{N}(0, 1).$$

As a result, under the alternative hypothesis $\mathbf{H}_1 : f(\bar{x}+) \neq f(\bar{x}-)$,

$$\lim_{n \rightarrow \infty} \mathbb{P}[|T_p(h)| \geq \Phi_{1-\alpha/2}] = 1.$$

Here $\Phi_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -quantile of the standard normal distribution.

5 Other Standard Error Estimators

The standard error $\hat{\sigma}_{p,v}(x)$ (see Theorem 2) is fully automatic and adapts to both interior and boundary regions. In this section we consider two other ways to construct a standard error.

5.1 Plug-in Standard Error

Take $v \geq 1$. Then the asymptotic variance of $\hat{F}_p^{(v)}(x)$ takes the following form:

$$\mathcal{V}_{p,v}(x) = (v!)^2 f(x) \mathbf{e}'_v \mathbf{S}_{p,x}^{-1} \mathbf{\Gamma}_{p,x} \mathbf{S}_{p,x}^{-1} \mathbf{e}_v.$$

One way of constructing estimate of the above quantity is to plug-in a consistent estimator of $f(x)$, which is simply the estimated density. Hence we can use

$$\hat{\mathcal{V}}_{p,v}(x) = (v!)^2 \hat{f}_p(x) \mathbf{e}'_v \mathbf{S}_{p,x}^{-1} \mathbf{\Gamma}_{p,x} \mathbf{S}_{p,x}^{-1} \mathbf{e}_v.$$

The next question is how $\mathbf{S}_{p,x}$ and $\mathbf{\Gamma}_{p,x}$ should be constructed. Note that they are related to the kernel, evaluation point x and the bandwidth h , but *not* the data generating process. Therefore the three matrices can be constructed by either analytical integration or numerical method.

5.2 Jackknife-based Standard Error

The standard error $\hat{\sigma}_{p,v}(x)$ is obtained by inspecting the asymptotic linear representation. It is fully automatic and adapts to both interior and boundaries. In this part, we present another standard error which resembles $\hat{\sigma}_{p,v}(x)$, albeit with a different motivation.

Recall that $\hat{\beta}_p(x)$ is essentially a second order U-statistic, and the following expansion is justified:

$$\begin{aligned} & \frac{1}{n} \mathbf{X}'_h \mathbf{K}_h (\mathbf{Y} - \mathbf{X} \beta_p(x)) \\ &= \frac{1}{n} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\tilde{F}(x_i) - \mathbf{r}_p(x_i - x)' \beta_p(x) \right) K_h(x_i - x) \\ &= \frac{1}{n} \sum_i \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\frac{1}{n-1} \sum_{j:j \neq i} \left(\mathbf{1}(x_j \leq x_i) - \mathbf{r}_p(x_i - x)' \beta_p(x) \right) \right) K_h(x_i - x) + O_{\mathbb{P}} \left(\frac{1}{n} \right) \\ &= \frac{1}{n(n-1)} \sum_{i,j:i \neq j} \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\mathbf{1}(x_j \leq x_i) - \mathbf{r}_p(x_i - x)' \beta_p(x) \right) K_h(x_i - x) + O_{\mathbb{P}} \left(\frac{1}{n} \right), \end{aligned}$$

where the remainder represents leave-in bias. Note that the above could be written as a U-statistic, and to apply the Hoeffding decomposition, define

$$\begin{aligned} \mathbf{U}(x_i, x_j) &= \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\mathbf{1}(x_j \leq x_i) - \mathbf{r}_p(x_i - x)' \beta_p(x) \right) K_h(x_i - x) \\ &\quad + \mathbf{r}_p \left(\frac{x_j - x}{h} \right) \left(\mathbf{1}(x_i \leq x_j) - \mathbf{r}_p(x_j - x)' \beta_p(x) \right) K_h(x_j - x), \end{aligned}$$

which is symmetric in its two arguments. Then

$$\begin{aligned} \frac{1}{n} \mathbf{X}'_h \mathbf{K}_h (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}_p(x)) &= \mathbb{E} [\mathbf{U}(x_i, x_j)] + \frac{1}{n} \sum_i \left(\mathbf{U}_1(x_i) - \mathbb{E} [\mathbf{U}(x_i, x_j)] \right) \\ &+ \binom{n}{2}^{-1} \sum_{i,j;i < j} \left(\mathbf{U}(x_i, x_j) - \mathbf{U}_1(x_i) - \mathbf{U}_1(x_j) + \mathbb{E} [\mathbf{U}(x_i, x_j)] \right). \end{aligned}$$

Here $\mathbf{U}_1(x_i) = \mathbb{E} [\mathbf{U}(x_i, x_j) | x_i]$. The second line in the above display is the analogue of $\hat{\mathbf{L}}$, which contributes to the leading variance, and the third line is negligible. The new standard error, we call the jackknife-based standard error, is given by the following:

$$\hat{\sigma}_{p,v}^{(\text{JK})}(x) \equiv (v!) \sqrt{\frac{1}{nh^{2v}} \mathbf{e}'_v \hat{\mathbf{S}}_{p,x}^{-1} \hat{\mathbf{\Gamma}}_{p,x}^{\text{JK}} \hat{\mathbf{S}}_{p,x}^{-1} \mathbf{e}_v},$$

with

$$\hat{\mathbf{\Gamma}}_{p,x}^{\text{JK}} = \frac{1}{n} \sum_i \left(\frac{1}{n-1} \sum_{j:j \neq i} \hat{\mathbf{U}}(x_i, x_j) \right) \left(\frac{1}{n-1} \sum_{j:j \neq i} \hat{\mathbf{U}}(x_i, x_j) \right)' - \left(\binom{n}{2}^{-1} \sum_{i,j;i \neq j} \hat{\mathbf{U}}(x_i, x_j) \right) \left(\binom{n}{2}^{-1} \sum_{i,j;i \neq j} \hat{\mathbf{U}}(x_i, x_j) \right)',$$

and

$$\begin{aligned} \hat{\mathbf{U}}(x_i, x_j) &= \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\mathbf{1}(x_j \leq x_i) - \mathbf{r}_p(x_i - x)' \hat{\boldsymbol{\beta}}_p(x) \right) K_h(x_i - x) \\ &+ \mathbf{r}_p \left(\frac{x_j - x}{h} \right) \left(\mathbf{1}(x_i \leq x_j) - \mathbf{r}_p(x_j - x)' \hat{\boldsymbol{\beta}}_p(x) \right) K_h(x_j - x). \end{aligned}$$

The name jackknife comes from the fact that we use leave-one-out “estimator” for $\mathbf{U}_1(x_i)$: with x_i fixed,

$$\left\langle \frac{1}{n-1} \sum_{j:j \neq i} \hat{\mathbf{U}}(x_i, x_j) \right\rangle \rightarrow_{\mathbb{P}} \mathbf{U}_1(x_i).$$

Under the same conditions specified in Theorem 2, one can show that the jackknife-based standard error is consistent.

6 Simulation Study

6.1 DGP 1: Truncated Normal Distribution

In this subsection, we conduct simulation study based on truncated normal distribution. To be more specific, the underlying distribution of x_i is the standard normal distribution truncated below at -0.8 . Therefore,

$$G(x) = F(x) = \frac{\Phi(x) - \Phi(-0.8)}{1 - \Phi(-0.8)}, \quad x \geq -0.8,$$

and zero otherwise. Equivalently, x_i has Lebesgue density $\Phi^{(1)}(x)/(1 - \Phi(-0.8))$ on $[-0.8, \infty]$.

In this simulation study, the target parameter is the density function evaluated at various points. Note that both the variance and the bias of our estimator depend on the evaluation point, and in particular, the magnitude of the bias depends on higher order derivatives of the distribution

function.

1. *Evaluation point.* We estimate the density at $x \in \{-0.8, -0.5, 0.5, 1.5\}$. Note that -0.8 is the boundary point, where classical density estimators such as the kernel density estimator has high bias. The point -0.5 , given our bandwidth choice, is fairly close to the boundary, hence should be understood as in the lower boundary region. The two points 0.5 and 1.5 are interior, but the curvature of the normal density is quite different at those two points, and we expect to see the estimators having different bias behaviors.
2. *Polynomial order.* We consider $p \in \{2, 3\}$. For density estimation using our estimators, $p = 2$ should be the default choice, since it corresponds to estimating conditional mean with local linear regression. Such choice is also recommended by Fan and Gijbels (1996), according to which one should always choose $p - s = 2 - 1 = 1$ to be an odd number. We include $p = 3$ for completeness.
3. *Kernel function.* For local polynomial regression, the choice of kernel function is usually not very important. We use the triangular kernel $k(u) = (1 - |u|) \vee 0$.
4. *Sample size.* The sample size used consists of $n \in \{1000, 2000\}$. For most empirical studies employing nonparametric density estimation, the sample size is well above 1000, hence $n = 2000$ is more representative.

Overall, we have $4 \times 2 \times 2 = 16$ designs, and for each design, we conduct 5000 Monte Carlo repetitions.

We consider a grid of bandwidth choices, which correspond to multiples of the MSE-optimal bandwidth, ranging from $0.1h_{\text{MSE}}$ to $2h_{\text{MSE}}$. We also consider the estimated bandwidth. The MSE-optimal bandwidth, h_{MSE} , is chosen by minimizing the asymptotic mean squared error, using the true underlying distribution.

For each design, we report the empirical bias of the estimator, $\mathbb{E}[\hat{f}_p(x) - f(x)]$, under bias. And empirical standard deviations, $\mathbb{V}^{1/2}[\hat{f}_p(x)]$, and empirical root-MSE, under sd and $\sqrt{\text{mse}}$, respectively. For the standard errors constructed from the variance estimators, we report their empirical average under mean, which should be compared to sd. We also report the empirical rejection rate of t-statistics at 5% nominal level, under size. The t-statistic is $(\hat{f}_p(x) - \mathbb{E}\hat{f}_p(x))/\text{se}$, which is exactly centered, hence rejection rate thereof is a measure of accuracy of normal approximation.

6.2 DGP 2: Exponential Distribution

In this subsection, we conduct simulation study based on exponential distribution. To be more specific, the underlying distribution of x_i is $F(x) = 1 - e^{-x}$. Equivalently, x_i has Lebesgue density e^{-x} for $x \geq 0$.

In this simulation study, the target parameter is the density function evaluated at various points. Note that both the variance and the bias of our estimator depend on the evaluation point, and

in particular, the magnitude of the bias depends on higher order derivatives of the distribution function.

1. *Evaluation point.* We estimate the density at $x \in \{0, 1, 1.5\}$. Note that 0 is the boundary point, where classical density estimators such as the kernel density estimator has high bias. The two points 1 and 1.5 are interior.
2. *Polynomial order.* We consider $p \in \{2, 3\}$. For density estimation using our estimators, $p = 2$ should be the default choice, since it corresponds to estimating conditional mean with local linear regression. Such choice is also recommended by Fan and Gijbels (1996), according to which one should always choose $p - s = 2 - 1 = 1$ to be an odd number. We include $p = 3$ for completeness.
3. *Kernel function.* For local polynomial regression, the choice of kernel function is usually not very important. We use the triangular kernel $k(u) = (1 - |u|) \vee 0$.
4. *Sample size.* The sample size used consists of $n \in \{1000, 2000\}$. For most empirical studies employing nonparametric density estimation, the sample size is well above 1000, hence $n = 2000$ is more representative.

Overall, we have $3 \times 2 \times 2 = 12$ designs, and for each design, we conduct 5000 Monte Carlo repetitions.

We consider a grid of bandwidth choices, which correspond to multiples of the MSE-optimal bandwidth, ranging from $0.1h_{\text{MSE}}$ to $2h_{\text{MSE}}$. We also consider the estimated bandwidth. The MSE-optimal bandwidth, h_{MSE} , is chosen by minimizing the asymptotic mean squared error, using the true underlying distribution.

For each design, we report the empirical bias of the estimator, $\mathbb{E}[\hat{f}_p(x) - f(x)]$, under bias. And empirical standard deviations, $\mathbb{V}^{1/2}[\hat{f}_p(x)]$, and empirical root-MSE, under sd and $\sqrt{\text{mse}}$, respectively. For the standard errors constructed from the variance estimators, we report their empirical average under mean, which should be compared to sd. We also report the empirical rejection rate of t-statistics at 5% nominal level, under size. The t-statistic is $(\hat{f}_p(x) - \mathbb{E}\hat{f}_p(x))/\text{se}$, which is exactly centered, hence rejection rate thereof is a measure of accuracy of normal approximation.

References

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, New York: Chapman & Hall/CRC.

7 Proof

7.1 Proof of Lemma 1

Proof. A generic element of the matrix $\frac{1}{n}\mathbf{X}'_h\mathbf{K}_h\mathbf{X}_h$ takes the form:

$$\frac{1}{n}\sum_i\frac{1}{h}\left(\frac{x_i-x}{h}\right)^sK\left(\frac{x_i-x}{h}\right), \quad 0 \leq s \leq 2p.$$

Then we compute the expectation:

$$\begin{aligned} \mathbb{E}\left[\frac{1}{n}\sum_i\frac{1}{h}\left(\frac{x_i-x}{h}\right)^sK\left(\frac{x_i-x}{h}\right)\right] &= \mathbb{E}\left[\frac{1}{h}\left(\frac{x_i-x}{h}\right)^sK\left(\frac{x_i-x}{h}\right)\right] \\ &= \int_{x_L}^{x_U}\frac{1}{h}\left(\frac{u-x}{h}\right)^sK\left(\frac{u-x}{h}\right)f(u)du = \int_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}}v^sK(v)f(x+vh)dv = \int_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}}v^sK(v)f(x+vh)dv, \end{aligned}$$

hence for x in the interior,

$$\mathbb{E}\left[\frac{1}{n}\sum_i\frac{1}{h}\left(\frac{x_i-x}{h}\right)^sK\left(\frac{x_i-x}{h}\right)\right] = f(x)\int_{\mathbb{R}}\mathbf{r}_p(v)\mathbf{r}_p(v)'K(v)dv + o(1),$$

and for $x = x_L + ch$ with $c \in [0, 1]$,

$$\mathbb{E}\left[\frac{1}{n}\sum_i\frac{1}{h}\left(\frac{x_i-x}{h}\right)^sK\left(\frac{x_i-x}{h}\right)\right] = f(x_L)\int_{-c}^{\infty}\mathbf{r}_p(v)\mathbf{r}_p(v)'K(v)dv + o(1),$$

and for $x = x_U - ch$ with $c \in [0, 1]$,

$$\mathbb{E}\left[\frac{1}{n}\sum_i\frac{1}{h}\left(\frac{x_i-x}{h}\right)^sK\left(\frac{x_i-x}{h}\right)\right] = f(x_U)\int_{-\infty}^c\mathbf{r}_p(v)\mathbf{r}_p(v)'K(v)dv + o(1),$$

provided that $F \in \mathcal{C}^1$.

The variance satisfies

$$\begin{aligned} \mathbb{V}\left[\frac{1}{n}\sum_i\frac{1}{h}\left(\frac{x_i-x}{h}\right)^sK\left(\frac{x_i-x}{h}\right)\right] &= \frac{1}{n}\mathbb{V}\left[\frac{1}{h}\left(\frac{x_i-x}{h}\right)^sK\left(\frac{x_i-x}{h}\right)\right] \\ &\leq \frac{1}{n}\mathbb{E}\left[\frac{1}{h^2}\left(\frac{x_i-x}{h}\right)^{2s}K\left(\frac{x_i-x}{h}\right)^2\right] = O\left(\frac{1}{nh}\right), \end{aligned}$$

provided that $F \in \mathcal{C}^1$. ■

7.2 Proof of Lemma 2

Proof. First consider the smoothing bias. The leading term can be easily obtain by taking expectation together with Taylor expansion of F to power $p+1$. The variance of this term has order $n^{-1}h^{-1}h^{2p+2}$, which gives the residual estimate $o_{\mathbb{P}}(h^{p+1})$ since it is assumed that $nh \rightarrow \infty$.

Next for the leave-in bias, note that it has expectation of order n^{-1} , and variance of order $n^{-3}h^{-1}$, hence overall this term of order $O_{\mathbb{P}}(n^{-1})$. ■

7.3 Proof of Lemma 3

Proof. We first compute the variance. Note that

$$\begin{aligned} & \int_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}} \mathbf{r}_p(u) \left(\tilde{F}(x+hu) - F(x+hu) \right) K(u) f(x+hu) du \\ &= \frac{1}{n} \int_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}} \mathbf{r}_p(u) \left(\mathbb{1}[x_i \leq x+hu] - F(x+hu) \right) K(u) f(x+hu) du, \end{aligned}$$

and

$$\begin{aligned} & \mathbb{V} \left[\int_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}} \mathbf{r}_p(u) \left(\mathbb{1}[x_i \leq x+hu] - F(x+hu) \right) K(u) f(x+hu) du \right] \\ &= \iint_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}} \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(x+hu) f(x+hv) \\ & \quad \times \left[\int_{\mathbb{R}} (\mathbb{1}[t \leq x+hu] - F(x+hu)) (\mathbb{1}[t \leq x+hv] - F(x+hv)) f(t) dt \right] dudv \\ &= \iint_{\frac{x_L-x}{h}}^{\frac{x_U-x}{h}} \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(x+hu) f(x+hv) \left(F(x+h(u \wedge v)) - F(x+hu)F(x+hv) \right) dudv. \quad (\text{I}) \end{aligned}$$

We first consider the interior case, where the above reduces to:

$$\begin{aligned} & (\text{I})_{\text{interior}} \\ &= \iint_{\mathbb{R}} \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(x)^2 \left(F(x) - F(x)^2 \right) dudv \\ & \quad + h \iint_{\mathbb{R}} (u \wedge v) \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(x)^3 dudv \\ & \quad - h \iint_{\mathbb{R}} (u+v) \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(x)^3 F(x) dudv \\ & \quad + h \iint_{\mathbb{R}} (u+v) \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(x) F^{(2)}(x) \left(F(x) - F(x)^2 \right) dudv + o(h) \\ &= f(x)^2 \left(F(x) - F(x)^2 \right) \mathbf{S}_{p,x} \mathbf{e}_0 \mathbf{e}_0' \mathbf{S}_{p,x} \\ & \quad - h f(x)^3 F(x) \mathbf{S}_{p,x} (\mathbf{e}_1 \mathbf{e}_0' + \mathbf{e}_0 \mathbf{e}_1') \mathbf{S}_{p,x} \\ & \quad + h f(x) F^{(2)}(x) \left(F(x) - F(x)^2 \right) \mathbf{S}_{p,x} (\mathbf{e}_1 \mathbf{e}_0' + \mathbf{e}_0 \mathbf{e}_1') \mathbf{S}_{p,x} \\ & \quad + h f(x)^3 \mathbf{\Gamma}_{p,x} + o(h). \end{aligned}$$

For $x = x_L + hc$ with $c \in [0, 1)$ in the lower boundary region,

$$\begin{aligned} & (\text{I})_{\text{lower boundary}} \\ &= h \iint_{\mathbb{R}} (u \wedge v + c) \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(x_L)^3 dudv + o(h) = h f(x_L)^3 \left(\mathbf{\Gamma}_{p,x} + c \mathbf{S}_{p,x} \mathbf{e}_0 \mathbf{e}_0' \mathbf{S}_{p,x} \right) + o(h). \end{aligned}$$

Finally, we have

$$\begin{aligned} & (\text{I})_{\text{upper boundary}} \\ &= h \iint_{\mathbb{R}} (u \wedge v - c) \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(x_U)^3 dudv - h \iint_{\mathbb{R}} (u+v-2c) \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(x_U)^3 dudv + o(h) \\ &= h f(x_U)^2 f(x_U) \left(\mathbf{\Gamma}_{p,x} + c \mathbf{S}_{p,x} \mathbf{e}_0 \mathbf{e}_0' \mathbf{S}_{p,x} - \mathbf{S}_{p,x} (\mathbf{e}_1 \mathbf{e}_0' + \mathbf{e}_0 \mathbf{e}_1') \mathbf{S}_{p,x} \right) + o(h). \end{aligned}$$

With the above results, it is easy to verify the variance formula, provided that we can show the asymptotic normality.

We first consider the interior case, and verify the Lindeberg condition on the fourth moment. Let $\boldsymbol{\alpha} \in \mathbb{R}^{p+1}$ be an

arbitrary nonzero vector, then

$$\begin{aligned}
& \sum_i \mathbb{E} \left(\frac{1}{\sqrt{n}} \boldsymbol{\alpha}' \mathbf{N}_x (f(x) \mathbf{S}_{p,x})^{-1} \int_{\frac{x_l-x}{h}}^{\frac{x_u-x}{h}} \mathbf{r}_p(u) \left(\mathbb{1}[x_i \leq x+hu] - F(x+hu) \right) K(u) f(x+hu) du \right)^4 \\
&= \frac{1}{n} \mathbb{E} \left(\boldsymbol{\alpha}' \mathbf{N}_x (f(x) \mathbf{S}_{p,x})^{-1} \int_{\frac{x_l-x}{h}}^{\frac{x_u-x}{h}} \mathbf{r}_p(u) \left(\mathbb{1}[x_i \leq x+hu] - F(x+hu) \right) K(u) f(x+hu) du \right)^4 \\
&= \frac{1}{n} \iiint \prod_{j=1,2,3,4} \left(\boldsymbol{\alpha}' \mathbf{N}_x (f(x) \mathbf{S}_{p,x})^{-1} \mathbf{r}_p(u_j) K(u_j) \right) f(x+hu_j) \\
&\quad \left[\int_{\mathbb{R}} \prod_{j=1,2,3,4} \left(\mathbb{1}[t \leq x+hu_j] - F(x+hu_j) \right) f(t) dt \right] du_1 du_2 du_3 du_4 \\
&\leq \frac{C}{n} \cdot \iiint \prod_{j=1,2,3,4} \left(\boldsymbol{\alpha}' \mathbf{N}_x (f(x) \mathbf{S}_{p,x})^{-1} \mathbf{r}_p(u_j) K(u_j) \right) f(x) du_1 du_2 du_3 du_4 + O\left(\frac{1}{nh}\right),
\end{aligned}$$

where $\mathcal{A} = [\frac{x_l-x}{h}, \frac{x_u-x}{h}]^4 \subset \mathbb{R}^4$. The first term in the above display is asymptotically negligible, since it takes the form $C \cdot (\boldsymbol{\alpha}' \mathbf{N}_x \mathbf{e}_0)^4 / n$ where the constant C depends on the DGP, and is finite. The order of the next term is $1/(nh)$, which comes from multiplying n^{-1} , h^{-2} (from the scaling matrix \mathbf{N}_x), and h (from linearization), hence is also negligible.

Under the assumption that $nh \rightarrow \infty$, the Lindeberg condition is verified for interior case. The same logic applies to the boundary case, whose proof is easier than the interior case, since the leading term in the calculation is identically zero for x in either the lower or upper boundary. \blacksquare

7.4 Proof of Lemma 4

Proof. For $\hat{\mathbf{R}}$, we rewrite it as a second order degenerate U-statistic:

$$\hat{\mathbf{R}} = \frac{1}{n^2} \sum_{i,j:i < j} \hat{\mathbf{U}}_{ij},$$

where

$$\begin{aligned}
\hat{\mathbf{U}}_{ij} &= \mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\mathbb{1}[x_j \leq x_i] - F(x_i) \right) K_h(x_i - x) + \mathbf{r}_p \left(\frac{x_j - x}{h} \right) \left(\mathbb{1}[x_i \leq x_j] - F(x_j) \right) K_h(x_j - x) \\
&\quad - \mathbb{E} \left[\mathbf{r}_p \left(\frac{x_i - x}{h} \right) \left(\mathbb{1}[x_j \leq x_i] - F(x_i) \right) K_h(x_i - x) \middle| x_j \right] - \mathbb{E} \left[\mathbf{r}_p \left(\frac{x_j - x}{h} \right) \left(\mathbb{1}[x_i \leq x_j] - F(x_j) \right) K_h(x_j - x) \middle| x_i \right].
\end{aligned}$$

To compute the leading term, it suffices to consider

$$\begin{aligned}
& 2\mathbb{E} \left[\mathbf{r}_p \left(\frac{x_i - x}{h} \right) \mathbf{r}_p \left(\frac{x_i - x}{h} \right)' \left(\mathbb{1}[x_j \leq x_i] - F(x_i) \right)^2 K_h(x_i - x)^2 \right] \\
&= 2\mathbb{E} \left[\mathbf{r}_p \left(\frac{x_i - x}{h} \right) \mathbf{r}_p \left(\frac{x_i - x}{h} \right)' \left(F(x_i) - F(x_i)^2 \right) K_h(x_i - x)^2 \right] \\
&= \frac{2}{h} \int_{\frac{x_l-x}{h}}^{\frac{x_u-x}{h}} \mathbf{r}_p(v) \mathbf{r}_p(v)' \left(F(x+hv) - F(x+hv)^2 \right) K(v)^2 f(x+hv) dv \\
&= \frac{2}{h} \int_{\frac{x_l-x}{h}}^{\frac{x_u-x}{h}} \mathbf{r}_p(v) \mathbf{r}_p(v)' \left(F(x) - F(x)^2 \right) K(v)^2 f(x) dv + O(1) \\
&=_{\text{interior}} \frac{2}{h} f(x) [F(x) - F(x)^2] \mathbf{T}_{p,x} + O(1), \\
&=_{\text{boundary}} O(1),
\end{aligned}$$

which closes the proof. \blacksquare

7.5 Proof of Theorem 1

Proof. This follows from previous lemmas. \blacksquare

7.6 Proof of Theorem 2

Proof. First we note that the second half of the theorem follows from the first half and the asymptotic normality result of Theorem 1, hence it suffices to prove the first half, i.e. the consistency of $\hat{Y}_{p,v}(x)$.

The analysis of this estimator is quite involved, since it takes the form of a third order V-statistic. Moreover, since the empirical d.f. \tilde{F} is involved in the formula, a full expansion leads to a fifth order V-statistic. However, some simple tricks will greatly simplify the problem.

We first split $\hat{\Gamma}_{p,x}$ into four terms, respectively

$$\begin{aligned}\hat{\Sigma}_{p,x,1} &= \frac{1}{n^3} \sum_{i,j,k} \mathbf{r}_p \left(\frac{x_j - x}{h} \right) \mathbf{r}_p \left(\frac{x_k - x}{h} \right)' K_h(x_j - x) K_h(x_k - x) \left(\mathbb{1}[x_i \leq x_j] - F(x_j) \right) \left(\mathbb{1}[x_i \leq x_k] - F(x_k) \right) \\ \hat{\Sigma}_{p,x,2} &= \frac{1}{n^3} \sum_{i,j,k} \mathbf{r}_p \left(\frac{x_j - x}{h} \right) \mathbf{r}_p \left(\frac{x_k - x}{h} \right)' K_h(x_j - x) K_h(x_k - x) \left(F(x_j) - \tilde{F}(x_j) \right) \left(\mathbb{1}[x_i \leq x_k] - \tilde{F}(x_k) \right) \\ \hat{\Sigma}_{p,x,3} &= \frac{1}{n^3} \sum_{i,j,k} \mathbf{r}_p \left(\frac{x_j - x}{h} \right) \mathbf{r}_p \left(\frac{x_k - x}{h} \right)' K_h(x_j - x) K_h(x_k - x) \left(\mathbb{1}[x_i \leq x_j] - \tilde{F}(x_j) \right) \left(F(x_k) - \tilde{F}(x_k) \right) \\ \hat{\Sigma}_{p,x,4} &= \frac{1}{n^3} \sum_{i,j,k} \mathbf{r}_p \left(\frac{x_j - x}{h} \right) \mathbf{r}_p \left(\frac{x_k - x}{h} \right)' K_h(x_j - x) K_h(x_k - x) \left(F(x_j) - \tilde{F}(x_j) \right) \left(F(x_k) - \tilde{F}(x_k) \right).\end{aligned}$$

Leaving $\hat{\Sigma}_{p,x,1}$ for a while, since it is the key component in this variance estimator. We first consider $\mathbf{N}_x \hat{\mathbf{S}}_{p,x}^{-1} \hat{\Sigma}_{p,x,4} \hat{\mathbf{S}}_{p,x}^{-1} \mathbf{N}_x$. By the uniform consistency of the empirical d.f., it can be shown easily that

$$\mathbf{N}_x \hat{\mathbf{S}}_{p,x}^{-1} \hat{\Sigma}_{p,x,4} \hat{\mathbf{S}}_{p,x}^{-1} \mathbf{N}_x = O_{\mathbb{P}}((nh)^{-1}).$$

Note that the extra h^{-1} comes from the scaling matrix \mathbf{N}_x , but not the kernel function K_h . Next we consider $\mathbf{N}_x \hat{\mathbf{S}}_{p,x}^{-1} \hat{\Sigma}_{p,x,2} \hat{\mathbf{S}}_{p,x}^{-1} \mathbf{N}_x$, which takes the following form (up to the negligible smoothing bias):

$$\begin{aligned}\mathbf{N}_x \hat{\mathbf{S}}_{p,x}^{-1} \hat{\Sigma}_{p,x,2} \hat{\mathbf{S}}_{p,x}^{-1} \mathbf{N}_x &= \mathbf{N}_x \mathbf{H}(\beta_p(x) - \hat{\beta}_p(x)) \left(\frac{1}{n^2} \sum_{i,k} \mathbf{r}_p \left(\frac{x_k - x}{h} \right)' K_h(x_k - x) \left(\mathbb{1}[x_i \leq x_k] - \tilde{F}(x_k) \right) \right) \hat{\mathbf{S}}_{p,x}^{-1} \mathbf{N}_x \\ &= O_{\mathbb{P}}((nh)^{-1/2}) = o_{\mathbb{P}}(1),\end{aligned}$$

where the last line uses the asymptotic normality of $\hat{\beta}_p(x)$. For $\hat{\Sigma}_{p,x,1}$, we make the observation that it is possible to ignore all ‘‘diagonal’’ terms, meaning that

$$\hat{\Sigma}_{p,x,1} = \frac{1}{n^3} \sum_{\substack{i,j,k \\ \text{distinct}}} \mathbf{r}_p \left(\frac{x_j - x}{h} \right) \mathbf{r}_p \left(\frac{x_k - x}{h} \right)' K_h(x_j - x) K_h(x_k - x) \left(\mathbb{1}[x_i \leq x_j] - F(x_j) \right) \left(\mathbb{1}[x_i \leq x_k] - F(x_k) \right) + o_{\mathbb{P}}(h),$$

under the assumption that $nh^2 \rightarrow \infty$. As a surrogate, define

$$\mathbf{U}_{i,j,k} = \mathbf{r}_p \left(\frac{x_j - x}{h} \right) \mathbf{r}_p \left(\frac{x_k - x}{h} \right)' K_h(x_j - x) K_h(x_k - x) \left(\mathbb{1}[x_i \leq x_j] - F(x_j) \right) \left(\mathbb{1}[x_i \leq x_k] - F(x_k) \right),$$

which means

$$\hat{\Sigma}_{p,x,1} = \frac{1}{n^3} \sum_{\substack{i,j,k \\ \text{distinct}}} \mathbf{U}_{i,j,k}.$$

The critical step is to further decompose the above into

$$\hat{\Sigma}_{p,x,1} = \frac{1}{n^3} \sum_{\substack{i,j,k \\ \text{distinct}}} \mathbb{E}[\mathbf{U}_{i,j,k} | x_i] \tag{I}$$

$$+ \frac{1}{n^3} \sum_{\substack{i,j,k \\ \text{distinct}}} \left(\mathbf{U}_{i,j,k} - \mathbb{E}[\mathbf{U}_{i,j,k} | x_i, x_j] \right) \tag{II}$$

$$+ \frac{1}{n^3} \sum_{\substack{i,j,k \\ \text{distinct}}} \left(\mathbb{E}[\mathbf{U}_{i,j,k} | x_i, x_j] - \mathbb{E}[\mathbf{U}_{i,j,k} | x_i] \right). \tag{III}$$

We already investigated the properties of term (I) in Lemma 3, hence it remains to show that both (II) and (III) are $o(h)$, hence does not affect the estimation of asymptotic variance. We consider (II) as an example, and the analysis of (III) is similar. Since (II) has zero expectation, we consider its variance (for simplicity treat \mathbf{U} as a scalar):

$$\mathbb{V}[(\text{II})] = \mathbb{E} \left[\frac{1}{n^6} \sum_{\substack{i,j,k \\ \text{distinct}}} \sum_{\substack{i',j',k' \\ \text{distinct}}} \left(\mathbf{U}_{i,j,k} - \mathbb{E}[\mathbf{U}_{i,j,k}|x_i, x_j] \right) \left(\mathbf{U}_{i',j',k'} - \mathbb{E}[\mathbf{U}_{i',j',k'}|x_{i'}, x_{j'}] \right) \right].$$

The expectation will be zero if the six indices are all distinct. Similarly, when there are only two indices among the six are equal, the expectation will be zero *unless* $k = k'$, hence

$$\begin{aligned} \mathbb{V}[(\text{II})] &= \mathbb{E} \left[\frac{1}{n^6} \sum_{\substack{i,j,k \\ \text{distinct}}} \sum_{\substack{i',j',k' \\ \text{distinct}}} \left(\mathbf{U}_{i,j,k} - \mathbb{E}[\mathbf{U}_{i,j,k}|x_i, x_j] \right) \left(\mathbf{U}_{i',j',k'} - \mathbb{E}[\mathbf{U}_{i',j',k'}|x_{i'}, x_{j'}] \right) \right] \\ &= \mathbb{E} \left[\frac{1}{n^6} \sum_{\substack{i,j,k,i'j' \\ \text{distinct}}} \left(\mathbf{U}_{i,j,k} - \mathbb{E}[\mathbf{U}_{i,j,k}|x_i, x_j] \right) \left(\mathbf{U}_{i',j',k} - \mathbb{E}[\mathbf{U}_{i',j',k}|x_{i'}, x_{j'}] \right) \right] \\ &\quad + \dots, \end{aligned}$$

where \dots represent cases where more than two indices among the six are equal. We can easily compute the order from the above as

$$\mathbb{V}[(\text{II})] = O(n^{-1}) + O((nh)^{-2}),$$

which shows that

$$(\text{II}) = O_{\mathbb{P}}(n^{-1/2} + (nh)^{-1}) = o_{\mathbb{P}}(h),$$

which closes the proof. ■

7.7 Proof of Lemma 5

Proof. We rely on Lemma 1 and 2 (note that whether the weights are estimated is irrelevant here), hence will not repeat arguments already established there. Instead, extra care will be given to ensure the characterization of higher order bias.

Consider the case where with enough smoothness on G , then the bias is characterized by

$$\begin{aligned} &h^{-v} v! \mathbf{e}'_v \left[f(x) \mathbf{S}_{p,x} + h F^{(2)}(x) \tilde{\mathbf{S}}_{p,x} + o(h) + O_{\mathbb{P}}(1/\sqrt{nh}) \right]^{-1} \\ &\quad \left[h^{p+1} \frac{F^{(p+1)}(x)}{(p+1)!} f(x) \mathbf{c}_{p,x} + h^{p+2} \left[\frac{F^{(p+2)}(x)}{(p+2)!} f(x) + \frac{F^{(p+1)}(x)}{(p+1)!} F^{(2)}(x) \right] \tilde{\mathbf{c}}_{p,x} + o(h^{p+2}) \right] \\ &= h^{-v} v! \mathbf{e}'_v \left[\frac{1}{f(x)} \mathbf{S}_{p,x}^{-1} - h \frac{F^{(2)}(x)}{[f(x)]^2} \mathbf{S}_{p,x}^{-1} \tilde{\mathbf{S}}_{p,x} \mathbf{S}_{p,x}^{-1} + O_{\mathbb{P}}(1/\sqrt{nh}) \right] \\ &\quad \left[h^{p+1} \frac{F^{(p+1)}(x)}{(p+1)!} f(x) \mathbf{c}_{p,x} + h^{p+2} \left[\frac{F^{(p+2)}(x)}{(p+2)!} f(x) + \frac{F^{(p+1)}(x)}{(p+1)!} F^{(2)}(x) \right] \tilde{\mathbf{c}}_{p,x} + o(h^{p+2}) \right] \{1 + o_{\mathbb{P}}(1)\}, \end{aligned}$$

which gives the desired result. Here $\tilde{\mathbf{S}}_{p,x} = \int_{\frac{x-h}{h}}^{\frac{x+h}{h}} \mathbf{u} \mathbf{r}_p(u) \mathbf{r}_p(u)' k(u) du$. And for the last line to hold, one needs the extra condition $nh^3 \rightarrow \infty$ so that $O_{\mathbb{P}}(1/\sqrt{nh}) = o_{\mathbb{P}}(h)$. See Fan and Gijbels (1996) (Theorem 3.1, pp. 62). ■

7.8 Proof of Lemma 6

Proof. The proof resembles that of Lemma 1, and is omitted here. ■

7.9 Proof of Theorem 3

Proof. The proof splits into two cases. We sketch one of them. Assume either x is boundary or $p - v$ is odd, the MSE-optimal bandwidth is asymptotically equivalent to the following:

$$\frac{\tilde{h}_{p,v}(x)}{h_{p,v}(x)} \rightarrow 1, \quad \tilde{h}_{p,v}(x) = \left(\frac{1}{n} \frac{(2v-1)f(x)\mathbf{e}'_v \mathbf{S}_{p,x}^{-1} \boldsymbol{\Gamma}_{p,x} \mathbf{S}_{p,x}^{-1} \mathbf{e}_v}{(2p-2v+2)\left(\frac{F^{(p+1)}(x)}{(p+1)!} \mathbf{e}'_v \mathbf{S}_{p,x}^{-1} \mathbf{c}_{p,x}\right)^2} \right)^{\frac{1}{2p+1}},$$

which is obtained by optimizing MSE ignoring the higher order bias term. With consistency of the preliminary estimates, it can be shown that

$$\hat{h}_{p,v}(x) = \left(\frac{1}{n} \frac{(2v-1)\hat{\sigma}_{p,v}(x)^2 n \ell^{2v-1}}{(2p-2v+2)\left(v! \frac{\tilde{F}^{(p+1)}(x)}{(p+1)!} \mathbf{e}'_v \mathbf{S}_{p,x}^{-1} \mathbf{c}_{p,x}\right)^2} \right)^{\frac{1}{2p+1}} \{1 + o_{\mathbb{P}}(1)\}.$$

Apply the consistency assumption of the preliminary estimates again, one can easily show that $\hat{h}_{p,v}(x)$ is consistent both in rate and constant.

A similar argument can be made for the other case, and is omitted here. \blacksquare

7.10 Proof of Lemma 7

Proof. This resembles the proof of Lemma 1, and we only perform the mean computation. To start,

$$\begin{aligned} \mathbb{E} \left[\frac{1}{n} \mathbf{X}'_h \mathbf{K}_h \mathbf{X}_h \right] &= \mathbb{E} \left[\mathbf{r}_p \left(\frac{x_i - \bar{x}}{h} \right) \mathbf{r}_p \left(\frac{x_i - \bar{x}}{h} \right)' \frac{1}{h} K \left(\frac{x_i - \bar{x}}{h} \right) \right] \\ &= \mathbb{E} \left[\mathbf{r}_p \left(\frac{x_i - \bar{x}}{h} \right) \mathbf{r}_p \left(\frac{x_i - \bar{x}}{h} \right)' \frac{1}{h} K \left(\frac{x_i - \bar{x}}{h} \right) \middle| x_i < \bar{x} \right] F(\bar{x}) \\ &\quad + \mathbb{E} \left[\mathbf{r}_p \left(\frac{x_i - \bar{x}}{h} \right) \mathbf{r}_p \left(\frac{x_i - \bar{x}}{h} \right)' \frac{1}{h} K \left(\frac{x_i - \bar{x}}{h} \right) \middle| x_i \geq \bar{x} \right] (1 - F(\bar{x})). \end{aligned}$$

Then by Lemma 1, the first term takes the form:

$$\begin{aligned} &\mathbb{E} \left[\mathbf{r}_p \left(\frac{x_i - \bar{x}}{h} \right) \mathbf{r}_p \left(\frac{x_i - \bar{x}}{h} \right)' \frac{1}{h} K \left(\frac{x_i - \bar{x}}{h} \right) \middle| x_i < \bar{x} \right] F(\bar{x}) \\ &= f(\bar{x} - |x_i < \bar{x}) F(\bar{x}) \int_{-1}^0 \mathbf{r}_{-,p}(u) \mathbf{r}_{-,p}(u)' K(u) du + O(h), \end{aligned}$$

where $f(\bar{x} - |x_i < \bar{x})$ is the one-sided density of x_i at the cutoff, conditional on $x_i < \bar{x}$. Alternatively, we can simplify by the fact that $f(\bar{x} | x_i < \bar{x}) F(\bar{x}) = f(\bar{x} -)$. Similarly, one has

$$\begin{aligned} &\mathbb{E} \left[\mathbf{r}_p \left(\frac{x_i - \bar{x}}{h} \right) \mathbf{r}_p \left(\frac{x_i - \bar{x}}{h} \right)' \frac{1}{h} K \left(\frac{x_i - \bar{x}}{h} \right) \middle| x_i \geq \bar{x} \right] (1 - F(\bar{x})) \\ &= f(\bar{x} + |x_i \geq \bar{x}) (1 - F(\bar{x})) \int_0^1 \mathbf{r}_{+,p}(u) \mathbf{r}_{+,p}(u)' K(u) du + O(h), \end{aligned}$$

and that $f(\bar{x} + |x_i \geq \bar{x}) (1 - F(\bar{x})) = f(\bar{x} +)$. The rest of the proof follows standard variance calculation, and is not repeated here. \blacksquare

7.11 Proof of Lemma 8

Proof. This follows from Lemma 2 by splitting the bias calculation for the two subsamples, below and above the cutoff \bar{x} . \blacksquare

7.12 Proof of Lemma 9

Proof. To start,

$$\int_{-1}^1 \mathbf{r}_p(u) \left(\tilde{F}(\bar{x} + hu) - F(\bar{x} + hu) \right) K(u) f(\bar{x} + hu) du = \frac{1}{n} \int_{-1}^1 \mathbf{r}_p(u) \left(\mathbf{1}[x_i \leq \bar{x} + hu] - F(\bar{x} + hu) \right) K(u) f(\bar{x} + hu) du,$$

and

$$\begin{aligned}
& \mathbb{V} \left[\int_{-1}^1 \mathbf{r}_p(u) \left(\mathbb{1}[x_i \leq \bar{x} + hu] - F(\bar{x} + hu) \right) K(u) f(\bar{x} + hu) du \right] \\
&= \iint_{-1}^1 \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(\bar{x} + hu) f(\bar{x} + hv) \\
&\quad \times \left[\int_{\mathbb{R}} (\mathbb{1}[t \leq \bar{x} + hu] - F(\bar{x} + hu)) (\mathbb{1}[t \leq \bar{x} + hv] - F(\bar{x} + hv)) f(t) dt \right] dudv \\
&= \iint_{-1}^1 \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(\bar{x} + hu) f(\bar{x} + hv) \left(F(\bar{x} + h(u \wedge v)) - F(\bar{x} + hu) F(\bar{x} + hv) \right) dudv. \quad (\text{I})
\end{aligned}$$

Now we split the integral of (I) into four regions.

$$\begin{aligned}
(u < 0, v < 0) \text{ (I)} &= \iint_{-1}^0 \mathbf{r}_{-,p}(u) \mathbf{r}_{-,p}(v)' K(u) K(v) f(\bar{x} + hu) f(\bar{x} + hv) \left(F(\bar{x} + h(u \wedge v)) - F(\bar{x} + hu) F(\bar{x} + hv) \right) dudv \\
&= f(\bar{x}-)^2 \left(F(\bar{x}) - F(\bar{x})^2 \right) \mathbf{S}_{-,p} \mathbf{e}_{0,-} \mathbf{e}'_{0,-} \mathbf{S}_{-,p} \\
&\quad - hf(\bar{x}-)^3 F(\bar{x}) \mathbf{S}_{-,p} (\mathbf{e}_{1,-} \mathbf{e}'_{0,-} + \mathbf{e}_{0,-} \mathbf{e}'_{1,-}) \mathbf{S}_{-,p} \\
&\quad + hf(\bar{x}-) F^{(2)}(\bar{x}-) \left(F(\bar{x}) - F(\bar{x})^2 \right) \mathbf{S}_{-,p} (\mathbf{e}_{1,-} \mathbf{e}'_{0,-} + \mathbf{e}_{0,-} \mathbf{e}'_{1,-}) \mathbf{S}_{-,p} \\
&\quad + hf(\bar{x}-)^3 \mathbf{\Gamma}_{-,p} + O(h^2),
\end{aligned}$$

and

$$\begin{aligned}
(u \geq 0, v \geq 0) \text{ (I)} &= \iint_0^1 \mathbf{r}_{+,p}(u) \mathbf{r}_{+,p}(v)' K(u) K(v) f(\bar{x} + hu) f(\bar{x} + hv) \left(F(\bar{x} + h(u \wedge v)) - F(\bar{x} + hu) F(\bar{x} + hv) \right) dudv \\
&= f(\bar{x}+)^2 \left(F(\bar{x}) - F(\bar{x})^2 \right) \mathbf{S}_{+,p} \mathbf{e}_{0,+} \mathbf{e}'_{0,+} \mathbf{S}_{+,p} \\
&\quad - hf(\bar{x}+)^3 F(\bar{x}) \mathbf{S}_{+,p} (\mathbf{e}_{1,+} \mathbf{e}'_{0,+} + \mathbf{e}_{0,+} \mathbf{e}'_{1,+}) \mathbf{S}_{+,p} \\
&\quad + hf(\bar{x}+) F^{(2)}(\bar{x}+) \left(F(\bar{x}) - F(\bar{x})^2 \right) \mathbf{S}_{+,p} (\mathbf{e}_{1,+} \mathbf{e}'_{0,+} + \mathbf{e}_{0,+} \mathbf{e}'_{1,+}) \mathbf{S}_{+,p} \\
&\quad + hf(\bar{x}+)^3 \mathbf{\Gamma}_{+,p} + O(h^2),
\end{aligned}$$

and

$$\begin{aligned}
(u < 0, v \geq 0) \text{ (I)} &= \iint_{[-1,0] \times [0,1]} \mathbf{r}_{-,p}(u) \mathbf{r}_{+,p}(v)' K(u) K(v) f(\bar{x} + hu) f(\bar{x} + hv) F(\bar{x} + hu) \left(1 - F(\bar{x} + hv) \right) dudv \\
&= \left[\int_{-1}^0 \mathbf{r}_{-,p}(u) K(u) f(\bar{x} + hu) F(\bar{x} + hu) du \right] \left[\int_0^1 \mathbf{r}_{+,p}(v)' K(v) f(\bar{x} + hv) \left(1 - F(\bar{x} + hv) \right) dv \right] \\
&= \left[f(\bar{x}-) F(\bar{x}) \mathbf{S}_{-,p} \mathbf{e}_{0,-} + h \left(f(\bar{x}-)^2 + F^{(2)}(\bar{x}-) F(\bar{x}) \right) \mathbf{S}_{-,p} \mathbf{e}_{1,-} + O(h^2) \right] \\
&\quad \left[f(\bar{x}+) (1 - F(\bar{x})) \mathbf{S}_{+,p} \mathbf{e}_{0,+} + h \left(-f(\bar{x}+)^2 + F^{(2)}(\bar{x}+) (1 - F(\bar{x})) \right) \mathbf{S}_{+,p} \mathbf{e}_{1,+} + O(h^2) \right]',
\end{aligned}$$

and

$$\begin{aligned}
(u \geq 0, v < 0) \text{ (I)} &= \iint_{[0,1] \times [-1,0]} \mathbf{r}_{-,p}(u) \mathbf{r}_{+,p}(v)' K(u) K(v) f(\bar{x} + hu) f(\bar{x} + hv) F(\bar{x} + hv) \left(1 - F(\bar{x} + hu) \right) dudv \\
&= \left[\int_0^1 \mathbf{r}_{+,p}(u) K(u) f(\bar{x} + hu) (1 - F(\bar{x} + hu)) du \right] \left[\int_{-1}^0 \mathbf{r}_{-,p}(v)' K(v) f(\bar{x} + hv) F(\bar{x} + hv) dv \right] \\
&= \left[f(\bar{x}+) (1 - F(\bar{x})) \mathbf{S}_{+,p} \mathbf{e}_{0,+} + h \left(-f(\bar{x}+)^2 + F^{(2)}(\bar{x}+) (1 - F(\bar{x})) \right) \mathbf{S}_{+,p} \mathbf{e}_{1,+} + O(h^2) \right] \\
&\quad \left[f(\bar{x}-) F(\bar{x}-) \mathbf{S}_{-,p} \mathbf{e}_{0,-} + h \left(f(\bar{x}-)^2 + F^{(2)}(\bar{x}-) F(\bar{x}) \right) \mathbf{S}_{-,p} \mathbf{e}_{1,-} + O(h^2) \right]'.
\end{aligned}$$

Let $\mathbf{S}_{-,p}^{-1}$ and $\mathbf{S}_{+,p}^{-1}$ be the MoorePenrose inverse of $\mathbf{S}_{-,p}$ and $\mathbf{S}_{+,p}$, respectively. Then

$$\begin{aligned} & \mathbb{V} \left[(\mathbf{e}_{1,+} - \mathbf{e}_{1,-})' \sqrt{\frac{n}{h}} (f(\bar{x}+) \mathbf{S}_{+,p} + f(\bar{x}-) \mathbf{S}_{-,p})^{-1} \hat{\mathbf{L}} \right] \\ &= f(\bar{x}-) \mathbf{e}'_{1,-} \mathbf{S}_{-,p}^{-1} \mathbf{\Gamma}_{-,p} \mathbf{S}_{-,p}^{-1} \mathbf{e}_{1,-} + f(\bar{x}+) \mathbf{e}'_{1,+} \mathbf{S}_{+,p}^{-1} \mathbf{\Gamma}_{+,p} \mathbf{S}_{+,p}^{-1} \mathbf{e}_{1,+} + O(h). \end{aligned}$$

■

7.13 Proof of Lemma 10

Proof. This follows from Lemma 4 by splitting the bias calculation for the two subsamples, below and above the cutoff \bar{x} . ■

7.14 Proof of Corollary 1

Proof. This follows from the previous lemmas and verifying the Lindeberg condition. See also the proof of Lemma 3, Theorem 1 and Theorem 2. ■

7.15 Proof of Lemma 11

Proof. This follows from Lemma 1 by splitting the bias calculation for the two subsamples, below and above the cutoff \bar{x} . See also the proof of Lemma 7. ■

7.16 Proof of Lemma 12

Proof. This follows from Lemma 2 by splitting the bias calculation for the two subsamples, below and above the cutoff \bar{x} . ■

7.17 Proof of Lemma 13

Proof. To start,

$$\int_{-1}^1 \mathbf{r}_p(u) \left(\tilde{F}(\bar{x} + hu) - F(\bar{x} + hu) \right) K(u) f(\bar{x} + hu) du = \frac{1}{n} \int_{-1}^1 \mathbf{r}_p(u) \left(\mathbb{1}[x_i \leq \bar{x} + hu] - F(\bar{x} + hu) \right) K(u) f(\bar{x} + hu) du,$$

and

$$\begin{aligned} & \mathbb{V} \left[\int_{-1}^1 \mathbf{r}_p(u) \left(\mathbb{1}[x_i \leq \bar{x} + hu] - F(\bar{x} + hu) \right) K(u) f(\bar{x} + hu) du \right] \\ &= \iint_{-1}^1 \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(\bar{x} + hu) f(\bar{x} + hv) \\ & \quad \times \left[\int_{\mathbb{R}} (\mathbb{1}[t \leq \bar{x} + hu] - F(\bar{x} + hu)) (\mathbb{1}[t \leq \bar{x} + hv] - F(\bar{x} + hv)) f(t) dt \right] dudv \\ &= \iint_{-1}^1 \mathbf{r}_p(u) \mathbf{r}_p(v)' K(u) K(v) f(\bar{x} + hu) f(\bar{x} + hv) \left(F(\bar{x} + h(u \wedge v)) - F(\bar{x} + hu) F(\bar{x} + hv) \right) dudv. \quad (\text{I}) \end{aligned}$$

Now we split the integral of (I) into four regions.

$$\begin{aligned} (u < 0, v < 0) \text{ (I)} &= \iint_{-1}^0 \mathbf{r}_{-,p}(u) \mathbf{r}_{-,p}(v)' K(u) K(v) f(\bar{x} + hu) f(\bar{x} + hv) \left(F(\bar{x} + h(u \wedge v)) - F(\bar{x} + hu) F(\bar{x} + hv) \right) dudv \\ &= f(\bar{x}-)^2 \left(F(\bar{x}) - F(\bar{x})^2 \right) \mathbf{S}_{-,p} \mathbf{e}_0 \mathbf{e}'_0 \mathbf{S}_{-,p} \\ & \quad - hf(\bar{x}-)^3 F(\bar{x}) \mathbf{S}_{-,p} (\mathbf{e}_{1,-} \mathbf{e}'_0 + \mathbf{e}_0 \mathbf{e}'_{1,-}) \mathbf{S}_{-,p} \\ & \quad + hf(\bar{x}-) F^{(2)}(\bar{x}) \left(F(\bar{x}) - F(\bar{x})^2 \right) \mathbf{S}_{-,p} (\mathbf{e}_{1,-} \mathbf{e}'_0 + \mathbf{e}_0 \mathbf{e}'_{1,-}) \mathbf{S}_{-,p} \\ & \quad + hf(\bar{x}-)^3 \mathbf{\Gamma}_{-,p} + O(h^2), \end{aligned}$$

and

$$\begin{aligned}
(u \geq 0, v \geq 0) \text{ (I)} &= \iint_0^1 \mathbf{r}_{+,p}(u) \mathbf{r}_{+,p}(v)' K(u)K(v) f(\bar{x} + hu) f(\bar{x} + hv) \left(F(\bar{x} + h(u \wedge v)) - F(\bar{x} + hu)F(\bar{x} + hv) \right) dudv \\
&= f(\bar{x} +)^2 \left(F(\bar{x}) - F(\bar{x})^2 \right) \mathbf{S}_{+,p} \mathbf{e}_0 \mathbf{e}'_0 \mathbf{S}_{+,p} \\
&\quad - hf(\bar{x} +)^3 F(\bar{x}) \mathbf{S}_{+,p} (\mathbf{e}_{1,+} \mathbf{e}'_0 + \mathbf{e}_0 \mathbf{e}'_{1,+}) \mathbf{S}_{+,p} \\
&\quad + hf(\bar{x} +) F^{(2)}(\bar{x}) \left(F(\bar{x}) - F(\bar{x})^2 \right) \mathbf{S}_{+,p} (\mathbf{e}_{1,+} \mathbf{e}'_0 + \mathbf{e}_0 \mathbf{e}'_{1,+}) \mathbf{S}_{+,p} \\
&\quad + hf(\bar{x} +)^3 \mathbf{\Gamma}_{+,p} + O(h^2),
\end{aligned}$$

and

$$\begin{aligned}
(u < 0, v \geq 0) \text{ (I)} &= \iint_{[-1,0] \times [0,1]} \mathbf{r}_{-,p}(u) \mathbf{r}_{+,p}(v)' K(u)K(v) f(\bar{x} + hu) f(\bar{x} + hv) F(\bar{x} + hu) \left(1 - F(\bar{x} + hv) \right) dudv \\
&= \left[\int_{-1}^0 \mathbf{r}_{-,p}(u) K(u) f(\bar{x} + hu) F(\bar{x} + hu) du \right] \left[\int_0^1 \mathbf{r}_{+,p}(v)' K(v) f(\bar{x} + hv) \left(1 - F(\bar{x} + hv) \right) dv \right] \\
&= \left[f(\bar{x} -) F(\bar{x}) \mathbf{S}_{-,p} \mathbf{e}_0 + h \left(f(\bar{x} -)^2 + F^{(2)}(\bar{x}) F(\bar{x}) \right) \mathbf{S}_{-,p} \mathbf{e}_{1,-} + O(h^2) \right] \\
&\quad \left[f(\bar{x} +) (1 - F(\bar{x})) \mathbf{S}_{+,p} \mathbf{e}_0 + h \left(-f(\bar{x} +)^2 + F^{(2)}(\bar{x}) (1 - F(\bar{x})) \right) \mathbf{S}_{+,p} \mathbf{e}_{1,+} + O(h^2) \right]',
\end{aligned}$$

and

$$\begin{aligned}
(u \geq 0, v < 0) \text{ (I)} &= \iint_{[0,1] \times [-1,0]} \mathbf{r}_{-,p}(u) \mathbf{r}_{+,p}(v)' K(u)K(v) f(\bar{x} + hu) f(\bar{x} + hv) F(\bar{x} + hv) \left(1 - F(\bar{x} + hu) \right) dudv \\
&= \left[\int_0^1 \mathbf{r}_{+,p}(u) K(u) f(\bar{x} + hu) (1 - F(\bar{x} + hu)) du \right] \left[\int_{-1}^0 \mathbf{r}_{-,p}(v)' K(v) f(\bar{x} + hv) F(\bar{x} + hv) dv \right] \\
&= \left[f(\bar{x} +) (1 - F(\bar{x})) \mathbf{S}_{+,p} \mathbf{e}_0 + h \left(-f(\bar{x} +)^2 + F^{(2)}(\bar{x}) (1 - F(\bar{x})) \right) \mathbf{S}_{+,p} \mathbf{e}_{1,+} + O(h^2) \right] \\
&\quad \left[f(\bar{x} -) F(\bar{x}) \mathbf{S}_{-,p} \mathbf{e}_0 + h \left(f(\bar{x} -)^2 + F^{(2)}(\bar{x}) F(\bar{x}) \right) \mathbf{S}_{-,p} \mathbf{e}_{1,-} + O(h^2) \right]'.
\end{aligned}$$

By collecting terms, one has

$$\begin{aligned}
\text{(I)} &= \left(f(\bar{x} +) \mathbf{S}_{+,p} + f(\bar{x} +) \mathbf{S}_{-,p} \right) \mathbf{e}_0 \mathbf{e}'_0 \left(f(\bar{x} +) \mathbf{S}_{+,p} + f(\bar{x} +) \mathbf{S}_{-,p} \right)' \\
&\quad - hf(\bar{x} -) F(\bar{x}) f(\bar{x} -) \mathbf{S}_{-,p} \mathbf{e}_{1,-} \mathbf{e}'_0 (f(\bar{x} +) \mathbf{S}_{+,p} + f(\bar{x} -) \mathbf{S}_{-,p}) \\
&\quad + h \frac{F^{(2)}(\bar{x})}{f(\bar{x} -)} F(\bar{x}) (1 - F(\bar{x})) f(\bar{x} -) \mathbf{S}_{-,p} \mathbf{e}_{1,-} \mathbf{e}'_0 (f(\bar{x} +) \mathbf{S}_{+,p} + f(\bar{x} -) \mathbf{S}_{-,p}) \\
&\quad - hf(\bar{x} -) F(\bar{x}) (f(\bar{x} +) \mathbf{S}_{+,p} + f(\bar{x} -) \mathbf{S}_{-,p}) \mathbf{e}_0 \mathbf{e}'_{1,-} f(\bar{x} -) \mathbf{S}_{-,p} \\
&\quad + h \frac{F^{(2)}(\bar{x})}{f(\bar{x} -)} F(\bar{x}) (1 - F(\bar{x})) (f(\bar{x} +) \mathbf{S}_{+,p} + f(\bar{x} -) \mathbf{S}_{-,p}) \mathbf{e}_0 \mathbf{e}'_{1,-} f(\bar{x} -) \mathbf{S}_{-,p} \\
&\quad - hf(\bar{x} +) F(\bar{x}) f(\bar{x} +) \mathbf{S}_{+,p} \mathbf{e}_{1,+} \mathbf{e}'_0 (f(\bar{x} +) \mathbf{S}_{+,p} + f(\bar{x} -) \mathbf{S}_{-,p}) \\
&\quad + h \frac{F^{(2)}(\bar{x})}{f(\bar{x} +)} (1 - F(\bar{x})) F(\bar{x}) f(\bar{x} +) \mathbf{S}_{+,p} \mathbf{e}_{1,+} \mathbf{e}'_0 (f(\bar{x} +) \mathbf{S}_{+,p} + f(\bar{x} -) \mathbf{S}_{-,p}) \\
&\quad - hf(\bar{x} +) F(\bar{x}) (f(\bar{x} +) \mathbf{S}_{+,p} + f(\bar{x} -) \mathbf{S}_{-,p}) \mathbf{e}_0 \mathbf{e}'_{1,+} f(\bar{x} +) \mathbf{S}_{+,p} \\
&\quad + h \frac{F^{(2)}(\bar{x})}{f(\bar{x} +)} F(\bar{x}) (1 - F(\bar{x})) (f(\bar{x} +) \mathbf{S}_{+,p} + f(\bar{x} -) \mathbf{S}_{-,p}) \mathbf{e}_0 \mathbf{e}'_{1,+} f(\bar{x} +) \mathbf{S}_{+,p} \\
&\quad + hf(\bar{x} -) f(\bar{x} -) \mathbf{S}_{-,p} \mathbf{e}_{1,-} \mathbf{e}'_0 f(\bar{x} +) \mathbf{S}_{+,p} \\
&\quad + hf(\bar{x} -) f(\bar{x} +) \mathbf{S}_{+,p} \mathbf{e}_0 \mathbf{e}'_{1,-} f(\bar{x} -) \mathbf{S}_{-,p} \\
&\quad + h(f(\bar{x} +)^3 \mathbf{\Gamma}_{+,p} + f(\bar{x} -)^3 \mathbf{\Gamma}_{-,p}).
\end{aligned}$$

Next, we note that

$$\mathbf{S}_{+,p} \mathbf{e}_{1,-} = \mathbf{S}_{-,p} \mathbf{e}_{1,+} = \mathbf{0},$$

Therefore,

$$\begin{aligned}
& \mathbb{V} \left[(\mathbf{e}_{1,+} - \mathbf{e}_{1,-})' \sqrt{\frac{n}{h}} (f(\bar{x}+) \mathbf{S}_{+,p} + f(\bar{x}-) \mathbf{S}_{-,p})^{-1} \hat{\mathbf{L}} \right] \\
&= (\mathbf{e}_{1,+} - \mathbf{e}_{1,-})' (f(\bar{x}+) \mathbf{S}_{+,p} + f(\bar{x}-) \mathbf{S}_{-,p})^{-1} (f(\bar{x}+)^3 \mathbf{\Gamma}_{+,p} \\
&\quad + f(\bar{x}-)^3 \mathbf{\Psi} \mathbf{\Gamma}_{+,p} \mathbf{\Psi}) (f(\bar{x}+) \mathbf{S}_{+,p} + f(\bar{x}-) \mathbf{S}_{-,p})^{-1} (\mathbf{e}_{1,+} - \mathbf{e}_{1,-}) + O(h).
\end{aligned}$$

■

7.18 Proof of Lemma 14

Proof. This follows from Lemma 4 by splitting the bias calculation for the two subsamples, below and above the cutoff \bar{x} . ■

7.19 Proof of Corollary 2

Proof. This follows from the previous lemmas and verifying the Lindeberg condition. See also the proof of Lemma 3, Theorem 1 and Theorem 2. ■

Table 1. Simulation (truncated Normal). $x = -0.8$, $p = 2$, triangular kernel.

(a) $n = 1000$

	\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$					
0.1	0.008	0.170	0.170	0.168	6.00
0.3	0.003	0.097	0.097	0.094	5.98
0.5	0.002	0.074	0.074	0.074	4.92
0.7	0.006	0.062	0.062	0.063	4.76
0.9	0.013	0.054	0.056	0.056	4.40
1	0.017	0.052	0.054	0.053	4.46
1.1	0.021	0.050	0.054	0.051	4.74
1.3	0.031	0.046	0.056	0.047	5.06
1.5	0.044	0.043	0.061	0.044	4.64
1.7	0.058	0.040	0.071	0.041	4.60
1.9	0.074	0.038	0.083	0.039	4.88
\hat{h}					
	0.013	0.070	0.071	0.066	8.24

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.4	0.475	0.607	0.861	1.277

(b) $n = 2000$

	\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$					
0.1	0.003	0.121	0.121	0.122	5.14
0.3	0.002	0.071	0.071	0.070	5.54
0.5	0.002	0.055	0.055	0.055	5.32
0.7	0.004	0.046	0.046	0.047	4.84
0.9	0.009	0.040	0.042	0.041	4.22
1	0.012	0.038	0.040	0.039	4.36
1.1	0.016	0.037	0.040	0.038	4.46
1.3	0.024	0.034	0.042	0.035	4.56
1.5	0.034	0.032	0.046	0.033	4.56
1.7	0.045	0.030	0.054	0.031	4.62
1.9	0.058	0.028	0.065	0.029	4.62
\hat{h}					
	0.012	0.054	0.055	0.048	9.50

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.407	0.492	0.638	0.906	1.36

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 2. Simulation (truncated Normal). $x = -0.8$, $p = 3$, triangular kernel.

(a) $n = 1000$						(b) $n = 2000$					
	\hat{f}_p			SE			\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size		bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$						$h_{\text{MSE}} \times$					
0.1	0.015	0.234	0.234	0.229	4.94	0.1	0.006	0.161	0.161	0.161	4.46
0.3	0.005	0.129	0.129	0.127	5.50	0.3	0.003	0.094	0.094	0.093	5.36
0.5	-0.001	0.100	0.100	0.100	4.96	0.5	-0.001	0.074	0.074	0.073	5.70
0.7	-0.004	0.085	0.085	0.085	4.58	0.7	-0.003	0.062	0.062	0.062	5.24
0.9	-0.004	0.075	0.075	0.076	4.58	0.9	-0.004	0.055	0.055	0.056	4.52
1	-0.005	0.071	0.071	0.072	4.50	1	-0.005	0.052	0.053	0.053	4.54
1.1	-0.006	0.068	0.069	0.069	4.86	1.1	-0.006	0.050	0.051	0.051	4.50
1.3	-0.007	0.064	0.064	0.064	5.12	1.3	-0.008	0.047	0.047	0.047	4.60
1.5	-0.007	0.059	0.060	0.060	4.92	1.5	-0.008	0.043	0.044	0.044	4.44
1.7	-0.005	0.056	0.056	0.057	4.64	1.7	-0.007	0.041	0.042	0.042	4.54
1.9	0.000	0.053	0.053	0.054	4.92	1.9	-0.004	0.039	0.039	0.040	4.78
\hat{h}						\hat{h}					
	0.001	0.114	0.114	0.110	5.06		0.000	0.081	0.081	0.078	5.58
	Quantile						Quantile				
	0.10	0.25	0.50	0.75	0.90		0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.317	0.344	0.387	0.462	0.59	\hat{h}/h_{MSE}	0.332	0.359	0.402	0.483	0.628

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 3. Simulation (truncated Normal). $x = -0.5$, $p = 2$, triangular kernel.

(a) $n = 1000$						(b) $n = 2000$					
	\hat{f}_p			SE			\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size		bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$						$h_{\text{MSE}} \times$					
0.1	0.003	0.067	0.067	0.068	5.34	0.1	0.002	0.052	0.052	0.051	5.46
0.3	0.001	0.037	0.037	0.037	5.14	0.3	0.000	0.028	0.028	0.029	5.28
0.5	-0.002	0.028	0.028	0.028	5.26	0.5	-0.002	0.021	0.021	0.021	5.76
0.7	-0.003	0.026	0.026	0.026	4.86	0.7	-0.003	0.018	0.019	0.019	5.48
0.9	-0.003	0.026	0.026	0.026	4.90	0.9	-0.003	0.018	0.018	0.018	5.46
1	-0.002	0.026	0.026	0.026	4.86	1	-0.003	0.018	0.018	0.018	5.30
1.1	-0.001	0.026	0.026	0.026	4.86	1.1	-0.003	0.018	0.018	0.018	5.28
1.3	0.001	0.026	0.026	0.026	4.76	1.3	-0.001	0.018	0.018	0.018	5.12
1.5	0.005	0.026	0.026	0.026	4.88	1.5	0.001	0.018	0.018	0.018	4.68
1.7	0.010	0.025	0.027	0.026	5.02	1.7	0.005	0.018	0.019	0.018	4.58
1.9	0.016	0.025	0.030	0.025	4.68	1.9	0.009	0.018	0.020	0.018	4.50
\hat{h}						\hat{h}					
	0.004	0.031	0.031	0.026	9.78		0.005	0.027	0.027	0.018	12.86

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.716	0.828	1.004	1.299	1.861

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.728	0.843	1.044	1.433	2.223

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 4. Simulation (truncated Normal). $x = -0.5$, $p = 3$, triangular kernel.

(a) $n = 1000$

	\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$					
0.1	0.001	0.061	0.061	0.061	5.08
0.3	0.001	0.036	0.036	0.035	5.46
0.5	0.001	0.027	0.027	0.027	4.94
0.7	0.001	0.025	0.025	0.025	4.90
0.9	0.001	0.025	0.025	0.025	5.00
1	0.002	0.026	0.026	0.026	5.16
1.1	0.002	0.026	0.026	0.026	5.18
1.3	0.005	0.026	0.027	0.026	4.80
1.5	0.010	0.026	0.028	0.026	4.74
1.7	0.017	0.026	0.031	0.026	4.70
1.9	0.025	0.026	0.036	0.026	4.66
\hat{h}					
	0.001	0.031	0.031	0.031	4.56

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}					
	0.289	0.314	0.361	0.437	0.587

(b) $n = 2000$

	\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$					
0.1	0.001	0.046	0.046	0.045	5.12
0.3	0.001	0.026	0.026	0.026	5.54
0.5	0.001	0.019	0.019	0.020	4.90
0.7	0.001	0.017	0.017	0.018	4.84
0.9	0.001	0.018	0.018	0.018	4.86
1	0.001	0.018	0.018	0.018	4.84
1.1	0.002	0.018	0.018	0.018	4.74
1.3	0.003	0.018	0.019	0.018	4.72
1.5	0.007	0.018	0.020	0.019	4.60
1.7	0.013	0.018	0.022	0.019	4.38
1.9	0.019	0.018	0.027	0.018	4.32
\hat{h}					
	0.001	0.022	0.022	0.023	4.74

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}					
	0.3	0.327	0.374	0.457	0.619

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 5. Simulation (truncated Normal). $x = 0.5$, $p = 2$, triangular kernel.

(a) $n = 1000$						(b) $n = 2000$					
	\hat{f}_p			SE			\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size		bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$						$h_{\text{MSE}} \times$					
0.1	0.004	0.068	0.068	0.068	5.26	0.1	0.002	0.051	0.051	0.051	5.10
0.3	0.000	0.037	0.037	0.037	5.04	0.3	0.000	0.028	0.028	0.028	4.56
0.5	-0.002	0.027	0.027	0.028	4.84	0.5	-0.001	0.021	0.021	0.021	4.62
0.7	-0.004	0.022	0.022	0.022	4.64	0.7	-0.003	0.017	0.017	0.017	4.86
0.9	-0.008	0.018	0.020	0.018	4.68	0.9	-0.006	0.014	0.016	0.014	4.90
1	-0.010	0.017	0.020	0.017	4.92	1	-0.007	0.013	0.015	0.013	4.98
1.1	-0.012	0.016	0.020	0.016	4.92	1.1	-0.009	0.013	0.015	0.012	5.08
1.3	-0.016	0.014	0.021	0.014	5.20	1.3	-0.013	0.011	0.017	0.011	5.00
1.5	-0.021	0.012	0.024	0.012	5.72	1.5	-0.016	0.010	0.019	0.010	5.16
1.7	-0.026	0.011	0.028	0.011	6.04	1.7	-0.020	0.009	0.022	0.009	5.24
1.9	-0.031	0.010	0.032	0.009	6.44	1.9	-0.025	0.008	0.026	0.008	5.46
\hat{h}						\hat{h}					
	-0.009	0.023	0.025	0.017	19.00		-0.007	0.018	0.020	0.013	17.38

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.748	0.829	0.971	1.256	1.785

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.772	0.849	0.976	1.214	1.703

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 6. Simulation (truncated Normal). $x = 0.5$, $p = 3$, triangular kernel.

(a) $n = 1000$						(b) $n = 2000$					
	\hat{f}_p			SE			\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size		bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$						$h_{\text{MSE}} \times$					
0.1	0.001	0.061	0.061	0.061	5.04	0.1	0.001	0.045	0.045	0.045	4.88
0.3	0.001	0.033	0.033	0.034	4.88	0.3	0.000	0.024	0.024	0.025	4.54
0.5	0.001	0.024	0.024	0.025	4.68	0.5	0.001	0.018	0.018	0.018	4.72
0.7	0.001	0.019	0.019	0.020	4.70	0.7	0.001	0.015	0.015	0.015	5.18
0.9	0.000	0.017	0.017	0.017	5.06	0.9	0.001	0.012	0.012	0.012	5.10
1	-0.002	0.016	0.016	0.016	5.18	1	-0.001	0.012	0.012	0.012	5.06
1.1	-0.004	0.015	0.016	0.015	5.30	1.1	-0.002	0.011	0.011	0.011	5.04
1.3	-0.009	0.014	0.017	0.014	5.64	1.3	-0.007	0.010	0.012	0.010	5.10
1.5	-0.014	0.013	0.019	0.013	5.60	1.5	-0.011	0.010	0.015	0.009	5.32
1.7	-0.018	0.012	0.022	0.012	5.80	1.7	-0.015	0.009	0.018	0.009	5.42
1.9	-0.021	0.012	0.024	0.011	6.10	1.9	-0.019	0.009	0.021	0.008	5.88
\hat{h}						\hat{h}					
	-0.001	0.020	0.020	0.018	7.92		-0.001	0.015	0.015	0.013	8.98

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}					
	0.641	0.697	0.784	0.926	1.156

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}					
	0.664	0.721	0.81	0.96	1.176

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 7. Simulation (truncated Normal). $x = 1.5$, $p = 2$, triangular kernel.

(a) $n = 1000$

	\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$					
0.1	0.005	0.042	0.042	0.043	6.48
0.3	0.002	0.024	0.024	0.024	5.12
0.5	0.003	0.018	0.018	0.018	4.68
0.7	0.004	0.015	0.016	0.015	4.24
0.9	0.005	0.013	0.014	0.013	4.36
1	0.006	0.012	0.014	0.013	4.22
1.1	0.007	0.012	0.014	0.012	4.20
1.3	0.009	0.011	0.014	0.011	4.26
1.5	0.011	0.010	0.015	0.011	4.16
1.7	0.013	0.010	0.016	0.010	3.96
1.9	0.014	0.010	0.017	0.010	4.02
\hat{h}	0.006	0.016	0.017	0.012	12.76

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.726	0.837	1.035	1.388	2.029

(b) $n = 2000$

	\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$					
0.1	0.002	0.032	0.032	0.032	5.74
0.3	0.001	0.018	0.018	0.018	5.08
0.5	0.002	0.014	0.014	0.014	5.18
0.7	0.003	0.012	0.012	0.012	5.04
0.9	0.004	0.010	0.011	0.010	5.08
1	0.005	0.010	0.011	0.010	4.90
1.1	0.006	0.009	0.011	0.009	4.76
1.3	0.007	0.008	0.011	0.008	4.68
1.5	0.009	0.008	0.012	0.008	4.60
1.7	0.011	0.007	0.013	0.008	4.54
1.9	0.012	0.007	0.014	0.007	4.44
\hat{h}	0.005	0.013	0.014	0.009	14.70

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.758	0.863	1.042	1.381	1.96

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 8. Simulation (truncated Normal). $x = 1.5$, $p = 3$, triangular kernel.

(a) $n = 1000$						(b) $n = 2000$					
	\hat{f}_p			SE			\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size		bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$						$h_{\text{MSE}} \times$					
0.1	0.002	0.045	0.045	0.045	6.12	0.1	0.001	0.033	0.033	0.033	5.82
0.3	0.001	0.026	0.026	0.026	5.34	0.3	0.000	0.019	0.019	0.019	5.08
0.5	0.001	0.020	0.020	0.020	4.90	0.5	0.000	0.015	0.015	0.015	5.10
0.7	0.002	0.017	0.017	0.017	4.34	0.7	0.001	0.013	0.013	0.013	5.10
0.9	0.003	0.015	0.016	0.016	4.38	0.9	0.002	0.011	0.011	0.011	4.80
1	0.004	0.015	0.015	0.015	4.12	1	0.003	0.011	0.011	0.011	4.54
1.1	0.006	0.014	0.016	0.015	4.30	1.1	0.004	0.010	0.011	0.011	4.60
1.3	0.010	0.013	0.017	0.014	4.38	1.3	0.008	0.010	0.013	0.010	4.82
1.5	0.017	0.013	0.021	0.013	4.36	1.5	0.013	0.009	0.016	0.009	4.86
1.7	0.024	0.012	0.027	0.012	4.34	1.7	0.019	0.009	0.021	0.009	4.78
1.9	0.032	0.011	0.034	0.012	4.70	1.9	0.026	0.008	0.028	0.009	4.38
\hat{h}						\hat{h}					
	0.006	0.018	0.019	0.015	10.70		0.005	0.014	0.015	0.011	12.12

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}					
	0.79	0.874	0.993	1.18	1.487

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}					
	0.827	0.91	1.033	1.235	1.533

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 9. Simulation (Exponential). $x = 0$, $p = 2$, triangular kernel.

(a) $n = 1000$						(b) $n = 2000$					
	\hat{f}_p			SE			\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size		bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$						$h_{\text{MSE}} \times$					
0.1	0.005	0.259	0.259	0.254	5.90	0.1	0.000	0.187	0.187	0.185	5.36
0.3	-0.002	0.142	0.142	0.140	5.46	0.3	0.000	0.103	0.103	0.103	5.54
0.5	-0.009	0.105	0.106	0.104	5.16	0.5	-0.007	0.078	0.078	0.077	5.08
0.7	-0.020	0.085	0.087	0.085	4.74	0.7	-0.017	0.063	0.065	0.063	4.90
0.9	-0.032	0.072	0.079	0.072	5.04	0.9	-0.028	0.054	0.061	0.053	4.92
1	-0.038	0.068	0.077	0.067	5.08	1	-0.034	0.051	0.061	0.050	4.88
1.1	-0.044	0.063	0.077	0.063	5.58	1.1	-0.039	0.048	0.062	0.047	4.90
1.3	-0.058	0.057	0.081	0.056	5.62	1.3	-0.052	0.042	0.067	0.041	5.12
1.5	-0.072	0.051	0.088	0.050	5.98	1.5	-0.065	0.038	0.075	0.037	5.36
1.7	-0.086	0.046	0.098	0.046	6.14	1.7	-0.078	0.035	0.085	0.034	5.86
1.9	-0.100	0.043	0.109	0.042	5.64	1.9	-0.091	0.032	0.097	0.031	5.88
\hat{h}						\hat{h}					
	-0.033	0.094	0.100	0.089	7.74		-0.031	0.073	0.079	0.064	8.70
<hr/>						<hr/>					
	Quantile						Quantile				
	0.10	0.25	0.50	0.75	0.90		0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.749	0.836	0.99	1.274	1.94	\hat{h}/h_{MSE}	0.815	0.916	1.089	1.438	2.155

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 10. Simulation (Exponential). $x = 0$, $p = 3$, triangular kernel.

(a) $n = 1000$						(b) $n = 2000$					
	\hat{f}_p			SE			\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size		bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$						$h_{\text{MSE}} \times$					
0.1	0.008	0.332	0.332	0.326	5.30	0.1	0.000	0.239	0.239	0.237	5.52
0.3	0.003	0.182	0.182	0.179	5.24	0.3	0.004	0.132	0.132	0.132	5.26
0.5	0.002	0.135	0.135	0.134	5.26	0.5	0.003	0.099	0.099	0.098	4.84
0.7	-0.003	0.108	0.108	0.109	4.52	0.7	-0.001	0.080	0.080	0.080	4.78
0.9	-0.007	0.092	0.093	0.093	4.58	0.9	-0.005	0.069	0.069	0.069	5.02
1	-0.009	0.086	0.087	0.087	5.02	1	-0.007	0.065	0.065	0.064	4.80
1.1	-0.012	0.081	0.082	0.081	4.90	1.1	-0.010	0.061	0.062	0.060	4.88
1.3	-0.018	0.073	0.075	0.073	5.60	1.3	-0.015	0.055	0.057	0.054	5.06
1.5	-0.024	0.067	0.071	0.066	5.58	1.5	-0.021	0.050	0.054	0.049	5.42
1.7	-0.031	0.061	0.068	0.060	6.16	1.7	-0.027	0.046	0.053	0.045	5.50
1.9	-0.039	0.056	0.068	0.055	5.76	1.9	-0.034	0.042	0.054	0.041	5.60
\hat{h}						\hat{h}					
	0.000	0.135	0.135	0.128	5.96		0.000	0.097	0.097	0.092	5.78

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}					
	0.47	0.511	0.574	0.678	0.877

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}					
	0.49	0.531	0.594	0.705	0.9

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 11. Simulation (Exponential). $x = 1$, $p = 2$, triangular kernel.

(a) $n = 1000$

	\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$					
0.1	0.006	0.065	0.065	0.065	5.88
0.3	0.003	0.036	0.036	0.036	5.30
0.5	0.004	0.027	0.027	0.027	5.32
0.7	0.006	0.022	0.023	0.022	5.22
0.9	0.009	0.019	0.021	0.018	5.00
1	0.011	0.017	0.020	0.017	5.10
1.1	0.013	0.016	0.020	0.016	4.90
1.3	0.017	0.014	0.022	0.014	4.74
1.5	0.023	0.012	0.026	0.012	4.66
1.7	0.028	0.011	0.030	0.011	4.42
1.9	0.033	0.010	0.034	0.010	4.10
\hat{h}	0.008	0.021	0.022	0.017	11.84

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.783	0.846	0.934	1.065	1.269

(b) $n = 2000$

	\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$					
0.1	0.004	0.049	0.049	0.049	5.26
0.3	0.002	0.027	0.028	0.027	5.16
0.5	0.003	0.021	0.021	0.021	5.34
0.7	0.005	0.017	0.018	0.017	5.36
0.9	0.007	0.014	0.016	0.014	5.20
1	0.008	0.013	0.016	0.013	5.32
1.1	0.010	0.013	0.016	0.013	5.28
1.3	0.013	0.011	0.017	0.011	5.02
1.5	0.017	0.010	0.020	0.010	4.92
1.7	0.022	0.009	0.024	0.009	4.76
1.9	0.027	0.008	0.028	0.008	4.50
\hat{h}	0.007	0.015	0.017	0.014	9.64

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.813	0.869	0.943	1.043	1.173

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 12. Simulation (Exponential). $x = 1$, $p = 3$, triangular kernel.

(a) $n = 1000$						(b) $n = 2000$					
	\hat{f}_p			SE			\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size		bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$						$h_{\text{MSE}} \times$					
0.1	0.003	0.064	0.064	0.064	5.72	0.1	0.002	0.047	0.047	0.047	5.62
0.3	0.001	0.036	0.036	0.036	5.04	0.3	0.001	0.026	0.026	0.026	5.20
0.5	0.000	0.027	0.027	0.027	5.04	0.5	0.000	0.020	0.020	0.020	5.38
0.7	-0.001	0.023	0.023	0.022	5.08	0.7	-0.001	0.017	0.017	0.016	5.16
0.9	-0.001	0.020	0.020	0.020	5.52	0.9	-0.002	0.015	0.015	0.015	5.18
1	-0.001	0.019	0.019	0.019	5.44	1	-0.002	0.014	0.014	0.014	5.20
1.1	0.000	0.018	0.018	0.018	5.32	1.1	-0.001	0.014	0.014	0.013	5.16
1.3	0.002	0.017	0.017	0.016	4.96	1.3	0.001	0.012	0.012	0.012	5.16
1.5	0.005	0.015	0.015	0.015	4.76	1.5	0.003	0.011	0.012	0.011	5.28
1.7	0.008	0.013	0.015	0.013	4.78	1.7	0.006	0.010	0.011	0.010	5.24
1.9	0.011	0.012	0.016	0.012	4.66	1.9	0.008	0.009	0.012	0.009	5.16
\hat{h}						\hat{h}					
	0.000	0.022	0.022	0.020	7.46		0.000	0.016	0.016	0.014	7.30

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}					
	0.708	0.765	0.862	1.021	1.271

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}					
	0.76	0.823	0.927	1.096	1.35

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 13. Simulation (Exponential). $x = 1.5, p = 2$, triangular kernel.

(a) $n = 1000$

	\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$					
0.1	0.003	0.048	0.049	0.048	6.04
0.3	0.002	0.027	0.027	0.027	5.30
0.5	0.002	0.020	0.020	0.020	4.76
0.7	0.004	0.017	0.017	0.017	4.34
0.9	0.006	0.014	0.016	0.015	4.30
1	0.008	0.013	0.015	0.014	4.38
1.1	0.009	0.013	0.016	0.013	4.52
1.3	0.013	0.011	0.017	0.012	4.46
1.5	0.016	0.010	0.019	0.011	4.38
1.7	0.021	0.009	0.023	0.010	4.02
1.9	0.025	0.009	0.026	0.009	4.12
\hat{h}	0.006	0.016	0.017	0.014	8.74

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.803	0.863	0.947	1.065	1.21

(b) $n = 2000$

	\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$					
0.1	0.001	0.037	0.037	0.036	6.12
0.3	0.001	0.021	0.021	0.020	5.96
0.5	0.002	0.016	0.016	0.016	5.62
0.7	0.003	0.013	0.013	0.013	5.14
0.9	0.005	0.011	0.012	0.011	4.92
1	0.006	0.011	0.012	0.011	5.04
1.1	0.007	0.010	0.012	0.010	5.00
1.3	0.010	0.009	0.013	0.009	4.94
1.5	0.013	0.008	0.015	0.008	4.90
1.7	0.016	0.008	0.018	0.008	4.54
1.9	0.020	0.007	0.021	0.007	4.60
\hat{h}	0.005	0.012	0.013	0.011	8.24

	Quantile				
	0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.83	0.88	0.952	1.046	1.157

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.

Table 14. Simulation (Exponential). $x = 1.5$, $p = 3$, triangular kernel.

(a) $n = 1000$						(b) $n = 2000$					
	\hat{f}_p			SE			\hat{f}_p			SE	
	bias	sd	$\sqrt{\text{mse}}$	mean	size		bias	sd	$\sqrt{\text{mse}}$	mean	size
$h_{\text{MSE}} \times$						$h_{\text{MSE}} \times$					
0.1	0.000	0.049	0.049	0.048	6.14	0.1	0.000	0.036	0.036	0.035	6.02
0.3	0.000	0.027	0.027	0.027	4.88	0.3	0.000	0.020	0.020	0.020	6.04
0.5	0.000	0.021	0.021	0.021	4.34	0.5	0.000	0.016	0.016	0.015	5.42
0.7	-0.001	0.018	0.018	0.018	4.58	0.7	-0.001	0.013	0.013	0.013	5.08
0.9	-0.003	0.016	0.016	0.016	4.62	0.9	-0.002	0.012	0.012	0.012	5.14
1	-0.004	0.015	0.016	0.015	4.76	1	-0.003	0.011	0.012	0.011	5.22
1.1	-0.006	0.015	0.016	0.015	4.68	1.1	-0.005	0.011	0.012	0.011	4.96
1.3	-0.007	0.014	0.016	0.014	4.78	1.3	-0.007	0.011	0.013	0.010	5.38
1.5	-0.006	0.014	0.015	0.014	4.78	1.5	-0.007	0.010	0.012	0.010	5.24
1.7	-0.004	0.013	0.014	0.013	4.86	1.7	-0.006	0.010	0.011	0.010	5.22
1.9	-0.002	0.012	0.012	0.012	4.92	1.9	-0.004	0.009	0.010	0.009	5.16
\hat{h}						\hat{h}					
	-0.003	0.015	0.016	0.014	7.32		-0.003	0.011	0.012	0.010	8.32
	Quantile						Quantile				
	0.10	0.25	0.50	0.75	0.90		0.10	0.25	0.50	0.75	0.90
\hat{h}/h_{MSE}	0.962	1.035	1.17	1.393	1.761	\hat{h}/h_{MSE}	1.015	1.091	1.225	1.454	1.91

Note. (i) bias: empirical bias of the estimators; (ii) sd: empirical standard deviation of the estimators; (iii) $\sqrt{\text{mse}}$: empirical MSE of the estimators; (iv) mean: empirical average of the estimated standard errors; (v) size: empirical size of testing the hypothesis at nominal 5% level, the test statistic is centered at $\mathbb{E}\hat{f}_p$.