SFB 823

Discussion Paper

Nonparametric estimation and testing on discontinuity of positive supported densities: A kernel truncation approach

Benedikt Funke, Masayuki Hirukawa

Nr. 42/2016

SFB 823

Nonparametric Estimation and Testing on Discontinuity of Positive Supported Densities: A Kernel Truncation Approach*

Benedikt Funke† Technical University of Dortmund

Masayuki Hirukawa‡ Setsunan University

July 2016

Abstract

Discontinuity in density functions is of economic importance and interest. For instance, in studies on regression discontinuity designs, discontinuity in the density of a running variable suggests violation of the no-manipulation assumption. In this paper we develop estimation and testing procedures on discontinuity in densities with positive support. Our approach is built on splitting the gamma kernel (Chen, 2000) into two parts at a given (dis)continuity point and constructing two truncated kernels. The jump-size magnitude of the density at the point can be estimated nonparametrically by two kernels and a multiplicative bias correction method. The estimator is easy to implement, and its convergence properties are delivered by various approximation techniques on incomplete gamma functions. Based on the jump-size estimator, two versions of test statistics for the null of continuity at a given point are also proposed. Moreover, estimation theory of the entire density in the presence of a discontinuity point is explored. Monte Carlo simulations confirm nice finite-sample properties of the jump-size estimator and the test statistics.

Keywords: boundary bias; density estimation; discontinuous probability density; gamma kernel; incomplete gamma functions; nonparametric kernel testing; regression discontinuity design.

JEL Classification Codes: C12; C13; C14.

MSC 2010 Codes: 62G07; 62G10; 62G20.

*The first author gratefully acknowledges financial support from the German Science Foundation (DFG) through the Collaborative Research Center SFB 823. The second author gratefully acknowledges financial support from Japan Society of the Promotion of Science (grant number 15K03405).

†Department of Mathematics, Technical University of Dortmund, Vogelpothsweg 87, 44227 Dortmund, Germany; e-mail: benedikt.funke@mathematik.tu-dortmund.de.

‡Faculty of Economics, Setsunan University, 17-8 Ikeda Nakamachi, Neyagawa, Osaka, 572-8508, Japan; e-mail: hirukawa@econ.setsunan.ac.jp.

1 Introduction

The objective of this paper is to develop new estimation and testing procedures of discontinuity in density functions with support on $\mathbb{R}_+$. Inference on possibly discontinuous densities has been explored in nonparametric statistics: examples include Liebscher (1990), Cline and Hart (1991), and Chu and Cheng (1996), to name a few. Discontinuity in densities is also of economic importance and interest. Local randomization of a continuous running variable is a key requirement for the validity of regression discontinuity designs ("RDD"); if the value of the running variable falls into the left and right of the cutoff strategically, then treatment effects are no longer point identified due to self-selection. Therefore, detection of discontinuity in the density of the running variable at the cutoff suggests evidence of such strategic behavior or manipulation in RDD. Nonetheless, estimation and inference on jump-size magnitudes of densities at discontinuity points have not attracted interest in econometrics up until recently. McCrary (2008) applies a bin-based local linear regression method to estimate jump sizes. Subsequently, Otsu, Xu and Matsushita (2013) propose two versions of empirical likelihood-based inference procedures grounded on binning and local likelihood methods. While our proposal can be viewed as an extension of these articles, it has a unique feature. In our approach, jump sizes are estimated by means of density estimation techniques using the kernels obtained through truncating asymmetric kernels at a given (dis)continuity point, unlike nonparametric regression or local likelihood approaches using standard symmetric kernels.

Before proceeding, it is worth explaining why we specialize in asymmetric kernel smoothing. Empirical studies on discontinuity in densities frequently pay attention to the distributions of economic variables such as (taxable or relative) incomes (Saez, 2010; Bertrand, Kamenica and Pan, 2015), wages (DiNardo, Fortin and Lemieux, 1996), school enrollment counts (Angrist and Lavy, 1999) and proportion of votes for proposed bills (McCrary, 2008). The distributions, if they are free of discontinuity points, can be empirically characterized by two stylized facts, namely, (i) existence of a lower bound in support (most possibly at the origin) and (ii) concentration of observations near the boundary and a long tail with sparse data. When estimating such densities nonparametrically using symmetric kernels, we must rely either on a boundary correction method and an adaptive smoothing technique (e.g., variable bandwidth methods) *simultaneously*, or on back-transforming the density estimator from the log-transformed data to the original scale. The former is apparently cumbersome, and density estimates by the latter often behave poorly (e.g., Cowell, Ferreira and Litchfield, 1998) although the method is popularly applied in empirical works. Asymmetric kernels with support on $\mathbb{R}_+$ have emerged as a viable alternative that can accommodate the stylized facts. Although there are various classes of asymmetric kernels, for the sake of simplicity and due to popularity this study focuses exclusively on the gamma kernel by Chen (2000)

$$K_{G(x,b)}(u) = \frac{u^{x/b}\exp(-u/b)}{b^{x/b+1}\Gamma(x/b+1)}\mathbf{1}(u \geq 0),$$

where $x\,(\geq 0)$ and $b\,(> 0)$ are the design point and smoothing parameter, respectively.

When the density has a discontinuity point, the jump-size magnitude at the point can be defined as the difference between left and right limits of the density at the point. While nonparametric regression (McCrary, 2008) and empirical likelihood (Otsu, Xu and Matsushita, 2013) methods have been applied to estimate the jump size, we attempt to have our jump-size estimator preserve appealing properties of the gamma kernel. Accordingly, we split the gamma kernel into two parts at the discontinuity point, and make each part a legitimate kernel by re-normalization. The left and right limits of the density can be estimated by two truncated kernels. Although the estimators are consistent and their variance convergences are usual $O\left(n^{-1}b^{-1/2}\right)$ where n is the sample size, their bias convergences are $O\left(b^{1/2}\right)$, not the usual $O\left(b\right)$ rate. Then, we apply the multiplicative bias correction technique by Terrell and Scott (1980) to eliminate the undesirable $O\left(b^{1/2}\right)$ biases without inflating the order of magnitude in variance. Moreover, we take particular care of choosing the smoothing parameter. Specifically, the method of power-optimality smoothing parameter selection by Kulasekera and Wang (1998) is tailored to inference problems on discontinuous densities.

Our proposal has three contributions to the literature. First, unlike the binned local linear ("BLL") estimation by McCrary (2008), our kernel truncation approach always generates nonnegative density estimates and is free from choosing bin widths. Our jump-size estimator is also easy to implement. Since it has a closed form, nonlinear optimization as in Otsu, Xu and Matsushita (2013) is unnecessary. While incomplete gamma functions are key ingredients in our estimator, standard statistical packages including GAUSS, Matlab and R prepare a command that can return values of the functions either directly or in the form of gamma cumulative distribution functions.

Second, in delivering convergence results of asymmetric kernel estimators, we utilize the mathematical tools and proof strategies that are totally different from those for nonparametric estimators smoothed by symmetric kernels. Asymptotic results throughout this paper are built upon a few different approximation techniques on incomplete gamma functions; such proof strategies are taken for the first time in the econometric literature, to the best of our knowledge.

Third, we also present estimation theory of the entire density in the presence of a discontinuity point. Indeed, Imbens and Lemieux (2008) argue importance of graphical analyses in empirical studies on RDD, including inspections of densities of running variables. It is demonstrated that density estimators smoothed by the truncated gamma kernels admit the same bias and variance approximations as the gamma kernel density estimator does. Furthermore, the truncated gamma-kernel density estimator is shown to be consistent even when the true density is unbounded at the origin.

The remainder of this paper is organized as follows. Section 2 presents estimation and testing procedures of the density at a known discontinuity point $c\,(>0)$. As an important practical problem, a smoothing parameter selection method is also developed. Our particular focus is on the choice method for power optimality. In Section 3, we discuss how to estimate the entire density when the density has a discontinuity point. Convergence properties of density estimates are also explored. Section 4 conducts Monte Carlo simulations to evaluate finite-sample properties of the proposed jump-size estimator and test statistic. An empirical application on the validity of RDD is presented in Section 5. Section 6 summarizes the main results of the paper. Proofs are provided in the Appendix.

This paper adopts the following notational conventions: for $a > 0$, $\Gamma(a) = \int_0^\infty t^{a-1}\exp(-t)\,dt$ is the gamma function; for $a, z > 0$, $\gamma(a,z) = \int_0^z t^{a-1}\exp(-t)\,dt$ and $\Gamma(a,z) = \int_z^\infty t^{a-1}\exp(-t)\,dt = \Gamma(a) - \gamma(a,z)$ denote the lower and upper incomplete gamma functions, respectively; $\mathbf{1}\{\cdot\}$ signifies an indicator function; and $\lfloor\cdot\rfloor$ denotes the integer part. Lastly, the expression '$X_n \sim Y_n$' is used whenever $X_n/Y_n \to 1$ as $n \to \infty$.

2 Estimation and Inference for Discontinuity in the Density

2.1 Setup

Suppose that we suspect discontinuity of the probability density function ("pdf") $f(x)$ at a given point $x = c\,(> 0)$, which is assumed to be interior throughout. Also let

$$f_{-}(c) := \lim_{x \uparrow c} f(x) \text{ and } f_{+}(c) := \lim_{x \downarrow c} f(x),$$

be the lower and upper limits of the pdf at $x = c$, respectively. Our parameter of interest is the jump-size magnitude of the density at c

$$J(c) := f_{+}(c) - f_{-}(c).$$

To check whether f is (dis)continuous at c, we first estimate $J(c)$ nonparametrically and then proceed to a hypothesis testing for the null of continuity of f at c, i.e., $H_0 : J(c) = 0$, against the two-sided alternative.

2.2 An Issue in Estimating Two Limits of the Density

To develop a consistent estimator of $J(c)$, we start our analysis from estimating two limits of the density at c. Let $\{X_i\}_{i=1}^{n}$ be a univariate random sample drawn from a distribution that has the pdf f. When f is indeed discontinuous at c, a reasonable method would be to estimate $f_{-}(c)$ and $f_{+}(c)$ using sub-samples $\{X_i^{-}\} := \{X_i : X_i < c\}$ and $\{X_i^{+}\} := \{X_i : X_i \geq c\}$, respectively. Instead of relying on nonparametric regression or local likelihood methods, we split the gamma kernel into two parts at c, namely,

$$K_{G(x,b)}(u) := K_{G(x,b;c)}^{L}(u) + K_{G(x,b;c)}^{U}(u),$$

where

$$
\begin{aligned}
K_{G(x,b;c)}^{L}(u) &= \frac{u^{x/b}\exp(-u/b)}{b^{x/b+1}\Gamma(x/b+1)}\mathbf{1}(0 \le u < c) \text{ and} \\
K_{G(x,b;c)}^{U}(u) &= \frac{u^{x/b}\exp(-u/b)}{b^{x/b+1}\Gamma(x/b+1)}\mathbf{1}(u \ge c).
\end{aligned}
$$

However, neither $K_{G(x,b;c)}^{L}(u)$ nor $K_{G(x,b;c)}^{U}(u)$ is a legitimate kernel function in the sense that

$$
\begin{aligned}
\int_0^\infty K_{G(x,b;c)}^{L}(u)\,du &= \frac{\gamma(x/b+1,c/b)}{\Gamma(x/b+1)} \text{ and} \\
\int_0^\infty K_{G(x,b;c)}^{U}(u)\,du &= \frac{\Gamma(x/b+1,c/b)}{\Gamma(x/b+1)}.
\end{aligned}
$$

Therefore, we make scale-adjustments to obtain the re-normalized truncated kernels as

$$
\begin{aligned}
K_{G(x,b;c)}^{-}(u) &= \frac{\Gamma(x/b+1)}{\gamma(x/b+1,c/b)}K_{G(x,b;c)}^{L}(u) \text{ and} \\
K_{G(x,b;c)}^{+}(u) &= \frac{\Gamma(x/b+1)}{\Gamma(x/b+1,c/b)}K_{G(x,b;c)}^{U}(u).
\end{aligned}
$$

These kernels yield estimators of $f_{-}(c)$ and $f_{+}(c)$ as

$$
\begin{aligned}
\hat{f}_{-}(c) &= \frac{1}{n}\sum_{i=1}^{n} K_{G(x,b;c)}^{-}(X_i)\Big|_{x=c} = \frac{1}{n}\sum_{i=1}^{n} K_{G(c,b;c)}^{-}(X_i) \text{ and} \\
\hat{f}_{+}(c) &= \frac{1}{n}\sum_{i=1}^{n} K_{G(x,b;c)}^{+}(X_i)\Big|_{x=c} = \frac{1}{n}\sum_{i=1}^{n} K_{G(c,b;c)}^{+}(X_i).
\end{aligned}
$$

To explore asymptotic properties of these estimators, we make the following assumptions. For notational conciseness, expressions such as "$f_{\pm}(c)$" are used throughout, whenever no confusions may occur.

Assumption 1. The random sample $\{X_i\}_{i=1}^{n}$ is drawn from a univariate distribution with a pdf f having support on $\mathbb{R}_{+}$.

Assumption 2. The second-order derivative of the pdf f is Hölder-continuous of order $\varsigma \in (0,1]$ on $\mathbb{R}_+ \backslash \{c\}$. Also let $f_-^{(j)}(c) := \lim_{x \uparrow c} d^j f(x)/dx^j$ and $f_+^{(j)}(c) := \lim_{x \downarrow c} d^j f(x)/dx^j$ for $j = 1, 2$. Then, $f_\pm(c) > 0$ and $\left| f_\pm^{(2)}(c) \right| < \infty$.

Assumption 3. The smoothing parameter $b\,(= b_n > 0)$ satisfies $b + (nb)^{-1} \to 0$ as $n \to \infty$.

Assumptions 1 and 3 are standard in the literature on asymmetric kernel smoothing (e.g., Chen, 2000; Hirukawa and Sakudo, 2015). The condition "$(nb)^{-1} \to 0$" in Assumption 3 is required for the estimation of the entire density that will be discussed in Section 3, whereas a weaker condition "$\left(nb^{1/2}\right)^{-1} \to 0$" suffices for Propositions 1 and 2 and Theorem 1 below. Moreover, an equivalent to Assumption 2 can be found in McCrary (2008) and Otsu, Xu and Matsushita (2013). In particular, Hölder-continuity of the second-order density derivative $f^{(2)}(\cdot)$ in Assumption 2 implies that there is a constant $L \in (0, \infty)$ such that

$$\begin{aligned} \left| f^{(2)}(s) - f^{(2)}(t) \right| &\leq L |s-t|^{\varsigma}, \ \forall s, t \in [0, c) \text{ and} \\ \left| f^{(2)}(s') - f^{(2)}(t') \right| &\leq L |s'-t'|^{\varsigma}, \ \forall s', t' \in [c, \infty). \end{aligned}$$

The proposition below refers to bias and variance approximations of $\hat{f}_\pm(c)$. It is worth emphasizing that all convergences results in this paper are built upon a few different approximation techniques on incomplete gamma functions; such proof strategies are taken for the first time in the econometric literature, to the best of our knowledge. Moreover, for the purpose of our subsequent analysis, the bias expansion is derived up to the second-order term.

Proposition 1. *Under Assumptions 1-3, as $n \to \infty$,*

$$\begin{aligned} Bias\left\{\hat{f}_{\pm}(c)\right\} &\sim \mp\sqrt{\frac{2}{\pi}}c^{1/2}f_{\pm}^{(1)}(c)\,b^{1/2} + \left\{\left(1-\frac{4}{3\pi}\right)f_{\pm}^{(1)}(c) + \frac{c}{2}f_{\pm}^{(2)}(c)\right\}b, \ \textit{and} \\ Var\left\{\hat{f}_{\pm}(c)\right\} &\sim \frac{1}{nb^{1/2}}\frac{f_{\pm}(c)}{\sqrt{\pi}c^{1/2}}. \end{aligned}$$

Proposition 1 implies that $\hat{f}_{\pm}(c)$ are consistent for $f_{\pm}(c)$, and that their variance convergence has a usual rate of $O\left(n^{-1}b^{-1/2}\right)$. Nevertheless, the bias convergence is $O\left(b^{1/2}\right)$, which is slower than the usual $O(b)$ rate. This is an outcome of one-sided smoothing. If f were continuous at c and smoothing were made on both sides of the design point c using the gamma kernel, the nearly symmetric shape of the kernel would cancel out the $O\left(b^{1/2}\right)$ bias.[1] In reality, because data points used for estimating $f_{\pm}(c)$ lie only on either the left or right side of c, the $O\left(b^{1/2}\right)$ bias never vanishes. It follows that when $J(c)$ is estimated by $\hat{J}(c) := \hat{f}_{+}(c) - \hat{f}_{-}(c)$, it also has an inferior $O\left(b^{1/2}\right)$ bias. Therefore, our goal is to propose an estimator of $J(c)$ with an $O(b)$ bias and an $O\left(n^{-1}b^{-1/2}\right)$ variance.

2.3 Bias-Corrected Estimation and Inference

To improve the bias convergence in estimators of $f_{\pm}(c)$ from $O\left(b^{1/2}\right)$ to $O(b)$ while the order of magnitude in variance remains unchanged, we propose to employ a multiplicative bias correction ("MBC") technique. As in Hirukawa (2010), Hirukawa and Sakudo (2014, 2015), and Funke and Kawka (2015), the MBC method proposed by Terrell and Scott (1980) is adopted.[2] The method eliminates the leading bias term by constructing a multiplicative combination of two density estimators with differ-

[1] This can be also seen by combining two estimators $\hat{f}_{\pm}(c)$ as a weighted sum.

[2] Aforementioned articles also apply another MBC method proposed by Jones, Linton and Nielsen (1995). However, it appears that the method fails to eliminate the $O\left(b^{1/2}\right)$ bias. Their MBC estimator of $f_{-}(c)$, for example, can be written as

$$\check{f}_{-}(c) := \hat{f}_{-}(c)\,\check{\alpha}_{-}(c) := \hat{f}_{-}(c)\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{K_{G(c,b;c)}^{-}(X_i)}{\hat{f}_{-}(X_i)}\right\},$$

ent smoothing parameters. In our context, for some constant $\delta \in (0,1)$, the MBC estimators of $f_{\pm}(c)$ can be defined as

$$\tilde{f}_{\pm}(c) = \left\{\hat{f}_{\pm,b}(c)\right\}^{1/\left(1-\delta^{1/2}\right)} \left\{\hat{f}_{\pm,b/\delta}(c)\right\}^{-\delta^{1/2}/\left(1-\delta^{1/2}\right)},$$

where $\hat{f}_{\bullet,b}(x)$ and $\hat{f}_{\bullet,b/\delta}(x)$ signify the density estimators using smoothing parameters b and b/δ, respectively. Not only are $\tilde{f}_{\pm}(c)$ nonnegative by construction, but also their bias and variance convergences are usual $O(b)$ and $O\left(n^{-1}b^{-1/2}\right)$ rates, respectively, as documented in the next proposition. The proof is similar to the one for Theorem 1 of Hirukawa and Sakudo (2014), and thus it is omitted.

Proposition 2. *Under Assumptions 1-3, as $n \to \infty$,*

$$Bias\left\{\tilde{f}_{\pm}(c)\right\} \sim \left(\frac{1}{\delta^{1/2}}\right)\left[\frac{c}{\pi}\left\{\frac{\left(f_{\pm}^{(1)}(c)\right)^2}{f_{\pm}(c)}\right\} - \left\{\left(1-\frac{4}{3\pi}\right)f_{\pm}^{(1)}(c) + \frac{c}{2}f_{\pm}^{(2)}(c)\right\}\right] b, \text{ and}$$

$$Var\left\{\tilde{f}_{\pm}(c)\right\} \sim \frac{1}{nb^{1/2}}\lambda(\delta)\frac{f_{\pm}(c)}{\sqrt{\pi}c^{1/2}},$$

where

$$\lambda(\delta) := \frac{\left(1+\delta^{3/2}\right)(1+\delta)^{1/2} - 2\sqrt{2}\delta}{(1+\delta)^{1/2}\left(1-\delta^{1/2}\right)^2}$$

is monotonously increasing in $\delta \in (0,1)$ with

$$\lim_{\delta\downarrow 0}\lambda(\delta) = 1 \text{ and } \lim_{\delta\uparrow 1}\lambda(\delta) = \frac{11}{4}.$$

Proposition 2 suggests that as $\delta \downarrow 0$ ($\delta \uparrow 1$) or in case of oversmoothing (undersmoothing), the bias increases (decreases) and the variance decreases (increases). It is a common practice in nonparametric kernel testing that the bias is made asymptotically negligible via undersmoothing, and thus what matters for inference is the

where $\breve{\alpha}_{-}(c)$ serves as the 'bias correction' term. However, $\hat{f}_{-}(x)$ $(x < c)$ has an $O(b)$ bias, as stated in Theorem 2, so does $\breve{\alpha}_{-}(c)$. Therefore, the $O\left(b^{1/2}\right)$ bias in $\hat{f}_{-}(c)$ never vanishes, and thus we do not pursue this type of MBC.

size of $\lambda(\delta)$. Because of no minimum in $\lambda(\delta)$, the choice of δ is left as an exercise in Monte Carlo simulations.

It also follows that $J(c)$ can be consistently estimated as $\tilde{J}(c) := \tilde{f}_{+}(c) - \tilde{f}_{-}(c)$. The next theorem refers to the limiting distribution of $\tilde{J}(c)$.[3]

Theorem 1. *Under Assumptions 1-3, as $n \to \infty$,*

$$\sqrt{nb^{1/2}}\left\{\tilde{J}(c) - J(c) - B(c)\,b + o(b)\right\} \xrightarrow{d} N(0, V(c)), \tag{1}$$

where

$$\begin{aligned} B(c) &= \left(\frac{1}{\delta^{1/2}}\right)\left[\frac{c}{\pi}\left\{\frac{\left(f_{+}^{(1)}(c)\right)^{2}}{f_{+}(c)} - \frac{\left(f_{-}^{(1)}(c)\right)^{2}}{f_{-}(c)}\right\}\right. \\ &\quad \left.-\left\{\left(1-\frac{4}{3\pi}\right)\left(f_{+}^{(1)}(c) - f_{-}^{(1)}(c)\right) + \frac{c}{2}\left(f_{+}^{(2)}(c) - f_{-}^{(2)}(c)\right)\right\}\right] \\ V(c) &= \lambda(\delta)\left\{\frac{f_{+}(c) + f_{-}(c)}{\sqrt{\pi}c^{1/2}}\right\}, \end{aligned}$$

and $\lambda(\delta)$ is defined in Proposition 2. In addition, if $nb^{5/2} \to 0$ as $n \to \infty$, then (1) reduces to

$$\sqrt{nb^{1/2}}\left\{\tilde{J}(c) - J(c)\right\} \xrightarrow{d} N(0, V(c)).$$

As indicated in Proposition 2, $\tilde{J}(c)$ has an $O(b)$ bias and an $O\left(n^{-1}b^{-1/2}\right)$ variance. Observe that for a given δ, the variance coefficient decreases as c increases, i.e., as the discontinuity point moves away from the origin. We can also find that the leading bias term $B(c)\,b$ cancels out if f has a continuous second-order derivative at c.

Theorem 1 also implies that given a smoothing parameter $b = Bn^{-q}$ for some constants $B \in (0, \infty)$ and $q \in (2/5, 1)$ and $\tilde{V}(c)$, a consistent estimate of $V(c)$, the

[3]It is possible to use different constants δ_{-} and δ_{+} and/or different smoothing parameters b_{-} and b_{+} for $\tilde{f}_{-}(c)$ and $\tilde{f}_{+}(c)$, as long as b_{-} and b_{+} shrink to zero at the same rate. For convenience, however, we choose to employ the same δ and b.

test statistic is

$$T(c) := \frac{\sqrt{nb^{1/2}}\tilde{J}(c)}{\sqrt{\tilde{V}(c)}} \xrightarrow{d} N(0,1) \text{ under } H_0 : J(c) = 0.$$

Moreover, as documented in the next proposition, the test is consistent. Observe that the power approaches one for local alternatives with convergence rates no faster than $n^{1/2}b^{1/4}$, as well as for fixed alternatives.

Proposition 3. *Under Assumptions 1-3, as* $n \to \infty$, $\Pr\{|T(c)| > B_n\} \to 1$ *under* $H_1 : J(c) \neq 0$ *for any non-stochastic sequence* B_n *satisfying* $B_n = o\left(n^{1/2}b^{1/4}\right)$.

Our remaining tasks are to present examples of $\tilde{V}(c)$ and to propose a choice method of b. The latter is discussed in the next section, whereas there are a few candidates of $\tilde{V}(c)$. Replacing $f_{\pm}(c)$ in $V(c)$ with their consistent estimates $\tilde{f}_{\pm}(c)$ immediately yields

$$\tilde{V}_1(c) := \lambda(\delta)\left\{\frac{\tilde{f}_+(c) + \tilde{f}_-(c)}{\sqrt{\pi}c^{1/2}}\right\}.$$

Alternatively, it is possible to compute the gamma kernel density estimator at c

$$\hat{f}(c) := \frac{1}{n}\sum_{i=1}^{n} K_{G(x,b)}(X_i)\Big|_{x=c} = \frac{1}{n}\sum_{i=1}^{n} K_{G(c,b)}(X_i).$$

By (A4) and (A6), we have

$$\frac{\gamma(c/b+1, c/b)}{\Gamma(c/b+1)} = \frac{1}{2} + O\left(b^{1/2}\right) \text{ and } \frac{\Gamma(c/b+1, c/b)}{\Gamma(c/b+1)} = \frac{1}{2} + O\left(b^{1/2}\right).$$

It follows that

$$\hat{f}(c) = \frac{\gamma(c/b+1, c/b)}{\Gamma(c/b+1)}\hat{f}_-(c) + \frac{\Gamma(c/b+1, c/b)}{\Gamma(c/b+1)}\hat{f}_+(c) \xrightarrow{p} \frac{f_+(c) + f_-(c)}{2}.$$

As a consequence, we can obtain another estimator of $V(c)$ as

$$\tilde{V}_2(c) := \lambda(\delta)\left\{\frac{2\hat{f}(c)}{\sqrt{\pi}c^{1/2}}\right\}.$$

2.4 Smoothing Parameter Selection

How to choose the value of the smoothing parameter b is an important practical problem. McCrary (2008) proposes the choice method which closely follows the literature on the BLL smoothing. Moreover, in the literature on RDD, Imbens and Kalyanaraman (2012) and Porter and Yu (2015, Section 5.4) discuss methods of choosing the smoothing parameter. All these proposals rely on either a cross-validation criterion or a plug-in approach, and thus they stand on the idea of estimation-optimality. However, once our priority is given to testing for continuity of the pdf f at a given point c, such approaches cannot be justified in theory or practice, because estimation-optimal values may not be equally optimal for testing purposes. Here we have a preference for test-optimality and thus adopt the power-optimality criterion by Kulasekera and Wang (1998), whose idea is also applied in Hirukawa and Sakudo (2016).

Below Procedure 1 of Kulasekera and Wang (1998) is tailored to our context. The procedure is a version of sub-sampling. Let n_- and n_+ be the numbers of observations in sub-samples $\{X_i^-\}$ and $\{X_i^+\}$, respectively, where $n \equiv n_- + n_+$. Also assume that $\{X_i^-\}_{i=1}^{n_-}$ and $\{X_i^+\}_{i=1}^{n_+}$ are ordered samples. Then, the entire sample $\{X_i\}_{i=1}^{n} = \left\{\{X_i^-\}_{i=1}^{n_-}, \{X_i^+\}_{i=1}^{n_+}\right\}$ can be split into M sub-samples, where $M = M_n$ is a non-stochastic sequence that satisfies $1/M + M/n \to 0$ as $n \to \infty$. Given such M, $(k_-, k_+) := (\lfloor n_-/M \rfloor, \lfloor n_+/M \rfloor)$ and $k := k_- + k_+$, the mth sub-sample is defined as

$$\{X_{m,i}\}_{i=1}^{k} := \left\{\left\{X_{m+(i-1)M}^-\right\}_{i=1}^{k_-}, \left\{X_{m+(i-1)M}^+\right\}_{i=1}^{k_+}\right\}, \ m = 1, \ldots, M.$$

The test statistic using the mth sub-sample $\{X_{m,i}\}_{i=1}^{k}$ becomes

$$T_m(c) := \frac{\sqrt{kb^{1/2}}\tilde{J}_m(c)}{\sqrt{\tilde{V}_m(c)}}, \ m = 1, \ldots, M,$$

where $\tilde{J}_m(c)$ and $\tilde{V}_m(c)$ (which is either $\tilde{V}_{1,m}(c)$ or $\tilde{V}_{2,m}(c)$) are the sub-sample analogues of $\tilde{J}(c)$ and $\tilde{V}(c)$, respectively. Also denote the set of admissible values for

$b = b_n$ as $H_n := \left[\underline{B}n^{-q}, \overline{B}n^{-q}\right]$ for some prespecified exponent $q \in (2/5, 1)$ and two constants $0 < \underline{B} < \overline{B} < \infty$. Moreover, let

$$\hat{\pi}_M(b_k) := \frac{1}{M}\sum_{m=1}^{M} \mathbf{1}\left\{T_m(c) > c_m(\alpha)\right\},$$

where $c_m(\alpha)$ is the critical value for the size α test using the mth sub-sample. We pick the power-maximized $\hat{b}_k = \hat{B}k^{-q} = \arg\max_{b_k \in H_k} \hat{\pi}_M(b_k)$, and the smoothing parameter value $\hat{b}_n := \hat{B}n^{-q}$ follows.

We conclude this section by stating how to obtain $\hat{b}_n$ in practice. Step 1 reflects that M should be divergent but smaller than both n_- and n_+ in finite samples. Step 3 follows from the implementation methods in Kulasekera and Wang (1998). Finally, Step 4 corresponds to the case for more than one maximizer of $\hat{\pi}_M(b_k)$.

Step 1: Choose some $p \in (0, 1)$ and specify $M = \lfloor \min\{n_-^p, n_+^p\} \rfloor$.
Step 2: Make M sub-samples of sizes $(k_-, k_+) = (\lfloor n_-/M \rfloor, \lfloor n_+/M \rfloor)$.
Step 3: Pick two constants $0 < \underline{H} < \overline{H} < 1$ and define $H_k = \left[\underline{H}, \overline{H}\right]$.
Step 4: Set $c_m(\alpha) \equiv z_\alpha$ and find $\hat{b}_k = \inf\{\arg\max_{b_k \in H_k} \hat{\pi}_M(b_k)\}$ by a grid search.
Step 5: Recover $\hat{B}$ by $\hat{B} = \hat{b}_k k^q$ and calculate $\hat{b}_n = \hat{B}n^{-q}$.

3 Estimation of the Entire Density in the Presence of a Discontinuity Point

3.1 Density Estimation by Truncated Kernels

We are typically interested in how the shape of the pdf looks like, as well as whether it has a discontinuity point. Imbens and Lemieux (2008) strongly recommend graphical analyses in empirical studies on RDD, including inspections of densities of running variables. If the test in the previous section fails to reject the null of continuity of the pdf f at the cutoff c, the entire density may be re-estimated by the gamma kernel, for example. How should we estimate the entire density if the test rejects the null?

The answer to this question is simple. It suffices to compute $\hat{f}_{-}(x)$ or $\hat{f}_{+}(x)$ as an estimate of $f(x)$, depending on the position of the design point x. To put it in another way, $\hat{f}_{-}(x)$ ($\hat{f}_{+}(x)$) can be employed whenever $x < c$ ($x > c$), provided that c is the only point of discontinuity in f, as documented in the theorem below. Although only the bias-variance trade-off is provided there, asymptotic normality of the estimators can be established similarly to Theorem 1.

Theorem 2. *Suppose that Assumptions 1-3 hold. Then, for $x > c$, as $n \to \infty$,*

$$\begin{aligned} Bias\left\{\hat{f}_{+}(x)\right\} &\sim \left\{f^{(1)}(x) + \frac{x}{2}f^{(2)}(x)\right\}b, \text{ and} \\ Var\left\{\hat{f}_{+}(x)\right\} &\sim \frac{1}{nb^{1/2}}\frac{f(x)}{2\sqrt{\pi}x^{1/2}}. \end{aligned}$$

On the other hand, for $x < c$, as $n \to \infty$,

$$\begin{aligned} Bias\left\{\hat{f}_{-}(x)\right\} &\sim \left\{f^{(1)}(x) + \frac{x}{2}f^{(2)}\right\}b, \text{ and} \\ Var\left\{\hat{f}_{-}(x)\right\} &\sim \begin{cases} \frac{1}{nb^{1/2}}\frac{f(x)}{2\sqrt{\pi}x^{1/2}} & \text{if } x/b \to \infty \\ \frac{1}{nb}\frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^{2}(\kappa+1)}f(x) & \text{if } x/b \to \kappa \in (0,\infty) \end{cases}. \end{aligned}$$

Theorem 2 indicates no adversity when $f(x)$ for $x \neq c$ is estimated by $\hat{f}_{\pm}(x)$. Observe that $\hat{f}_{\pm}(x)$ admit the same bias and variance expansions as the gamma kernel density estimator $\hat{f}(x)$ does. A rationale is that as the design point x moves away from the truncation point c, data points tend to lie on both sides of x and each truncated kernel is likely to behave like the gamma kernel. We can also see that the variance coefficient decreases as x increases. The shrinking variance coefficient as the design point x moves away from the origin reflects that more data points can be pooled to smooth in areas with fewer observations. This property is particularly advantageous to estimating the distributions that have a long tail with sparse data, such as those of the economic variables mentioned in Section 1.

3.2 Convergence Properties of $\hat{f}_-(x)$ When the Density Is Unbounded at the Origin

Clusterings of observations near the boundary are frequently observed in the distributions with positive supports. In the study of RDD, Figure 1 of Bertrand, Kamenica and Pan (2015) suggests that the distribution of wives' relative income within households has a clustering of observations near the origin, as well as a sharp drop at the point of 1/2 (i.e., the point at which wives' income shares exceed their husbands'). Similarly, in Figure 5 of McCrary (2008), the distribution of proportion of votes for proposed bills in the US House of Representatives appears to be unbounded at the boundary of 100%, as well as a sharp discontinuity at the point of 50%.[4]

The following two theorems document weak consistency and the relative convergence of $\hat{f}_-(x)$ when $f(x)$ is unbounded at $x = 0$.

Theorem 3. *If $f(x)$ is unbounded at $x = 0$, Assumptions 1 holds and $b + (nb^2)^{-1} \to 0$ as $n \to \infty$, then $\hat{f}_-(0) \xrightarrow{p} \infty$.*

Theorem 4. *Suppose that $f(x)$ is unbounded at $x = 0$ and continuously differentiable in the neighborhood of the origin. In addition, if Assumption 1 holds and $b + \{nb^2 f(x)\}^{-1} \to 0$ as $n \to \infty$ and $x \to 0$, then*

$$\left| \frac{\hat{f}_-(x) - f(x)}{f(x)} \right| \xrightarrow{p} 0$$

as $x \to 0$.

It has been demonstrated by Bouezmarni and Scaillet (2005) and Hirukawa and Sakudo (2015) that the weak consistency and relative convergence for densities unbounded at the origin are peculiar to the density estimators smoothed by the gamma

[4] The arguments in this section are still valid for this case, if we transform the original data X to $X' := 1 - X$ and apply them to the transformed data X'.

and generalized gamma kernels. The theorems ensure that $\hat{f}_{-}(x)$ is also a proper estimate for unbounded densities. We can deduce from Theorems 2-4 that all in all, appealing properties of the gamma kernel density estimator are inherited to $\hat{f}_{\pm}(x)$.

4 Finite-Sample Performance

It is widely recognized that asymptotic results on kernel-smoothed tests are not well transmitted to their finite-sample distributions, which reflects that omitted terms in the first-order asymptotics on the test statistics are highly sensitive to their smoothing parameter values in finite samples. On the other hand, there is growing literature that reports nice finite-sample properties of the estimators and test statistics smoothed by asymmetric kernels. Examples include Kristensen (2010) and Gospodinov and Hirukawa (2012) for estimation and Fernandes and Grammig (2005), Fernandes, Mendes and Scaillet (2015), and Hirukawa and Sakudo (2016) for testing. To see which perspective dominates, this section investigates finite-sample performance of the estimator of the jump-size magnitude and the test statistic for discontinuity of the density via Monte Carlo simulations.

4.1 Jump-Size Estimation

First, we focus on the estimator of the jump-size magnitude $J(c)$. As true densities, those of the following two asymmetric distributions are considered:

1. Gamma: $f(x) = x^{\alpha-1}\exp(-x/\beta)\mathbf{1}(x \geq 0)/\{\beta^{\alpha}\Gamma(\alpha)\}$, $(\alpha, \beta) = (2.75, 1)$.
2. Weibull: $f(x) = (\alpha/\beta)(x/\beta)^{\alpha-1}\exp\{-(x/\beta)^{\alpha}\}\mathbf{1}(x \geq 0)$, $(\alpha, \beta) = (1.75, 3.5)$.

Shapes of these densities can be found in Figure 1. For each distribution we choose two suspected discontinuity points c, namely, 30% quantile ("30%") and median ("Med"); see Table 1 for exact values of the points. Because the gamma and Weibull densities have modes at 1.7500 and 2.1567, respectively, the two points for each density are located on the left- and right-hand sides of the mode. The sample size is $n \in \{500, 1000, 2000\}$, and $1,000$ replications are drawn for each combination of the sample size n and the distribution.

The simulation study compares finite-sample performance of our jump-size estimator $\tilde{J}(c)$ with McCrary's (2008) BLL estimator $\hat{J}_M(c)$. The latter employs the triangular kernel $K(u) = (1 - |u|)\mathbf{1}(|u| \leq 1)$, and the bandwidth is chosen by the method described on p.705 of McCrary (2008). For the former, the smoothing parameter b is selected by the power-optimality criterion for two test statistics $T_i(c) := \sqrt{nb^{1/2}}\tilde{J}(c)/\sqrt{\tilde{V}_i(c)}$ for $i = 1, 2$, where the definition of $\tilde{V}_i(c)$ is given in Section 2.3. Implementation details are as follows: (i) all critical values in $\hat{\pi}_M(b_k)$ are set equal to $z_{0.025} = 1.96$; (ii) (p, q) are predetermined by $(p, q) = (1/2, 4/9)$; (iii) the interval for b_k is $H_k = [0.05, 0.50]$; and (iv) three different values of the mixing exponent δ are considered, namely, $\delta \in \{0.49, 0.64, 0.81\}$, so that the exponents on $\hat{f}_{\pm,b}(c)$ and $\hat{f}_{\pm,b/\delta}(c)$ to generate $\tilde{f}_{\pm}(c)$ are $(10/3, -7/3)$, $(5, -4)$ and $(10, -9)$, respectively.

FIGURE 1 AND TABLE 1 ABOUT HERE

Table 1 presents as performance measures the bias ("Bias"), standard deviation ("StdDev") and root-mean squared error ("RMSE") of each estimator over 1000 Monte Carlo samples. Since the densities are continuous at c actually, the performance measures are calculated on the basis of $J(c) = 0$. Moreover, only the performance measures with the smoothing parameter b selected for $T_2(c)$ are reported, because there is no substantial difference between values of b chosen for $T_1(c)$ and $T_2(c)$.

It can be immediately found that the RMSE shrinks with the sample size, which indicates consistency of each estimator. Although the Bias of $\hat{J}_M(c)$ is larger than that of $\tilde{J}(c)$, the StdDev of the former is smaller, and as a consequence it tends to yield a smaller RMSE. We can also see that $\tilde{J}(c)$ has extremely small biases for all cases, which confirms that the MBC technique leads to huge bias reduction. The bias-variance trade-off in terms of δ within $\tilde{J}(c)$ (as Proposition 2 suggests) can be also observed.

4.2 Testing for Discontinuity

Second, size and power properties of the test statistic $T(c)$ are investigated. In what follows, $T_1(c)$ and $T_2(c)$ are compared with McCrary's (2008) test statistic based on the difference between logarithms of two density estimates, denoted as $T_M(c)$. Implementation details of each test statistic are the same as described above. The Monte Carlo design in this section is inspired by Otsu, Xu and Matsushita (2013). Let X be drawn with probability γ from the truncated gamma or Weibull distribution with support on $[0, c)$ and with probability $1-\gamma$ from the one with support on (c, ∞). Unless $\gamma = \Pr(X \leq c)$, the gamma or Weibull pdf is discontinuous at c. Also denote the measure of discontinuity as $d := \Pr(X \leq c) - \gamma$, where $d \in \{0.00, 0.02, 0.04, 0.06, 0.08, 0.10\}$ and $d > 0$ ($\Leftrightarrow J(c) > 0$) suggests a jump of the pdf at c. For each statistic, the empirical rejection frequencies of the null $H_0 : J(c) = 0$ for $d = 0$ and $d > 0$ indicate its size and power properties, respectively.

TABLES 2-3 ABOUT HERE

Table 2 presents size properties of $T_1(c)$ and $T_2(c)$. Each test statistic exhibits mild under-rejection of the null except a few cases, and the rejection frequencies of $T_2(c)$ are closer to the nominal ones. The rejection frequencies tend to decrease with δ, and substantial over-rejection of the null is not observed for $\delta = 0.81$. Considering that $\delta = 0.81$ also yields nearly unbiased estimates of $J(c)$, we set δ equal to this value for power comparisons.

Table 3 reports power properties of $T_1(c)$ and $T_2(c)$, in comparison with $T_M(c)$. Panel (A) refers to the results from the gamma distribution. It can be observed that the rejection frequency of each test statistic for a given $d > 0$ approaches to one with the sample size n, which indicates consistency of each test. Both $T_1(c)$ and $T_2(c)$ exhibit good power properties without inflating their sizes, and $T_2(c)$ appears to be more powerful than $T_1(c)$. In contrast, $T_M(c)$ exhibits considerable size distortions, and nonetheless its power properties look inferior to those of $T_1(c)$ and $T_2(c)$. It may be argued that the gamma distribution is too advantageous to $T_1(c)$ and $T_2(c)$ in that both rely on the gamma kernel. Hence, the simulation study based on the Weibull distribution could be fair, and the results are reported in Panel (B). Indeed, the size properties of $T_M(c)$ are dramatically improved. However, it is still outperformed in terms of power properties by $T_1(c)$ and $T_2(c)$. Again in this case, it appears that $T_2(c)$ has better power properties than $T_1(c)$. A possible rationale is that because $\tilde{V}_2(c)$ tends to be smaller than $\tilde{V}_1(c)$, as suggested in Proposition 2, $T_2(c)$ is likely to have a large value (i.e., tends to reject the null more often) than $T_1(c)$ under the alternative.

In sum, Monte Carlo results confirm the following two respects. First, the MBC technique achieves huge bias reduction, and the jump-size estimator $\tilde{J}(c)$ yields nearly unbiased estimates. Second, the test statistics $T_1(c)$ and $T_2(c)$ exhibit nice power properties without sacrificing their size properties, whereas the latter appears to be more powerful than the former. It is also worth emphasizing that the superior performance is based simply on first-order asymptotic results. Therefore, assistance of size-adjusting devices such as bootstrapping appears to be unnecessary, unlike most of the smoothed tests employing conventional symmetric kernels.

5 Empirical Illustration

This section applies our estimation and testing procedures of discontinuity in densities to real data. We employ the data sets on fourth and fifth graders of Israeli elementary schools used by Angrist and Lavy (1999). The data sets are made public on the Angrist Data Archive web page, and they are often utilized in empirical application parts of the closely related literature (e.g., Otsu, Xu and Matsushita, 2013; Feir, Lemieux and Marmer, 2016).

Following Maimonides' rule, Israeli public schools make each class size no greater than 40. As a result of strategic behavior on schools' and/or parents' sides, the density of school enrollment counts for each grade may be discontinuous at multiples of 40. Then, setting the cutoff $c = 40, 80, 120, 160$ for enrollment densities of fourth and fifth graders, we estimate the jump size and conduct the test for the null of continuity at each cutoff. Specifically, the results from our truncated gamma-kernel approach are compared with those from McCrary's (2008) BLL method. $T_2(c)$ with $\delta = 0.81$ is chosen as our test statistic because of its better finite-sample properties. The smoothing parameter for our approach and the bandwidth for the BLL method are chosen in the same manners as in Section 4.

FIGURE 2 AND TABLE 4 ABOUT HERE

Table 4 presents estimation and testing results on discontinuity in enrollment densities, where $\hat{f}_{-}^{M}(c)$ and $\hat{f}_{+}^{M}(c)$ are BLL estimates of left and right limits of the density at the cutoff c, respectively. For convenience, density estimates with (possible) discontinuity points at $c = 40, 120$ are plotted in Figure 2. Table 4 shows remarkable differences between estimation results from McCrary's (2008) and our procedures. The former finds upward jump estimates only at $c = 40$ for each grade. On the other hand, the latter yields upward jump estimates at $c = 40, 80$ and downward jump (or drop) estimates at $c = 120, 160$ for each grade. In addition, Figure 2 illustrates that the truncated gamma density estimators tend to capture peaks and troughs more clearly. Testing results also differ. While McCrary's (2008) test rejects the null of continuity at the cutoff only for three cases (i.e., $c = 40, 120$ for fourth graders and $c = 40$ for fifth graders), rejections of the null by our test include additional two cases (i.e., $c = 160$ for fourth graders and $c = 120$ for fifth graders) as well as the three cases. This appears to reflect better finite-sample power properties of $T_2(c)$ reported in Section 4.

6 Conclusion

This paper has developed estimation and testing procedures on discontinuity in densities with positive support. Our proposal is built on smoothing by the gamma kernel. To preserve its appealing properties, we split the gamma kernel into two parts at a given (dis)continuity point and construct two truncated kernels after re-normalization. The jump-size magnitude of the density at the point can be estimated nonparametrically by two truncated kernels and the MBC technique by Terrell and Scott (1980). The estimator is easy to implement, and its convergence properties are explored by means of various approximation techniques on incomplete gamma functions. Given the jump-size estimator, two versions of test statistics for the null of continuity at a given point are also proposed, and a smoothing parameter selection method under the power-optimality criterion is tailored to our testing procedure. Furthermore, estimation theory of the entire density in the presence of a discontinuity point is provided. It is demonstrated that density estimators smoothed by the truncated gamma kernels admit the same bias and variance approximations as the gamma kernel density estimator does. Monte Carlo simulations indicate that the jump-size estimator is nearly unbiased when there is no jump in the true density, and that the test statistics with power-optimal smoothing parameter values plugged in enjoy more power than McCrary's (2008) BLL-based test does, without sacrificing their size properties.

We conclude this paper by noting a few research extensions. First, the assumption of a single (known) point of discontinuity may be relaxed. It is worth investigating the cases for more than one (known) point of discontinuity or those for even unknown (finite) number of discontinuity points. For the latter, locations of discontinuity points are estimated first and then the corresponding upper and lower limits of the density can be evaluated at each estimated location. Second, while our focus has been exclusively on univariate densities, the discontinuity analysis may be extended to multivariate densities.

A Appendix

A.1 List of Useful Formulae

The formulae below are frequently used in the technical proofs.

Stirling's formula.

$$\Gamma(a+1) = \sqrt{2\pi} a^{a+1/2} \exp(-a) \left\{1 + \frac{1}{12a} + O\left(a^{-2}\right)\right\} \text{ as } a \to \infty. \tag{A1}$$

Recursive formulae on incomplete gamma functions.

$$\gamma(a+1,z) = a\gamma(a,z) - z^a \exp(-z) \text{ for } a, z > 0. \tag{A2}$$

$$\Gamma(a+1,z) = a\Gamma(a,z) + z^a \exp(-z) \text{ for } a, z > 0. \tag{A3}$$

Identity among gamma and incomplete gamma functions.

$$\gamma(a,z) + \Gamma(a,z) = \Gamma(a) \text{ for } a, z > 0. \tag{A4}$$

A.2 Proof of Proposition 1

To save space, we only provide approximations to the bias and variance of $\hat{f}_{-}(c)$. Using (A3), (A4) and (A5) gives the results for $\hat{f}_{+}(c)$ in the same manner. The proof utilizes the following asymptotic expansion:

$$\frac{\gamma(a,a)}{\Gamma(a)} = \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\left\{\frac{1}{3a^{1/2}} + \frac{1}{540a^{3/2}} + O\left(a^{-5/2}\right)\right\} \text{ as } a \to \infty. \qquad \text{(A5)}$$

This can be obtained by either letting $x \downarrow 0$ in equation (1) of Pagurova (1965) or putting $\eta = 0$ in equation (1.4) of Temme (1979). Then, putting $z = a$ in (A2) and then substituting (A1) and (A5), we have

$$\begin{aligned}\frac{\gamma(a+1,a)}{\Gamma(a+1)} &= \frac{\gamma(a,a)}{\Gamma(a)} - \frac{a^a \exp(-a)}{\Gamma(a+1)} \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\left(-\frac{2}{3}a^{-1/2} + \frac{23}{270}a^{-3/2}\right) + O\left(a^{-5/2}\right). \qquad \text{(A6)}\end{aligned}$$

Bias. By the change of variable $v := u/b$,

$$E\left\{\hat{f}_{-}(c)\right\} = \int_0^c \frac{u^{c/b}\exp(-u/b)}{b^{c/b+1}\gamma(c/b+1,c/b)} f(u)\,du = \int_0^a f(bv)\left\{\frac{v^a \exp(-v)}{\gamma(a+1,a)}\right\} dv,$$

where $a := c/b$ and the object inside brackets of the right-hand side is a pdf on the interval $[0,a]$. Then, a second-order Taylor expansion of $f(bv)$ around $bv = c$ (from below) yields

$$\begin{aligned}E\left\{\hat{f}_{-}(c)\right\} &= f_{-}(c) + bf_{-}^{(1)}(c)\left\{\frac{\gamma(a+2,a)}{\gamma(a+1,a)} - a\right\} \\ &\quad + \frac{b^2}{2}f_{-}^{(2)}(c)\left\{\frac{\gamma(a+3,a)}{\gamma(a+1,a)} - 2a\frac{\gamma(a+2,a)}{\gamma(a+1,a)} + a^2\right\} + R_{\hat{f}_{-}(c)}, \qquad \text{(A7)}\end{aligned}$$

where

$$R_{\hat{f}_{-}(c)} := \frac{b^2}{2}\int_0^a \left\{f_{-}^{(2)}(\xi) - f_{-}^{(2)}(c)\right\}(v-a)^2\left\{\frac{v^a\exp(-v)}{\gamma(a+1,a)}\right\} dv$$

is the remainder term with $\xi = \theta(bv) + (1-\theta)c$ for some $\theta \in (0,1)$.

We approximate the leading bias terms first. Using (A2) recursively, we have

$$
\begin{aligned}
\gamma(a+2,a) &= (a+1)\gamma(a+1,a) - a^{a+1}\exp(-a), \text{ and} \\
\gamma(a+3,a) &= (a+2)(a+1)\gamma(a+1,a) - 2(a+1)a^{a+1}\exp(-a).
\end{aligned}
$$

It follows from (A1) and (A6) that

$$
\begin{aligned}
\frac{\gamma(a+2,a)}{\gamma(a+1,a)} - a &= 1 - \frac{a^{a+1}\exp(-a)}{\Gamma(a+1)}\left\{\frac{\gamma(a+1,a)}{\Gamma(a+1)}\right\}^{-1} \\
&= -\sqrt{\frac{2}{\pi}}a^{1/2} + \left(1-\frac{4}{3\pi}\right) + O\left(a^{-1/2}\right), \text{ and} \\
\frac{\gamma(a+3,a)}{\gamma(a+1,a)} - 2a\frac{\gamma(a+2,a)}{\gamma(a+1,a)} + a^2 &= a + 2 - 2\frac{a^{a+1}\exp(-a)}{\Gamma(a+1)}\left\{\frac{\gamma(a+1,a)}{\Gamma(a+1)}\right\}^{-1} \\
&= a + O\left(a^{1/2}\right).
\end{aligned}
$$

Substituting these into the second and third terms on the right-hand side of (A7) and recognizing that $a = c/b$, we obtain

$$
\begin{aligned}
& bf_{-}^{(1)}(c)\left\{\frac{\gamma(a+2,a)}{\gamma(a+1,a)} - a\right\} + \frac{b^2}{2}f_{-}^{(2)}(c)\left\{\frac{\gamma(a+3,a)}{\gamma(a+1,a)} - 2a\frac{\gamma(a+2,a)}{\gamma(a+1,a)} + a^2\right\} \\
= & -\sqrt{\frac{2}{\pi}}c^{1/2}f_{-}^{(1)}(c)\,b^{1/2} + \left\{\left(1-\frac{4}{3\pi}\right)f_{-}^{(1)}(c) + \frac{c}{2}f_{-}^{(2)}(c)\right\}b + o(b).
\end{aligned}
$$

The remaining task is to demonstrate that $R_{\hat{f}_{-}(c)} = o(b)$. It follows from Hölder-continuity of $f^{(2)}(\cdot)$ and $v \le c/b = a$ that

$$
\left|f^{(2)}(\xi) - f_{-}^{(2)}(c)\right| \le L|\xi - c|^{\varsigma} = L\theta^{\varsigma}b^{\varsigma}(a-v)^{\varsigma}.
$$

Using Hölder's inequality and the fact that $v^{a}\exp(-v)/\gamma(a+1,a)$ is a density on $[0,a]$, we have

$$
\begin{aligned}
\left|R_{\hat{f}_{-}(c)}\right| &\le \frac{L\theta^{\varsigma}}{2}b^{2+\varsigma}\int_0^a (a-v)^{2+\varsigma}\left\{\frac{v^{a}\exp(-v)}{\gamma(a+1,a)}\right\}dv \\
&\le \frac{L\theta^{\varsigma}}{2}b^{2+\varsigma}\left[\int_0^a (a-v)^{3}\left\{\frac{v^{a}\exp(-v)}{\gamma(a+1,a)}\right\}dv\right]^{(2+\varsigma)/3},
\end{aligned}
$$

where

$$\int_0^a (a-v)^3 \left\{ \frac{v^a \exp(-v)}{\gamma(a+1,a)} \right\} dv = a^3 - 3a^2 \frac{\gamma(a+2,a)}{\gamma(a+1,a)} + 3a \frac{\gamma(a+3,a)}{\gamma(a+1,a)} - \frac{\gamma(a+4,a)}{\gamma(a+1,a)}$$
$$= O\left(a^{3/2}\right)$$

by using (A1) and (A6) repeatedly. Finally, substituting $a = c/b$ yields

$$\left| R_{\hat{f}_-(c)} \right| \le O\left(b^{2+\varsigma}\right) O\left\{b^{-(1+\varsigma/2)}\right\} = O\left(b^{1+\varsigma/2}\right) = o(b),$$

which establishes the bias approximation.

Variance. In

$$Var\left\{\hat{f}_-(c)\right\} = \frac{1}{n} E\left\{K^-_{G(c,b;c)}(X_i)\right\}^2 + O\left(n^{-1}\right),$$

we make an approximation to $E\left\{K^-_{G(c,b;c)}(X_i)\right\}^2$. By the change of variable $w := 2u/b$ and $a = c/b$,

$$E\left\{K^-_{G(c,b;c)}(X_i)\right\}^2 = \int_0^c \frac{u^{2c/b} \exp(-2u/b)}{b^{2(c/b+1)} \gamma^2 (c/b+1, c/b)} f(u)\, du$$
$$= b^{-1} \frac{\gamma(2a+1,2a)}{2^{2a+1}\gamma^2(a+1,a)} \int_0^{2a} f\left(\frac{bw}{2}\right) \left\{ \frac{w^{2a}\exp(-w)}{\gamma(2a+1,2a)} \right\} dw,$$

where the object inside brackets of the right-hand side is again a pdf. As before, the integral part can be approximated by $f_-(c) + O\left(b^{1/2}\right)$. Moreover, it follows from (A6), the argument on p.474 of Chen (2000) and $a = c/b$ that the multiplier part is

$$\left\{\frac{\gamma(2a+1,2a)}{\Gamma(2a+1)}\right\} \left\{\frac{\gamma(a+1,a)}{\Gamma(a+1)}\right\}^{-2} \left\{\frac{b^{-1}\Gamma(2a+1)}{2^{2a+1}\Gamma^2(a+1)}\right\} = \frac{b^{-1/2}}{\sqrt{\pi} c^{1/2}} + o\left(b^{-1/2}\right).$$

Therefore,

$$Var\left\{\hat{f}_-(c)\right\} = \frac{1}{nb^{1/2}} \frac{f_-(c)}{\sqrt{\pi} c^{1/2}} + o\left(n^{-1}b^{-1/2}\right). \blacksquare$$

A.3 Proof of Theorem 1

The proof requires the following lemma.

Lemma A1.

$$E\left\{K^{\pm}_{G(c,b;c)}\left(X_i\right)\right\}^3 = O\left(b^{-1}\right).$$

A.3.1 Proof of Lemma A1

To save space, we concentrate only on $E\left\{K^{-}_{G(c,b;c)}\left(X_i\right)\right\}^3$. By the change of variable $t := 3u/b$ and $a = c/b$,

$$\begin{aligned} E\left\{K^{-}_{G(c,b;c)}\left(X_i\right)\right\}^3 &= \int_0^c \frac{u^{3c/b}\exp\left(-3u/b\right)}{b^{3(c/b+1)}\gamma^3\left(c/b+1,c/b\right)} f\left(u\right) du \\ &= b^{-2}\frac{\gamma\left(3a+1,3a\right)}{3^{3a+1}\gamma^3\left(a+1,a\right)}\int_0^{3a} f\left(\frac{bt}{3}\right)\left\{\frac{t^{3a}\exp\left(-t\right)}{\gamma\left(3a+1,3a\right)}\right\} dt, \end{aligned}$$

where the integral part is $f_{-}\left(c\right) + O\left(b^{1/2}\right)$ as before. On the other hand, by (A1) and (A6), the multiplier part can be approximated by

$$\left\{\frac{\gamma\left(3a+1,3a\right)}{\Gamma\left(3a+1\right)}\right\}\left\{\frac{\gamma\left(a+1,a\right)}{\Gamma\left(a+1\right)}\right\}^{-3}\left\{\frac{b^{-2}\Gamma\left(3a+1\right)}{3^{3a+1}\Gamma^3\left(a+1\right)}\right\} = \frac{2}{\sqrt{3}\pi c}b^{-1} + o\left(b^{-1}\right),$$

which establishes the stated result. ■

A.3.2 Proof of Theorem 1

Let

$$\begin{aligned} \hat{f}_{\pm,b}\left(c\right) &= E\left\{\hat{f}_{\pm,b}\left(c\right)\right\} + \left[\hat{f}_{\pm,b}\left(c\right) - E\left\{\hat{f}_{\pm,b}\left(c\right)\right\}\right] := I^{\pm}_b\left(c\right) + Z^{\pm}, \text{ and} \\ \hat{f}_{\pm,b/\delta}\left(c\right) &= E\left\{\hat{f}_{\pm,b/\delta}\left(c\right)\right\} + \left[\hat{f}_{\pm,b/\delta}\left(c\right) - E\left\{\hat{f}_{\pm,b/\delta}\left(c\right)\right\}\right] := I^{\pm}_{b/\delta}\left(c\right) + W^{\pm}. \end{aligned}$$

Then, by a similar argument to the proof for Theorem 1 of Hirukawa and Sakudo (2014) and Proposition 2,

$$\begin{aligned} \tilde{J}\left(c\right) &= \left\{I^{+}_b\left(c\right)\right\}^{\frac{1}{1-\delta^{1/2}}}\left\{I^{+}_{b/\delta}\left(c\right)\right\}^{-\frac{\delta^{1/2}}{1-\delta^{1/2}}} - \left\{I^{-}_b\left(c\right)\right\}^{\frac{1}{1-\delta^{1/2}}}\left\{I^{-}_{b/\delta}\left(c\right)\right\}^{-\frac{\delta^{1/2}}{1-\delta^{1/2}}} \\ &\quad + \left(\frac{1}{1-\delta^{1/2}}\right)\left\{\left(Z^{+} - \delta^{1/2}W^{+}\right) - \left(Z^{-} - \delta^{1/2}W^{-}\right)\right\} + R_{\tilde{J}(c)}, \end{aligned}$$

where it can be shown that the remainder term $R_{\tilde{J}(c)} = o_p\left(n^{-1/2}b^{-1/4}\right)$. Because $E\left(Z^{\pm}\right) = E\left(W^{\pm}\right) = 0$,

$$\begin{aligned} E\left\{\tilde{J}(c)\right\} &\sim \left\{I_b^{+}(c)\right\}^{\frac{1}{1-\delta^{1/2}}}\left\{I_{b/\delta}^{+}(c)\right\}^{-\frac{\delta^{1/2}}{1-\delta^{1/2}}} - \left\{I_b^{-}(c)\right\}^{\frac{1}{1-\delta^{1/2}}}\left\{I_{b/\delta}^{-}(c)\right\}^{-\frac{\delta^{1/2}}{1-\delta^{1/2}}} \\ &\sim J(c) + B(c)\,b, \end{aligned}$$

where

$$\begin{aligned} B(c) &= \left(\frac{1}{\delta^{1/2}}\right)\left[\frac{c}{\pi}\left\{\frac{\left(f_+^{(1)}(c)\right)^2}{f_+(c)} - \frac{\left(f_-^{(1)}(c)\right)^2}{f_-(c)}\right\}\right. \\ &\left. -\left\{\left(1-\frac{4}{3\pi}\right)\left(f_+^{(1)}(c) - f_-^{(1)}(c)\right) + \frac{c}{2}\left(f_+^{(2)}(c) - f_-^{(2)}(c)\right)\right\}\right]. \end{aligned}$$

Therefore,

$$\begin{aligned} \sqrt{nb^{1/2}}\left\{\tilde{J}(c) - J(c)\right\} &= \sqrt{nb^{1/2}}\left[\tilde{J}(c) - E\left\{\tilde{J}(c)\right\}\right] + \sqrt{nb^{1/2}}\left[E\left\{\tilde{J}(c)\right\} - J(c)\right] \\ &= \sqrt{nb^{1/2}}\left(\frac{1}{1-\delta^{1/2}}\right)\left\{\left(Z^{+} - \delta^{1/2}W^{+}\right) - \left(Z^{-} - \delta^{1/2}W^{-}\right)\right\} \\ &\quad + \sqrt{nb^{1/2}}\left\{B(c)\,b + o(b)\right\} + o_p(1), \end{aligned}$$

where the second term on the right hand side becomes asymptotically negligible if $nb^{5/2} \to 0$.

The remaining task is to establish the asymptotic normality of the first term. Due to the disjunction of two truncated kernels $K_{G(c,b;c)}^{\pm}(\cdot)$, the asymptotic variance of the term, denoted as $V(c)$, is just the sum of asymptotic variances of $\tilde{f}_{\pm}(c)$ given in Proposition 2. Hence, we need only to establish Liapunov's condition. Denoting

$$\begin{aligned} Z^{\pm} &= \sum_{i=1}^{n}\left(\frac{1}{n}\right)\left[K_{G(c,b;c)}^{\pm}(X_i) - E\left\{K_{G(c,b;c)}^{\pm}(X_i)\right\}\right] := \sum_{i=1}^{n}\left(\frac{1}{n}\right)Z_i^{\pm}, \text{ and} \\ W^{\pm} &= \sum_{i=1}^{n}\left(\frac{1}{n}\right)\left[K_{G(c,b/\delta;c)}^{\pm}(X_i) - E\left\{K_{G(c,b/\delta;c)}^{\pm}(X_i)\right\}\right] := \sum_{i=1}^{n}\left(\frac{1}{n}\right)W_i^{\pm}, \end{aligned}$$

we can rewrite the term as

$$
\begin{aligned}
& \sqrt{nb^{1/2}}\left(\frac{1}{1-\delta^{1/2}}\right)\left\{\left(Z^{+}-\delta^{1/2}W^{+}\right)-\left(Z^{-}-\delta^{1/2}W^{-}\right)\right\} \\
= & \sum_{i=1}^{n}\sqrt{\frac{b^{1/2}}{n}}\left(\frac{1}{1-\delta^{1/2}}\right)\left\{\left(Z_i^{+}-\delta^{1/2}W_i^{+}\right)-\left(Z_i^{-}-\delta^{1/2}W_i^{-}\right)\right\}:=\sum_{i=1}^{n}Y_i.
\end{aligned}
$$

It follows from $0<\delta<1$ that

$$
E\left|Y_i\right|^3 \le \frac{b^{3/4}}{n^{3/2}}\left(\frac{1}{1-\delta^{1/2}}\right)^3 E\left(\left|Z_i^{+}\right|+\left|W_i^{+}\right|+\left|Z_i^{-}\right|+\left|W_i^{-}\right|\right)^3.
$$

Because the expected value part is $O\left(b^{-1}\right)$ by Lemma A1, $E\left|Y_i\right|^3=O\left(n^{-3/2}b^{-1/4}\right)$. It is also straightforward to see that $Var\left(Y_i\right)=O\left(n^{-1}\right)$. Therefore,

$$
\frac{\sum_{i=1}^{n}E\left|Y_i\right|^3}{\left\{\sum_{i=1}^{n}Var\left(Y_i\right)\right\}^{3/2}}=O\left(n^{-1/2}b^{-1/4}\right)\to 0,
$$

or Liapunov's condition holds. This completes the proof. ■

A.4 Proof of Proposition 3

The proof closely follows the one for Proposition 1 of Hirukawa and Sakudo (2016). It follows from Theorem 1 that $E\left\{\tilde{J}(c)\right\}=J(c)+O(b)$, $Var\left\{\tilde{J}(c)\right\}=O\left(n^{-1}b^{-1/2}\right)$ and $\tilde{V}(c)\xrightarrow{p}V(c)$, regardless of whether H_0 or H_1 may be true. Therefore, $\tilde{J}(c)=J(c)+O(b)+O_p\left(n^{-1/2}b^{-1/4}\right)\xrightarrow{p}J(c)\neq 0$ under H_1, and thus $|T(c)|$ is a divergent stochastic sequence with an expansion rate of $n^{1/2}b^{1/4}$. The result immediately follows. ■

A.5 Proof of Theorem 2

To demonstrate this theorem, we must rely on different asymptotic expansions, depending on the positions of the design point x and the truncation point c. For notational convenience, put $(a,z)=(x/b,c/b)$. The proof requires the following lemma.

Lemma A2. *For $a > 0$ and $z > \max\{1, a\}$,*

$$\Gamma(a+1,z) \le \begin{cases} z^a \exp(-z) + \exp(-z) & \text{for } 0 < a \le 1 \\ (a+1) z^a \exp(-z) + \Gamma(a+1)\exp(-z) & \text{for } a > 1 \end{cases}.$$

A.5.1 Proof of Lemma A2

For $0 < a \le 1$, it follows from an elementary inequality on the upper incomplete gamma function (e.g., equation (1.05) on p.67 of Olver, 1974) and $z > 1$ that

$$\Gamma(a,z) \le z^{a-1}\exp(-z) \le \exp(-z). \tag{A8}$$

Then, by (A3),

$$\Gamma(a+1,z) = z^a \exp(-z) + a\Gamma(a,z) \le z^a \exp(-z) + 1 \cdot \exp(-z).$$

Next, for $a > 1$ and $a \in \mathbb{N}$, using (A3) recursively yields

$$\begin{aligned}\Gamma(a+1,z) &= z^a \exp(-z)\left\{1 + \frac{a}{z} + \frac{a(a-1)}{z^2} + \cdots + \frac{a(a-1)\cdots 2}{z^{a-1}}\right\} \\ &\quad + a(a-1)\cdots 2 \cdot 1 \cdot \Gamma(1,z),\end{aligned}$$

where the sum inside the brackets is bounded by $a\,(\le a+1)$. Then, by (A8),

$$\Gamma(a+1,z) \le (a+1) z^a \exp(-z) + \Gamma(a+1)\exp(-z).$$

Finally, for $a > 1$ and $a \notin \mathbb{N}$, we have

$$\begin{aligned}\Gamma(a+1,z) &= z^a \exp(-z)\left\{1 + \frac{a}{z} + \frac{a(a-1)}{z^2} + \cdots + \frac{a(a-1)\cdots(a-\lfloor a\rfloor+1)}{z^{\lfloor a\rfloor}}\right\} \\ &\quad + a(a-1)\cdots(a-\lfloor a\rfloor)\,\Gamma(a-\lfloor a\rfloor, z).\end{aligned}$$

where the sum inside the brackets is bounded by $\lfloor a\rfloor + 1\,(\le a+1)$. Because $0 < a - \lfloor a\rfloor < 1$, $\Gamma(a-\lfloor a\rfloor) > 1$ and thus

$$a(a-1)\cdots(a-\lfloor a\rfloor) = \frac{\Gamma(a+1)}{\Gamma(a-\lfloor a\rfloor)} \le \Gamma(a+1).$$

Therefore, again by (A8),

$$\Gamma(a+1,z) \le (a+1) z^a \exp(-z) + \Gamma(a+1)\exp(-z). \quad \blacksquare$$

A.5.2 Proof of Theorem 2

(i) On $\hat{f}_-(x)$:

We consider different approximations to incomplete gamma functions depending on the position of x. When $x/b \to \infty$, $z > a$ and $a, z \to \infty$ hold. Hence, the case for $a > 1$ of Lemma A2 applies, and thus

$$\frac{\Gamma(a+1,z)}{\Gamma(a+1)} \leq (a+1)\left\{\frac{z^a \exp(-z)}{\Gamma(a+1)}\right\} + \exp(-z).$$

It follows from (A1) and $\rho := a/z \in (0,1)$ that

$$\begin{aligned} \frac{z^a \exp(-z)}{\Gamma(a+1)} &= \left\{\frac{1+O(a^{-1})}{\sqrt{2\pi}}\right\} a^{-1/2} \exp\left\{a \ln\left(\frac{e}{\rho e^{1/\rho}}\right)\right\} \\ &= O\left[a^{-1/2} \exp\left\{a \ln\left(\frac{e}{\rho e^{1/\rho}}\right)\right\}\right], \end{aligned} \tag{A9}$$

where $e/\left(\rho e^{1/\rho}\right) \in (0,1)$ holds. Then,

$$\frac{\Gamma(a+1,z)}{\Gamma(a+1)} = O\left[a^{1/2} \exp\left\{a \ln\left(\frac{e}{\rho e^{1/\rho}}\right)\right\}\right].$$

On the other hand, when $x/b \to \kappa \in (0,\infty)$, putting $a \to \kappa$ and $z \to \infty$ in Lemma A2 yields

$$\frac{\Gamma(a+1,z)}{\Gamma(a+1)} = O\left\{z^\kappa \exp(-z)\right\}.$$

By (A4), we finally have

$$\frac{\gamma(a+1,z)}{\Gamma(a+1)} = 1 + \begin{cases} O\left[a^{1/2} \exp\left\{a \ln\left(e/\left(\rho e^{1/\rho}\right)\right)\right\}\right] & \text{if } x/b \to \infty \\ O\left\{z^\kappa \exp(-z)\right\} & \text{if } x/b \to \kappa \end{cases}. \tag{A10}$$

Bias. By (A9), (A10), and $(a, z) = (x/b, c/b)$,

$$\begin{aligned}
&\frac{\gamma(a+2,z)}{\gamma(a+1,z)} - a \\
&= 1 - \frac{z^{a+1}\exp(-z)}{\Gamma(a+1)}\left\{\frac{\gamma(a+1,z)}{\Gamma(a+1)}\right\}^{-1} \\
&= 1 + \begin{cases} O\left[a^{1/2}\exp\left\{a\ln\left(e/\left(\rho e^{1/\rho}\right)\right)\right\}\right] & \text{if } x/b \to \infty \\ O\left\{z^{\kappa}\exp(-z)\right\} & \text{if } x/b \to \kappa \end{cases} \\
&= 1 + \begin{cases} O\left[b^{-1/2}\exp\left\{(x/b)\ln\left(e/\left(\rho e^{1/\rho}\right)\right)\right\}\right] & \text{if } x/b \to \infty \\ O\left\{b^{-\kappa}\exp(-c/b)\right\} & \text{if } x/b \to \kappa \end{cases}, \text{ and} \\
&\frac{\gamma(a+3,z)}{\gamma(a+1,z)} - 2a\frac{\gamma(a+2,z)}{\gamma(a+1,z)} + a^2 \\
&= a + 2 - (z - a + 2)\frac{z^{a+1}\exp(-z)}{\Gamma(a+1)}\left\{\frac{\gamma(a+1,z)}{\Gamma(a+1)}\right\}^{-1} \\
&= a + 2 + \begin{cases} O\left[a^{3/2}\exp\left\{a\ln\left(e/\left(\rho e^{1/\rho}\right)\right)\right\}\right] & \text{if } x/b \to \infty \\ O\left\{z^{\kappa+1}\exp(-z)\right\} & \text{if } x/b \to \kappa \end{cases} \\
&= \frac{x}{b} + 2 + \begin{cases} O\left[b^{-3/2}\exp\left\{(x/b)\ln\left(e/\left(\rho e^{1/\rho}\right)\right)\right\}\right] & \text{if } x/b \to \infty \\ O\left\{b^{-\kappa-1}\exp(-c/b)\right\} & \text{if } x/b \to \kappa \end{cases}.
\end{aligned}$$

Then, by the argument in the proof of Proposition 1, in either case,

$$E\left\{\hat{f}_-(x)\right\} = f(x) + \left\{f^{(1)}(x) + \frac{x}{2}f^{(2)}(x)\right\}b + o(b).$$

Variance. In

$$E\left\{K^-_{G(x,b;c)}(X_i)\right\}^2 = b^{-1}\frac{\gamma(2a+1,2z)}{2^{2a+1}\gamma^2(a+1,z)}\int_0^{2z} f\left(\frac{bw}{2}\right)\left\{\frac{w^{2a}\exp(-w)}{\gamma(2a+1,2z)}\right\}dw,$$

the integral part is $f(x) + O(b)$ in either case. It also follows from (A10) and the argument on p.474 of Chen (2000) that the multiplier part is

$$\begin{aligned}
&\left\{\frac{\gamma(2a+1,2z)}{\Gamma(2a+1)}\right\}\left\{\frac{\gamma(a+1,z)}{\Gamma(a+1)}\right\}^{-2}\left\{\frac{b^{-1}\Gamma(2a+1)}{2^{2a+1}\Gamma^2(a+1)}\right\} \\
&= \begin{cases} \frac{b^{-1/2}}{2\sqrt{\pi}x^{1/2}} + o\left(b^{-1/2}\right) & \text{if } x/b \to \infty \\ \frac{b^{-1}\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} + o\left(b^{-1}\right) & \text{if } x/b \to \kappa \end{cases}.
\end{aligned}$$

Therefore,

$$Var\left\{\hat{f}_-(x)\right\} = \begin{cases} \frac{1}{nb^{1/2}}\frac{f(x)}{2\sqrt{\pi}x^{1/2}} + o\left(n^{-1}b^{-1/2}\right) & \text{if } x/b \to \infty \\ \frac{1}{nb}\frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}f(x) + o\left(n^{-1}b^{-1}\right) & \text{if } x/b \to \kappa \end{cases}. \blacksquare$$

(ii) On $\hat{f}_+(x)$:

We may focus only on the case for interior x. However, it seems difficult to derive a sharp bound on $\gamma(a+1,z)$ or $\Gamma(a+1,z)$ for the case of $a > z$ and $a, z \to \infty$ based directly on (A2) or (A3). Instead, we turn to the series expansion described in Section 3 of Ferreira, López and Pérez-Sinusía (2005), which is valid for the case of $a > z$, $a, z \to \infty$ and $a - z = O(a)$. The expansion is

$$\gamma(a+1,z) = z^{a+1}\exp(-z)\sum_{k=0}^{\infty} c_k(a)\,\Phi_k(z-a),$$

where the definitions of $\{c_k(a)\}$ and $\{\Phi_k(z-a)\}$ can be found therein. Because the sum is shown to be convergent, the order of magnitude in $\gamma(a+1,z)/\Gamma(a+1)$ is determined by the one in $z^{a+1}\exp(-z)/\Gamma(a+1)$. It follows from (A1) and $\rho' := z/a \in (0,1)$ that

$$\begin{aligned}\frac{z^{a+1}\exp(-z)}{\Gamma(a+1)} &= \left[\frac{\rho'\{1+O(a^{-1})\}}{\sqrt{2\pi}}\right]a^{1/2}\exp\left\{a\ln\left(\frac{\rho' e}{e^{\rho'}}\right)\right\}\\ &= O\left[a^{1/2}\exp\left\{a\ln\left(\frac{\rho' e}{e^{\rho'}}\right)\right\}\right],\end{aligned}$$

where $\rho' e/e^{\rho'} \in (0,1)$ is again the case. Then, by (A4),

$$\frac{\Gamma(a+1,z)}{\Gamma(a+1)} = 1 + O\left[a^{1/2}\exp\left\{a\ln\left(\frac{\rho' e}{e^{\rho'}}\right)\right\}\right].$$

The bias and variance of $\hat{f}_+(x)$ can be approximated as above. ■

A.6 Proof of Theorem 3

Both this proof and the proof of Theorem 4 require three lemmata below.

Lemma A3. *For $\alpha > 0$ and a sufficiently small $b > 0$, pick some design point $x \in [0, \alpha b]$. Then, for $\eta \in (0,c)$,*

$$\int_0^{\eta} K^{-}_{G(x,b;c)}(u)\,du = \int_0^{\eta}\frac{u^{x/b}\exp(-u/b)}{b^{x/b+1}\gamma(x/b+1,c/b)}du \to 1$$

as $b \to 0$.

Lemma A4. *For the design point x defined in Lemma A3, let*

$$\{K_i\}_{i=1}^n := \left\{bK_{G(x,b;c)}^{-}(X_i)\right\}_{i=1}^n.$$

Then,

$$0 \leq K_i \leq C := \max\{1, \alpha^\alpha\} \left\{\frac{\Gamma(\alpha+1)}{\gamma(\alpha+1,\alpha)}\right\} \left\{\frac{1}{\Gamma(a^*)}\right\},$$

where $\Gamma(a^*) := \min_{a>0} \Gamma(a) \approx 0.8856$ for $a^* \approx 1.4616$.

Lemma A5.(Hoeffding, 1963, Theorem 2) *Let $\{X_i\}_{i=1}^n$ be independent and $a_i \leq X_i \leq b_i$ for $i = 1, 2, \ldots, n$. Also write $\bar{X} := (1/n)\sum_{i=1}^n X_i$ and $\mu := E(\bar{X})$. Then, for $\epsilon > 0$,*

$$\Pr\left(\left|\bar{X} - \mu\right| \geq \epsilon\right) \leq 2\exp\left\{-\frac{2n^2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right\}.$$

A.6.1 Proof of Lemma A3

By the change of variable $v := u/b$, the integral can be rewritten as

$$\int_0^{\eta/b} \frac{v^{x/b}\exp(-v)}{\gamma(x/b+1, c/b)} dv = \frac{\gamma(x/b+1, \eta/b)}{\gamma(x/b+1, c/b)}.$$

Because $\eta/b \uparrow \infty$ and $0 \leq x/b \leq \alpha$, (A10) establishes that

$$\frac{\gamma(x/b+1, \eta/b)}{\gamma(x/b+1, c/b)} = \frac{\Gamma(x/b+1) + O\{b^{-\alpha}\exp(-\eta/b)\}}{\Gamma(x/b+1) + O\{b^{-\alpha}\exp(-c/b)\}} \to 1. \blacksquare$$

A.6.2 Proof of Lemma A4

By construction, $K_i \geq 0$ holds. In addition, since the gamma kernel has its mode at the design point x (Chen, 2000, p.473), K_i is bounded by

$$bK_{G(x,b;c)}^{-}(x) = \left(\frac{x}{b}\right)^{x/b} \exp\left(-\frac{x}{b}\right) \left\{\frac{\Gamma(x/b+1)}{\gamma(x/b+1, c/b)}\right\} \left\{\frac{1}{\Gamma(x/b+1)}\right\}. \quad \text{(A11)}$$

For $0 \leq x/b \leq \alpha$, $(x/b)^{x/b} \leq \max\{1, \alpha^\alpha\}$ and $\exp(-x/b) \leq 1$. Moreover, $\gamma(a,z)/\Gamma(a)$ for $a, z > 0$ is monotonously increasing in z and decreasing in a; see, for example, Tricomi (1950, p.276) for details. Because c is an interior point, $\alpha b \leq c$ or $\alpha \leq c/b$ holds. Hence,

$$\frac{\Gamma(x/b+1)}{\gamma(x/b+1,c/b)} \leq \frac{\Gamma(\alpha+1)}{\gamma(\alpha+1,\alpha)}.$$

Finally, it is known that $\Gamma(a^*) := \min_{a>0}\Gamma(a) \approx 0.8856$ for $a^* \approx 1.4616$. Therefore, the right-hand side of (A11) has the upper bound

$$\max\{1,\alpha^{\alpha}\}\cdot 1\cdot\left\{\frac{\Gamma(\alpha+1)}{\gamma(\alpha+1,\alpha)}\right\}\left\{\frac{1}{\Gamma(a^*)}\right\} := C. \blacksquare$$

A.6.3 Proof of Theorem 3

This proof largely follows the one for Theorem 5 of Hirukawa and Sakudo (2015). Without loss of generality, for $\alpha > 0$ and a sufficiently small $b > 0$, pick some design point $x \in [0, \alpha b]$. Then, the proof completes if the following statements hold:

$$\hat{f}_-(x) = E\left\{\hat{f}_-(x)\right\} + o_p(1). \tag{A12}$$

$$E\left\{\hat{f}_-(x)\right\} = E\left\{\hat{f}_-(0)\right\} + o(1). \tag{A13}$$

$$E\left\{\hat{f}_-(0)\right\} \to \infty. \tag{A14}$$

Below we demonstrate (A12)-(A14) one by one. First, (A13) immediately follows from the continuity of $K^-_{G(x,b;c)}(u)$ in x. Second, when $f(x) \to \infty$ as $x \to 0$, it holds that for any $A > 0$, there is some $\eta \in (0,c)$ such that $f(x) > A$ for all $x < \eta$. For the given η, Lemma A3 implies that

$$E\left\{\hat{f}_-(0)\right\} > \int_0^{\eta} K^-_{G(0,b;c)}(u) f(u)\,du > A\int_0^{\eta} K^-_{G(0,b;c)}(u)\,du \to A,$$

which establishes (A14). Third, for $\{K_i\}_{i=1}^n$ defined in Lemma A4, denote their sample average as $\bar{K} := (1/n)\sum_{i=1}^n K_i$. Then, it follows from Lemmata A4 and A5 that for $\epsilon > 0$,

$$\begin{aligned}\Pr\left(\left|\hat{f}_-(x) - E\left\{\hat{f}_-(x)\right\}\right| \geq \epsilon\right) &= \Pr\left(\left|\bar{K} - E(K_i)\right| \geq b\epsilon\right)\\ &\leq 2\exp\left\{-2\left(\frac{\epsilon}{C}\right)^2 nb^2\right\} \to 0.\end{aligned}$$

Therefore, (A12) is also demonstrated, and thus the proof is completed. ■

A.7 Proof of Theorem 4

This proof largely follows the one for Theorem 5.3 of Bouezmarni and Scaillet (2005). As in the proof of Theorem 3, pick some $x \in [0, \alpha b]$. Then, the proof is boiled down to establishing the following statements:

$$\left| \frac{E\left\{\hat{f}_{-}(x)\right\} - f(x)}{f(x)} \right| \to 0, \text{ and} \tag{A15}$$

$$\left| \frac{\hat{f}_{-}(x) - E\left\{\hat{f}_{-}(x)\right\}}{f(x)} \right| \xrightarrow{p} 0, \tag{A16}$$

as $n \to \infty$ and $b, x \to 0$.

We demonstrate (A15) first. An inspection of the proof for Theorem 5.3 of Bouezmarni and Scaillet (2005) reveals that (A15) is shown if their conditions A.2, A.3 and A.5 are fulfilled. Now we check the validity of three conditions. First, because $\int_0^\infty f(x)\,dx = 1$ and $f(x) \to \infty$ as $x \to 0$, there are constants $0 < \underline{C} < \overline{C} < \infty$ such that $\underline{C}x^{-d} \leq f(x) \leq \overline{C}x^{-d}$ for some $d \in (0,1)$ as $x \to 0$. Accordingly, $f^{(1)}(x) = O\left(x^{-d-1}\right)$ for a small value of x. These imply that $x\left|f^{(1)}(x)\right|/f(x) \leq O(1)$, and thus A.2 follows. Second, A.3 has been already established as Lemma A1. Third, let the random variable U be drawn from the distribution with the pdf $K^{-}_{G(x,b;c)}(u)$. Then, by $0 \leq x/b \leq \alpha$ and the expansion techniques used in the proof of Theorem 2, $Var(U) \leq O(b) \to 0$, and thus A.5 also holds.

Furthermore, it follows from Lemmata A4 and A5 that for $\bar{K}$ defined in the proof of Theorem 3 and for $\epsilon > 0$,

$$\begin{aligned} \Pr\left(\left| \frac{\hat{f}_{-}(x) - E\left\{\hat{f}_{-}(x)\right\}}{f(x)} \right| \geq \epsilon \right) &= \Pr\left(\left|\bar{K} - E(K_i)\right| \geq bf(x)\,\epsilon \right) \\ &\leq 2\exp\left\{-2\left(\frac{\epsilon}{C}\right)^2 nb^2 f^2(x)\right\} \to 0. \end{aligned}$$

Therefore, (A16) is also demonstrated, and thus the proof is completed. ■

References

[1] Angrist, J. D., and V. Lavy (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114, 533 - 575.

[2] Bertrand, M., E. Kamenica, and J. Pan (2015): "Gender Identity and Relative Income within Households," *Quarterly Journal of Economics*, 130, 571 - 614.

[3] Bouezmarni, T., and O. Scaillet (2005): "Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data," *Econometric Theory*, 21, 390 - 412.

[4] Chen, S. X. (2000): "Probability Density Function Estimation Using Gamma Kernels," *Annals of the Institute of Statistical Mathematics*, 52, 471 - 480.

[5] Chu, C. K., and P. E. Cheng (1996): "Estimation of Jump Points and Jump Values of a Density Function," *Statistica Sinica*, 6, 79 - 95.

[6] Cline, D. B., and J. D. Hart (1991): "Kernel Estimation of Densities with Discontinuities or Discontinuous Derivatives," *Statistics: A Journal of Theoretical and Applied Statistics*, 22, 69 - 84.

[7] Cowell, F. A., F. H. G. Ferreira, and J. A. Litchfield (1999): "Income Distribution in Brazil 1981-1990: Parametric and Nonparametric Approaches," *Journal of Income Distribution*, 8, 63 - 76.

[8] DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001 - 1044.

[9] Feir, D., T. Lemieux, and V. Marmer (2016): "Weak Identification in Fuzzy Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 34, 185 - 196.

[10] Fernandes, M., and J. Grammig (2005): "Nonparametric Specification Tests for Conditional Duration Models," *Journal of Econometrics*, 127, 35 - 68.

[11] Fernandes, M., E. F. Mendes, and O. Scaillet (2015): "Testing for Symmetry and Conditional Symmetry Using Asymmetric Kernels," *Annals of the Institute of Statistical Mathematics*, 67, 649 - 671.

[12] Ferreira, C., J. L. López, and E. Pérez-Sinusía (2005): "Incomplete Gamma Functions for Large Values of Their Variables," *Advances in Applied Mathematics*, 34, 467 - 485.

[13] Funke, B., and R. Kawka (2015): "Nonparametric Density Estimation for Multivariate Bounded Data Using Two Non-Negative Multiplicative Bias Correction Methods," *Computational Statistics & Data Analysis*, 92, 148 - 162.

[14] Gospodinov, N., and M. Hirukawa (2012): "Nonparametric Estimation of Scalar Diffusion Models of Interest Rates Using Asymmetric Kernels," *Journal of Empirical Finance*, 19, 595 - 609.

[15] Hirukawa, M. (2010): "Nonparametric Multiplicative Bias Correction for Kernel-Type Density Estimation on the Unit Interval," *Computational Statistics & Data Analysis*, 54, 473 - 495.

[16] Hirukawa, M., and M. Sakudo (2014): "Nonnegative Bias Reduction Methods for Density Estimation Using Asymmetric Kernels," *Computational Statistics & Data Analysis*, 75, 112 - 123.

[17] Hirukawa, M., and M. Sakudo (2015): "Family of the Generalised Gamma Kernels: A Generator of Asymmetric Kernels for Nonnegative Data," *Journal of Nonparametric Statistics*, 27, 41 - 63.

[18] Hirukawa, M., and M. Sakudo (2016): "Testing Symmetry of Unknown Densities via Smoothing with the Generalized Gamma Kernels," *Econometrics*, 4, Article No. 28.

[19] Hoeffding, W. (1963): "Probability Inequalities for Sums of Bounded Random Variables," *Journal of the American Statistical Association*, 58, 13 - 30.

[20] Imbens, G. W., and K. Kalyanaraman (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, 933 - 959.

[21] Imbens, G. W., and T. Lemieux (2008): "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615 - 635.

[22] Jones, M. C., O. Linton, and J. P. Nielsen (1995): "A Simple Bias Reduction Method for Density Estimation," *Biometrika*, 82, 327 - 338.

[23] Kristensen, D. (2010): "Nonparametric Filtering of the Realized Spot Volatility: A Kernel-Based Approach," *Econometric Theory*, 26, 60 - 93.

[24] Kulasekera, K. B., and J. Wang (1998): "Bandwidth Selection for Power Optimality in a Test of Equality of Regression Curves," *Statistics & Probability Letters*, 37, 287 - 293.

[25] Liebscher, E. (1990): "Kernel Estimators for Probability Densities with Discontinuities," *Statistics: A Journal of Theoretical and Applied Statistics*, 21, 185 - 196.

[26] McCrary, J. (2008): "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142, 698 - 714.

[27] Olver, F. W. J. (1974): *Asymptotics and Special Functions*. New York: Academic Press.

[28] Otsu, T., K.-L. Xu, and Y. Matsushita (2013): "Estimation and Inference of Discontinuity in Density," *Journal of Business & Economic Statistics*, 31, 507 - 524.

[29] Pagurova, V. I. (1965): "An Asymptotic Formula for the Incomplete Gamma Function," *USSR Computational Mathematics and Mathematical Physics*, 5, 162 - 166.

[30] Porter, J., and P. Yu (2015): "Regression Discontinuity Designs with Unknown Discontinuity Points: Testing and Estimation," *Journal of Econometrics*, 189, 132 - 147.

[31] Saez, E. (2010): "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, 2, 180 - 212.

[32] Temme, N. M. (1979): "The Asymptotic Expansion of the Incomplete Gamma Functions," *SIAM Journal on Mathematical Analysis*, 10, 757 - 766.

[33] Terrell, G. R., and D. W. Scott (1980): "On Improving Convergence Rates for Nonnegative Kernel Density Estimators," *Annals of Statistics*, 8, 1160 - 1163.

[34] Tricomi, F. G. (1950): "Sulla Funzione Gamma Incompleta," *Annali di Matematica Pura ed Applicata*, 31, 263 - 279.

Figure 1: Shapes of True Densities for Monte Carlo Simulations

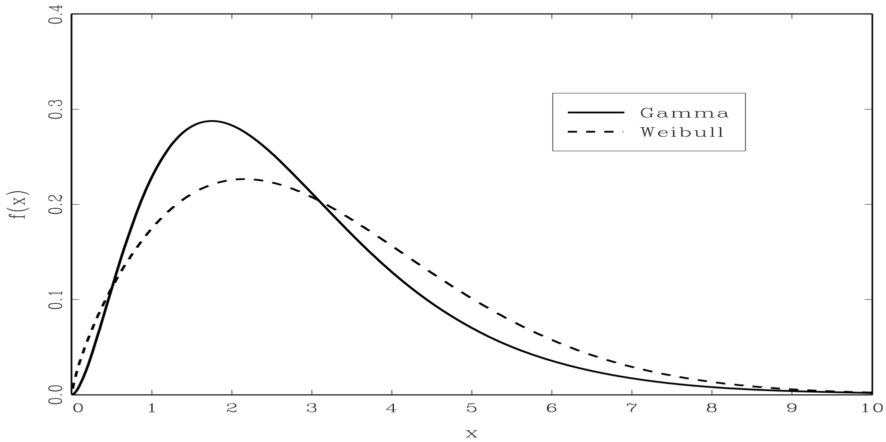
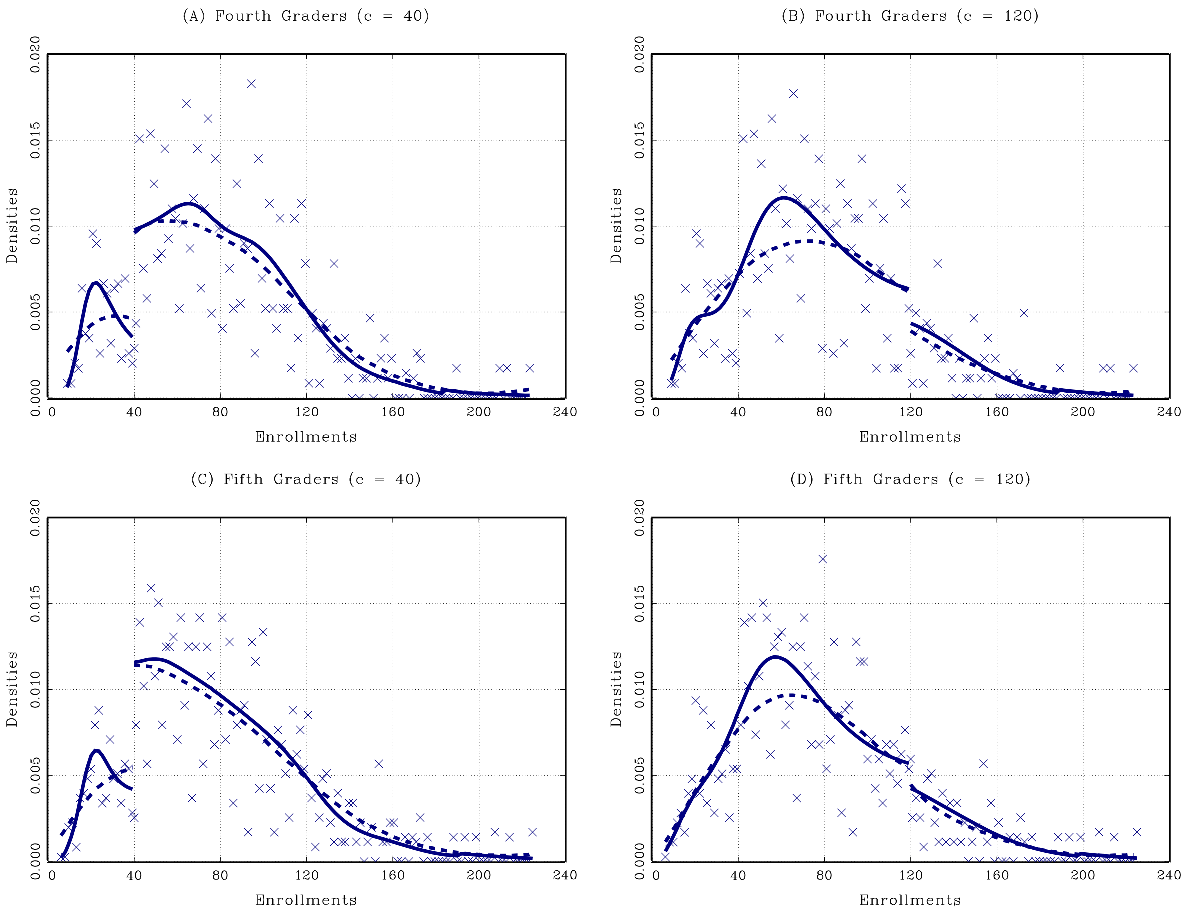


Figure 2: Density Estimates of School Enrollments



Note. In each panel, solid and dashed lines are density estimates via the truncated gamma kernels and the binned local linear method, respectively. The "×" symbols indicate binned data points.

Table 1: Biases, Standard Deviations and RMSEs of Estimators of $J(c)$

				Estimator			
					$\tilde{J}(c)$ with δ		
Distribution	c	n		$\hat{J}_M(c)$	0.49	0.64	0.81
Gamma	1.7057	500	Bias	-0.0381	0.0019	0.0011	0.0006
	(30%)		StdDev	0.0461	0.0786	0.0812	0.0837
			RMSE	0.0598	0.0786	0.0812	0.0837
		1000	Bias	-0.0335	0.0019	0.0010	0.0006
			StdDev	0.0331	0.0588	0.0607	0.0626
			RMSE	0.0471	0.0588	0.0607	0.0626
		2000	Bias	-0.0283	0.0006	0.0002	-0.0000
			StdDev	0.0250	0.0430	0.0445	0.0458
			RMSE	0.0377	0.0430	0.0445	0.0458
	2.4248	500	Bias	-0.0407	-0.0011	-0.0011	-0.0011
	(Med)		StdDev	0.0480	0.0626	0.0648	0.0668
			RMSE	0.0629	0.0626	0.0648	0.0668
		1000	Bias	-0.0323	-0.0012	-0.0012	-0.0013
			StdDev	0.0353	0.0463	0.0479	0.0494
			RMSE	0.0479	0.0463	0.0479	0.0494
		2000	Bias	-0.0240	-0.0004	-0.0004	-0.0004
			StdDev	0.0271	0.0351	0.0363	0.0374
			RMSE	0.0362	0.0351	0.0363	0.0374
Weibull	1.9419	500	Bias	-0.0235	0.0024	0.0012	0.0005
	(30%)		StdDev	0.0416	0.0665	0.0684	0.0704
			RMSE	0.0478	0.0665	0.0684	0.0704
		1000	Bias	-0.0187	0.0035	0.0013	0.0005
			StdDev	0.0302	0.0500	0.0509	0.0523
			RMSE	0.0355	0.0502	0.0509	0.0523
		2000	Bias	-0.0144	0.0017	0.0003	0.0001
			StdDev	0.0225	0.0367	0.0372	0.0383
			RMSE	0.0267	0.0367	0.0372	0.0383
	2.8386	500	Bias	-0.0246	0.0006	0.0004	0.0002
	(Med)		StdDev	0.0405	0.0534	0.0552	0.0569
			RMSE	0.0474	0.0534	0.0552	0.0569
		1000	Bias	-0.0195	0.0002	-0.0001	-0.0003
			StdDev	0.0290	0.0394	0.0408	0.0421
			RMSE	0.0350	0.0394	0.0408	0.0421
		2000	Bias	-0.0149	0.0007	0.0005	0.0004
			StdDev	0.0218	0.0299	0.0309	0.0319
			RMSE	0.0264	0.0299	0.0309	0.0319

Table 2: Finite-Sample Size Properties of Test Statistics for Discontinuity

(%)

				$T_1(c)$ with δ			$T_2(c)$ with δ		
Distribution	c	n	Nominal	0.49	0.64	0.81	0.49	0.64	0.81
Gamma	1.7057	500	5%	3.2	3.1	3.1	4.4	4.0	3.7
	(30%)		10%	7.5	7.5	7.8	8.8	8.8	8.6
		1000	5%	3.9	3.9	3.9	6.1	4.6	4.4
			10%	8.4	8.2	8.2	10.7	9.2	8.9
		2000	5%	3.5	3.6	3.7	4.2	3.9	3.9
			10%	8.1	8.2	8.4	8.8	8.5	8.7
	2.4248	500	5%	3.3	3.6	3.6	3.8	3.9	4.0
	(Med)		10%	7.9	7.8	7.7	8.7	8.6	8.5
		1000	5%	3.7	3.8	3.9	4.1	4.2	4.3
			10%	8.0	8.2	8.0	8.6	8.6	8.6
		2000	5%	4.7	4.7	4.8	4.9	5.0	5.1
			10%	8.8	8.9	9.0	9.4	9.4	9.5
Weibull	1.9419	500	5%	3.2	3.2	3.3	6.2	4.9	4.1
	(30%)		10%	7.7	7.8	7.9	10.7	9.4	9.0
		1000	5%	4.0	4.2	4.2	10.2	6.4	5.2
			10%	8.2	8.3	8.4	14.7	10.7	9.4
		2000	5%	3.8	3.7	3.8	7.7	4.4	4.0
			10%	8.3	8.4	8.3	12.4	9.0	8.5
	2.8386	500	5%	3.6	3.6	3.5	3.9	4.0	3.9
	(Med)		10%	7.8	7.7	7.7	8.5	8.5	8.3
		1000	5%	3.7	3.8	3.8	4.0	4.2	4.2
			10%	8.1	8.1	8.2	8.7	8.4	8.6
		2000	5%	4.7	4.7	4.8	4.9	5.0	5.0
			10%	8.9	9.0	9.2	9.4	9.4	9.6

Table 3: Finite-Sample Power Properties of Test Statistics for Discontinuity

(A) Gamma Distribution (%)

c	n	Test	Nominal	d 0.00	0.02	0.04	0.06	0.08	0.10
1.7057	500	$T_M(c)$	5%	4.6	1.4	2.2	9.2	28.9	55.1
(30%)			10%	10.1	4.1	6.1	17.9	43.5	68.7
		$T_1(c)$	5%	3.1	4.5	10.7	17.9	85.9	98.1
			10%	7.8	9.4	17.0	26.8	88.5	99.2
		$T_2(c)$	5%	3.7	14.2	44.7	60.7	93.8	98.8
			10%	8.6	19.0	53.6	63.9	97.2	99.7
	1000	$T_M(c)$	5%	6.9	1.5	4.4	22.0	58.7	87.5
			10%	13.6	4.1	9.5	35.3	72.9	93.1
		$T_1(c)$	5%	3.9	6.6	12.8	37.1	98.7	100.0
			10%	8.2	12.4	21.4	46.0	99.0	100.0
		$T_2(c)$	5%	4.4	13.9	50.3	90.8	99.5	100.0
			10%	8.9	19.2	54.9	92.7	99.9	100.0
	2000	$T_M(c)$	5%	9.9	1.6	11.7	52.1	90.1	99.5
			10%	19.2	4.5	21.1	66.6	95.6	99.8
		$T_1(c)$	5%	3.7	8.4	36.8	98.8	100.0	100.0
			10%	8.4	15.2	44.1	99.2	100.0	100.0
		$T_2(c)$	5%	3.9	25.1	90.2	99.5	99.9	100.0
			10%	8.7	30.4	94.7	99.9	99.9	100.0
2.4248	500	$T_M(c)$	5%	9.3	3.5	2.6	4.4	9.1	18.7
(Med)			10%	17.4	8.8	6.0	8.7	16.5	29.7
		$T_1(c)$	5%	3.6	4.4	7.1	12.3	20.6	30.1
			10%	7.7	9.1	13.4	21.1	31.0	42.5
		$T_2(c)$	5%	4.0	4.8	7.7	13.6	21.9	32.7
			10%	8.5	9.8	14.5	22.4	32.3	44.2
	1000	$T_M(c)$	5%	11.5	4.0	3.5	9.2	23.1	46.4
			10%	20.1	8.9	7.5	16.0	36.1	61.7
		$T_1(c)$	5%	3.9	5.0	10.4	20.7	35.5	53.2
			10%	8.0	10.4	18.3	32.1	49.0	65.8
		$T_2(c)$	5%	4.3	5.6	11.3	22.2	36.8	55.0
			10%	8.6	11.1	19.2	33.1	50.3	67.0
	2000	$T_M(c)$	5%	12.0	3.6	7.1	23.9	55.6	83.9
			10%	20.7	8.0	13.8	36.0	68.3	91.6
		$T_1(c)$	5%	4.8	7.7	18.1	37.8	60.7	80.2
			10%	9.0	13.9	28.3	50.2	72.6	87.9
		$T_2(c)$	5%	5.1	8.2	18.9	38.8	61.7	85.5
			10%	9.5	14.6	29.2	51.4	73.5	90.7

Table 3 *(Continued)*

(B) Weibull Distribution (%)

c	n	Test	Nominal	d: 0.00	0.02	0.04	0.06	0.08	0.10
1.9419	500	$T_M(c)$	5%	3.4	1.9	4.0	11.8	28.5	48.8
(30%)			10%	7.6	5.0	9.0	21.2	41.2	61.8
		$T_1(c)$	5%	3.3	4.9	14.4	19.1	85.0	97.1
			10%	7.9	9.5	19.8	26.9	88.8	98.6
		$T_2(c)$	5%	4.1	16.7	42.4	57.5	90.6	97.8
			10%	9.0	21.8	53.0	61.0	95.5	99.2
	1000	$T_M(c)$	5%	4.2	2.2	7.5	26.3	55.0	79.3
			10%	8.8	5.2	15.4	39.3	68.2	87.4
		$T_1(c)$	5%	4.2	6.5	12.9	42.1	98.4	99.9
			10%	8.4	12.4	21.2	49.1	99.1	100.0
		$T_2(c)$	5%	5.2	17.5	51.0	88.5	98.9	99.9
			10%	9.4	23.2	56.5	91.1	99.7	100.0
	2000	$T_M(c)$	5%	4.5	3.1	18.2	53.7	84.9	97.5
			10%	9.5	7.1	30.1	66.6	91.5	99.0
		$T_1(c)$	5%	3.8	8.3	53.7	98.8	100.0	100.0
			10%	8.3	14.8	58.0	99.5	100.0	100.0
		$T_2(c)$	5%	4.0	33.2	87.4	98.8	99.9	100.0
			10%	8.5	39.1	93.0	99.6	100.0	100.0
2.8386	500	$T_M(c)$	5%	4.6	2.5	3.0	6.0	12.6	24.4
(Med)			10%	9.4	6.1	6.6	11.8	22.2	36.8
		$T_1(c)$	5%	3.5	4.5	7.2	12.3	19.9	29.1
			10%	7.7	9.2	13.7	21.2	30.3	41.1
		$T_2(c)$	5%	3.9	4.8	7.6	14.6	22.1	43.6
			10%	8.3	9.7	14.4	22.9	32.2	51.9
	1000	$T_M(c)$	5%	5.7	2.5	4.8	12.7	31.0	55.2
			10%	11.2	6.2	9.1	22.5	45.0	68.7
		$T_1(c)$	5%	3.8	5.1	10.2	20.2	34.3	50.9
			10%	8.2	10.7	18.1	31.5	47.3	63.8
		$T_2(c)$	5%	4.2	5.7	11.3	21.3	36.7	61.7
			10%	8.6	11.3	19.0	32.5	48.9	70.6
	2000	$T_M(c)$	5%	6.6	3.1	10.0	31.8	63.9	87.6
			10%	12.7	6.7	17.7	46.0	76.1	93.3
		$T_1(c)$	5%	4.8	7.8	18.0	36.2	58.6	79.8
			10%	9.2	14.1	28.1	49.0	70.8	87.1
		$T_2(c)$	5%	5.0	8.1	18.6	37.8	64.7	99.2
			10%	9.6	14.6	28.7	49.9	74.6	99.5

Note. The value of δ for each of $T_1(c)$ and $T_2(c)$ is set equal to 0.81.

Table 4: Estimation and Testing for the Discontinuity of Densities of School Enrollments

		Binned Local Linear Method				Truncated Kernel Method			
n	c	$\hat{f}^M_-(c)$	$\hat{f}^M_+(c)$	$\hat{J}_M(c)$	$T_M(c)$	$\tilde{f}_-(c)$	$\tilde{f}_+(c)$	$\tilde{J}(c)$	$T_2(c)$
					(a) Fourth Graders:				
2059	40	0.0046	0.0096	0.0050	**5.61**	0.0034	0.0098	0.0064	**5.76**
	80	0.0103	0.0097	-0.0006	-0.62	0.0086	0.0090	0.0003	0.24
	120	0.0061	0.0039	-0.0022	**-3.35**	0.0063	0.0044	-0.0020	**-3.55**
	160	0.0011	0.0009	-0.0003	-0.84	0.0013	0.0005	-0.0008	**-2.88**
					(b) Fifth Graders:				
2029	40	0.0055	0.0114	0.0059	**6.29**	0.0042	0.0116	0.0074	**6.28**
	80	0.0107	0.0098	-0.0009	-0.98	0.0087	0.0103	0.0017	1.25
	120	0.0054	0.0045	-0.0009	-1.20	0.0057	0.0043	-0.0014	**-2.84**
	160	0.0014	0.0011	-0.0003	-0.80	0.0014	0.0010	-0.0004	-1.28

Note. The value of δ for $T_2(c)$ is set equal to 0.81. Values of test statistics in bold faces indicate significance at the 5% level.