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

Benedikt Funke *TU Dortmund*

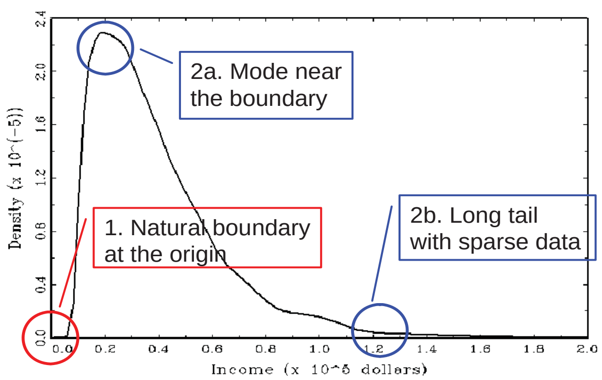
Masayuki Hirukawa *Setsunan University*

27 January 2017
Kanazawa University

Motivation

1. Nonparametric statistics:
   - Inference on possibly discontinuous densities (e.g., Liebscher, 1990; Cline and Hart, 1991; Chu and Cheng, 1996).
2. Economics:
   - In regression discontinuity designs ("RDD"), detection of discontinuity in the density of the running variable at the cutoff suggests evidence of strategic behavior or manipulation.
   - Little attention has been paid to estimation and inference on jump-size magnitudes of densities at discontinuity points up until recently.
     - McCrary (2008) ⇒ binned local linear ("BLL") regression
     - Otsu, Xu and Matsushita (2013) ⇒ empirical likelihood

Distributions of Positive Economic Variables: Two Stylized Facts



How Do We Estimate the Jump-Size Magnitude?

- Our estimation strategy is based 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.

- Estimating the difference between left and right limits of the density at a discontinuity point takes the following steps:
  1. Split the gamma kernel into two parts at the discontinuity point.
  2. Make each part a legitimate kernel by re-normalization.
  3. Estimate the left and right limits of the density by corresponding truncated kernels.

- Why not employ the beta (Chen, 1999) and gamma kernels for the left and right limits of the density, instead?
  - The variance convergences of the beta and gamma estimates in the vicinity of the discontinuity point slow down to $O\left(n^{-1}b^{-1}\right)$ [$\Rightarrow$ power loss].
    - It is hard (or even impossible) to improve the rate to usual $O\left(n^{-1}b^{-1/2}\right)$.
  - The estimates using the truncated kernels have the usual variance convergence of $O\left(n^{-1}b^{-1/2}\right)$.
    - Their bias convergences in turns slow down to $O\left(b^{1/2}\right)$ [$\Rightarrow$ size distortion].
    - The rate can be improved to usual $O\left(b\right)$ with no additional conditions (e.g., extra smoothness in the density), by means of a multiplicative bias correction method.

Contributions

1. Our proposal is easy to implement.
   - Unlike BLL by McCrary (2008), our approach always generates nonnegative density estimates and is free from choosing bin widths.
   - Unlike EL by Otsu, Xu and Matsushita (2013), nonlinear optimization is unnecessary.
   - Standard statistical packages including GAUSS, Matlab and R prepare a command that can return values of incomplete gamma functions.
2. Our proof strategies based on a few different approximation techniques for incomplete gamma functions are new to the literature.
3. Estimation theory of the entire density in the presence of a discontinuity point is also presented.

Plan of Talk

1. Estimation and Inference for Discontinuity in the Density
2. Smoothing Parameter Selection
3. Estimation of the Entire Density in the Presence of a Discontinuity Point
4. Finite-Sample Performance
5. Empirical Illustration
6. Conclusion

1 Estimation and Inference for Discontinuity in the Density

- Let
$$f_{-}\left(c\right):=\lim_{x\uparrow c}f\left(x\right)\ \text{and}\ f_{+}\left(c\right):=\lim_{x\downarrow c}f\left(x\right),$$
be the lower and upper limits of the pdf at a given point $x=c$, respectively.
- The parameter of interest is the jump-size magnitude of the density at c
$$J\left(c\right):=f_{+}\left(c\right)-f_{-}\left(c\right).$$
- To check whether f is (dis)continuous at c, we would do the followings:
  1. Estimate $J\left(c\right)$ nonparametrically.
  2. Test for the null of continuity of f at c, i.e., $H_0: J\left(c\right)=0$, against the two-sided alternative.

1.1 Estimation Strategy for Two Limits of the Density

1. Split the gamma kernel into two parts at c, namely,

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

where

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

- An entire univariate random sample $\{X_i\}_{i=1}^{n}$ is also split into two sub-samples

$$\left\{X_i^-\right\} := \{X_i : X_i < c\} \text{ and } \left\{X_i^+\right\} := \{X_i : X_i \ge c\}.$$

2. Make scale-adjustments to obtain the re-normalized truncated kernels

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

where

$$\gamma(a,z) = \int_0^z t^{a-1}\exp(-t)\,dt \text{ and } \Gamma(a,z) = \int_z^\infty t^{a-1}\exp(-t)\,dt$$

for $a, z > 0$ are **the lower and upper incomplete gamma functions**.

3. Estimate $f_-(c)$ and $f_+(c)$ as

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

1.2 Regularity Conditions

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 $\mathbb{R}_+ \setminus \{c\}$. Also let

$$f_-^{(j)}(c) := \lim_{x \uparrow c}\frac{d^j f(x)}{dx^j} \text{ and } f_+^{(j)}(c) := \lim_{x \downarrow c}\frac{d^j f(x)}{dx^j}$$

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

$$b + \frac{1}{nb} \to 0$$

as $n \to \infty$.

1.3 Bias-Corrected Estimation of the Jump-Size

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

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

Remark 1.

- The bias convergence of $O\left(b^{1/2}\right)$ is due to one-sided smoothing.
  - 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.

Remark 2.

- Proposition 1 can be demonstrated by combining the following formulae:

Stirling's formula.

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

Recursive formulae on incomplete gamma functions.

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

Series expansion of the lower incomplete gamma function by Pagurova (1965) and Temme (1979).

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

- The bias convergence in estimators of $f_{\pm}(c)$ can be improved to $O(b)$ via the multiplicative bias correction ("MBC") technique by Terrell and Scott (1980).
  - The MBC method has been applied to asymmetric kernel density estimation (e.g., Hirukawa, 2010; Hirukawa and Sakudo, 2014, 2015; Funke and Kawka, 2015).

- 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.

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

$$\begin{aligned}Bias\left\{\tilde{f}_{\pm}(c)\right\} &\sim \left(\frac{1}{\delta^{1/2}}\right)\left[\frac{c}{\pi}\left\{\frac{\left(f_{\pm}^{(1)}(c)\right)^{2}}{f_{\pm}(c)}\right\}\right.\\ &\quad\left.-\left\{\left(1-\frac{4}{3\pi}\right)f_{\pm}^{(1)}(c)+\frac{c}{2}f_{\pm}^{(2)}(c)\right\}\right]b,\text{ and}\\ Var\left\{\tilde{f}_{\pm}(c)\right\} &\sim \frac{1}{nb^{1/2}}\lambda(\delta)\frac{f_{\pm}(c)}{\sqrt{\pi}c^{1/2}},\end{aligned}$$

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

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

Theorem 1. *It holds for $\tilde{J}(c):=\tilde{f}_{+}(c)-\tilde{f}_{-}(c)$ that under Assumptions 1-3, as $n\to\infty$,*

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

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

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

1.4 Test Statistic for the (Dis)Continuity of the Density

- 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 test statistic is

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

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 $B_n = o\left(n^{1/2}b^{1/4}\right)$.

- How to estimate $V(c)$?

  Method 1. Replacing $f_{\pm}(c)$ in $V(c)$ with their consistent estimates $\tilde{f}_{\pm}(c)$ immediately yields

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

  Method 2. It follows from

$$\begin{aligned} \hat{f}(c) &= \frac{1}{n}\sum_{i=1}^{n} K_{G(x,b)}(X_i)\Big|_{x=c} = \frac{1}{n}\sum_{i=1}^{n} K_{G(c,b)}(X_i) \\ &= \frac{\gamma(c/b+1, c/b)}{\Gamma(c/b+1)}\hat{f}_-(c) + \frac{\Gamma(c/b+1, c/b)}{\Gamma(c/b+1)}\hat{f}_+(c) \\ &\xrightarrow{p} \frac{f_+(c) + f_-(c)}{2} \end{aligned}$$

  that 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 Smoothing Parameter Selection

- Based on our preference for test-optimality, we adopt the power-optimality criterion by Kulasekera and Wang (1998).
  - All previous proposals on bandwidth selection in the literature on RDD (e.g., McCrary, 2008; Imbens and Kalyanaraman, 2012; Porter and Yu, 2015) stand on the idea of estimation-optimality.

- The entire sample $\{X_i\}_{i=1}^{n} = \left\{\left\{X_i^-\right\}_{i=1}^{n_-}, \left\{X_i^+\right\}_{i=1}^{n_+}\right\}$ 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

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

- The test statistic using the mth sub-sample $\left\{X_{m,i}\right\}_{i=1}^{k}$ 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.
  - 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 $0 < \underline{B} < \overline{B} < \infty$.
  - 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 smoothing parameter value

$$\hat{b}_k = \hat{B}k^{-q} = \arg\max_{b_k \in H_k} \hat{\pi}_M(b_k) \Rightarrow \hat{b}_n := \hat{B}n^{-q}.$$

Selection Procedure in Practice.

Step 1: Choose some $p \in (0,1)$ and specify $M = \left\lfloor \min\left\{n_-^p, n_+^p\right\}\right\rfloor$.

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 $c_m(\alpha) \equiv z_{\alpha/2}$ (i.e., $\Pr\left\{N(0,1) > z_{\alpha/2}\right\} = \alpha/2$) and find $\hat{b}_k = \inf\left\{\arg\max_{b_k \in H_k} \hat{\pi}_M(b_k)\right\}$ 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 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.
- How should we estimate the entire density if the test rejects the null of continuity of the pdf f at the cutoff c?
  - 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.

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

$$Bias\left\{\hat{f}_+(x)\right\} \sim \left\{f^{(1)}(x) + \frac{x}{2}f^{(2)}(x)\right\}b, \text{ and}$$

$$Var\left\{\hat{f}_+(x)\right\} \sim \frac{1}{nb^{1/2}}\frac{f(x)}{2\sqrt{\pi}x^{1/2}}.$$

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

$$Bias\left\{\hat{f}_-(x)\right\} \sim \left\{f^{(1)}(x) + \frac{x}{2}f^{(2)}\right\}b, \text{ and}$$

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

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

Theorem 3. *If $f(x)$ is unbounded at $x = 0$, Assumptions 1 holds and $b + \left(nb^2\right)^{-1} \to 0$ 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*

$$b + \frac{1}{nb^2 f(x)} \to 0$$

as $n \to \infty$ and $x \to 0$, then

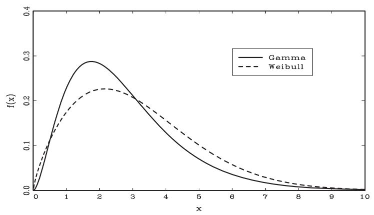
$$\left|\frac{\hat{f}_-(x) - f(x)}{f(x)}\right| \xrightarrow{p} 0$$

as $x \to 0$.

4 Finite-Sample Performance

4.1 Jump-Size Estimation

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



- Jump-size estimators:
  1. McCrary's (2008) BLL estimator $\hat{J}_M(c)$ using the triangular kernel $K(u) = (1 - |u|) \mathbf{1}(|u| \leq 1)$ and the bandwidth choice method therein.
  2. Our estimator $\tilde{J}(c)$ with the smoothing parameter b selected by the power-optimality criterion for two test statistics $T_i(c) := \frac{\sqrt{nb^{1/2}}\tilde{J}(c)}{\sqrt{\tilde{V}_i(c)}}, i = 1, 2$, where:
     (a) all critical values in $\hat{\pi}_M(b_k)$ are set equal to $z_{0.025} = 1.96$;
     (b) (p, q) are predetermined by $(p, q) = (1/2, 4/9)$;
     (c) the interval for b_k is $H_k = [0.05, 0.50]$; and
     (d) the mixing exponent $\delta \in \{0.49, 0.64, 0.81\}$, 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.

Result ($n = 2000$; $\#\{$**replications**$\} = 1000$).

			Estimator			
				$\tilde{J}(c)$ with δ		
Distribution	c		$\hat{J}_M(c)$	0.49	0.64	0.81
Gamma	1.7057	Bias	-0.0283	0.0006	0.0002	-0.0000
	(30%)	StdDev	0.0250	0.0430	0.0445	0.0458
		RMSE	0.0377	0.0430	0.0445	0.0458
	2.4248	Bias	-0.0240	-0.0004	-0.0004	-0.0004
	(Med)	StdDev	0.0271	0.0351	0.0363	0.0374
		RMSE	0.0362	0.0351	0.0363	0.0374
Weibull	1.9419	Bias	-0.0144	0.0017	0.0003	0.0001
	(30%)	StdDev	0.0225	0.0367	0.0372	0.0383
		RMSE	0.0267	0.0367	0.0372	0.0383
	2.8386	Bias	-0.0149	0.0007	0.0005	0.0004
	(Med)	StdDev	0.0218	0.0299	0.0309	0.0319
		RMSE	0.0264	0.0299	0.0309	0.0319

4.2 Testing for Discontinuity

- 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.
- Denote the measure of discontinuity as
$$d := \Pr(X \leq c) - \gamma,$$
where
$$d \in \{0.00, 0.02, 0.04, 0.06, 0.08, 0.10\}.$$
  - $d > 0$ ($\Leftrightarrow J(c) > 0$) suggests a jump of the pdf at c.

Result A: Size $(n = 2000; \#\{\textbf{replications}\} = 1000)$.

(%)

Distribution	c	Nominal	$T_1(c)$ with δ 0.49	0.64	0.81	$T_2(c)$ with δ 0.49	0.64	0.81
Gamma	1.7057	5%	3.5	3.6	3.7	4.2	3.9	3.9
	(30%)	10%	8.1	8.2	8.4	8.8	8.5	8.7
	2.4248	5%	4.7	4.7	4.8	4.9	5.0	5.1
	(Med)	10%	8.8	8.9	9.0	9.4	9.4	9.5
Weibull	1.9419	5%	3.8	3.7	3.8	7.7	4.4	4.0
	(30%)	10%	8.3	8.4	8.3	12.4	9.0	8.5
	2.8386	5%	4.7	4.7	4.8	4.9	5.0	5.0
	(Med)	10%	8.9	9.0	9.2	9.4	9.4	9.6

Result B: Power $(n = 2000; \#\{\textbf{replications}\} = 1000; \delta = 0.81)$.

(%)

Distribution	c	Test	Nominal	d 0.00	0.02	0.04	0.06	0.08	0.10
Gamma	1.7057	$T_M(c)$	5%	9.9	1.6	11.7	52.1	90.1	99.5
	(30%)		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	$T_M(c)$	5%	12.0	3.6	7.1	23.9	55.6	83.9
	(Med)		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
Weibull	1.9419	$T_M(c)$	5%	4.5	3.1	18.2	53.7	84.9	97.5
	(30%)		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	$T_M(c)$	5%	6.6	3.1	10.0	31.8	63.9	87.6
	(Med)		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

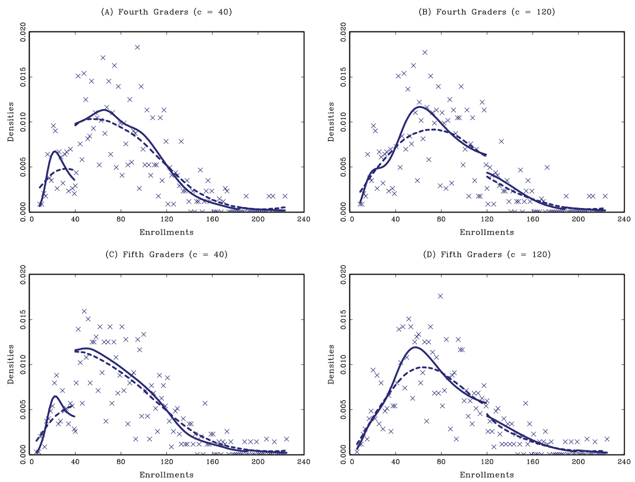
5 Empirical Illustration

- We reexamine the data sets on fourth and fifth graders of Israeli elementary schools used by Angrist and Lavy (1999).
- 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.
- 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.
  - $T_2(c)$ with $\delta = 0.81$ is chosen as our test statistic.

Estimation and Testing Results.

n	c	Binned Local Linear Method $\hat{f}^M_-(c)$	$\hat{f}^M_+(c)$	$\hat{J}_M(c)$	$T_M(c)$	Truncated Kernel Method $\tilde{f}_-(c)$	$\tilde{f}_+(c)$	$\tilde{J}(c)$	$T_2(c)$
					(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

Density Estimates.



6 Conclusion

- New estimation and testing procedures on discontinuity in densities with positive support are developed.
  - 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).
  - Two versions of test statistics for the null of continuity at a given point are proposed.
    - ∗ A smoothing parameter selection method under the power-optimality criterion is tailored to our testing procedure.
  - Estimation theory of the entire density in the presence of a discontinuity point is provided.

- Monte Carlo simulations indicate the followings:
  - The jump-size estimator is nearly unbiased when there is no jump in the true density.
  - 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.