Taisuke Otsu, Ke-Li Xu and Yukitoshi Matsushita

Estimation and inference of discontinuity in density

Article (Accepted version)
(Refereed)

Original citation:
Otsu, Taisuke, Xu, Ke-Li and Matsushita, Yukitoshi (2013). *Estimation and inference of discontinuity in density*. Journal of Business and Economic Statistics 31, (4) pp. 507-524 ISSN 0735-0015

DOI: 10.1080/07350015.2013.818007

Estimation and Inference of Discontinuity in Density

Taisuke Otsu[*] London School of Economics

Ke-Li Xu[†] Texas A&M University

Yukitoshi Matsushita[‡] University of Tsukuba

June 2013

Abstract

Continuity or discontinuity of probability density functions of data often plays a fundamental role in empirical economic analysis. For example, for identification and inference of causal effects in regression discontinuity designs it is typically assumed that the density function of a conditioning variable is continuous at a cutoff point that determines assignment of a treatment. Also, discontinuity in density functions can be a parameter of economic interest, such as in analysis of bunching behaviors of taxpayers. In order to facilitate researchers to conduct valid inference for these problems, this paper extends the binning and local likelihood approaches to estimate discontinuity of density functions and proposes empirical likelihood-based tests and confidence sets for the discontinuity. In contrast to the conventional Wald-type test and confidence set using the binning estimator, our empirical likelihood-based methods (i) circumvent asymptotic variance estimation to construct the test statistics and confidence sets; (ii) are invariant to nonlinear transformations of the parameters of interest; (iii) offer confidence sets whose shapes are automatically determined by data; and (iv) admit higher-order refinements, so-called Bartlett corrections. First- and second-order asymptotic theories are developed. Simulations demonstrate the superior finite sample behaviors of the proposed methods. In an empirical application, we assess the identifying assumption of no manipulation of class sizes in the regression discontinuity design studied by Angrist and Lavy (1999).

Keywords: Discontinuity in density; Empirical likelihood; Local likelihood; Nonparametric inference; Regression Discontinuity design; Bartlett correction.

JEL classification: C14; C21.

[*]Department of Economics, London School of Economics and Political Science. *Mailing address*: Houghton Street, London, WC2A 2AE, UK. *E-mail address*: t.otsu@lse.ac.uk.

[†]Department of Economics, Texas A&M University. *Mailing address*: 3063 Allen, 4228 TAMU, Department of Economics, College Station, TX 77843, USA. *E-mail address*: keli.xu@econmail.tamu.edu.

[‡]Graduate School of Humanities and Social Science, University of Tsukuba. *Mailing address*: 1-1-1, Tennoudai, Tsukuba, Ibaraki 305-8571, Japan. *E-mail address*: matsushita@dpipe.tsukuba.ac.jp.

1 Introduction

Continuity or discontinuity of probability density functions of data often play a fundamental role in empirical economic analysis. For example, for identification and inference of causal effects in regression discontinuity designs it is typically assumed that the density function of the conditioning variable is continuous at a cutoff point of interest (see, e.g., Hahn, Todd and van der Klaauw, 2001, Porter, 2003, Imbens and Lemieux, 2008, Lee, 2008, and McCrary, 2008). Given the continuity (or no manipulation) of the conditioning variable, discontinuity of the conditional mean function enables us to identify a local average treatment effect. Also, discontinuity (i.e. spread at a given discontinuity point) in the density function can be a parameter of economic interest. For example, Saez (2010) investigated bunching behaviors of taxpayers at kinked points in the US income tax schedule. In this case, the discontinuity of the income density becomes an economic parameter of interest and is used to derive the compensated reported income elasticity with respect to the marginal tax rate. For these empirical problems, effective estimation and inference of such (dis)continuities in the density functions are of central importance, which are the focus of the current paper.

This paper makes two contributions for inference problems on (dis)continuities of densities. First, we suggest a nonparametric estimator for discontinuities of densities based on the local likelihood approach (Loader, 1996, and Hjort and Jones, 1996). In the literature, McCrary (2008) proposed to estimate discontinuities by applying a local polynomial regression method for binned data (Cheng, 1994, 1997). Like Cheng and McCrary's local linear binning estimator, the proposed local likelihood estimator shares attractive performance in the presence of edge effects (i.e., estimation bias of densities at boundary points), which is crucial in the current setup. On the other hand, unlike the local linear binning estimator, the local likelihood density estimator is guaranteed to be non-negative by construction and is free from choosing bin widths to create binned data. This non-negativity of the local likelihood estimator is important when we are interested in regions with low densities. Simulations demonstrate the superior finite sample behavior of the new estimator.

Second and more importantly, we provide a general framework for conducting inference on discontinuities of densities based on the idea of empirical likelihood. We construct empirical likelihood functions from the estimating equations of both the binning and local likelihood estimators, and propose empirical likelihood tests and confidence sets for discontinuities of densities. Our empirical likelihood approach has at least six attractive features. First, we do not need to specify parametric functional forms of density functions since we construct the empirical likelihood functions from the first-order conditions of the binning and local likelihood estimators. Second, we do not need to estimate the asymptotic variance which is required in the Wald or t-statistic of McCrary (2008). The asymptotic variance estimation is automatically incorporated in the construction of empirical likelihood (i.e., internally studentized) and the derived empirical likelihood statistics are asymptotically pivotal, having chi-square limiting distributions. Third, our empirical likelihood-based inference methods are invariant to the formulations of the parameter of interest. In contrast, the Wald statistic of McCrary (2008) depends on how the parameter of interest or null hypothesis is specified by the researcher. Fourth, the shapes of the empirical likelihood confidence sets are automatically determined by data. In contrast, the Wald-type confidence sets are restricted to be symmetric around the point estimates. Fifth, the empirical likelihood confidence sets are well defined even if the local linear binning estimator of McCrary (2008) yields negative estimates for the densities. Finally, our empirical likelihood tests admit higher-order refinements, so-called Bartlett corrections. Simulation results indicate that the empirical likelihood tests have accurate finite sample sizes, and are generally more powerful (especially those based on the local likelihood approach) than the Wald test.

Angrist and Lavy (1999) exploited the so-called Maimonides' rule, which stipulates that a class with more than 40 pupils should be split into two, as an exogenous source of variation in class size to identify the effects of class size on the scholastic achievement of pupils in Israel. An important assumption of their study is no manipulation of class size by parents. Evidence of such manipulation casts doubt on the identification strategy of the regression discontinuity design. Angrist and Lavy (1999) provided intuitive arguments that manipulation is unlikely to happen in Israel. In this paper, we statistically re-examine the assumption of no manipulation of class sizes by testing continuity of the enrollment density (i.e., density continuity of the running variable). Using the proposed local likelihood estimator and the associated empirical likelihood inference procedure, we find significant evidence of manipulation at the first multiple of 40 but not clearly at other multiples. The progressively smaller estimates and weaker evidence of discontinuity our empirical results discover at multiples of 40 coincides with the fact that the parents are more likely to selectively manipulate the class size as just above 40 because they could place their children in schools with smaller class sizes if the manipulation is successful, than they do as just above 80, 120, or 160. These findings are not shared by using McCrary's binning estimator and Wald test.

This paper also contributes to the rapidly growing literature on empirical likelihood (see Owen, 2001, for a review). In particular, we extend the empirical likelihood approach to the density discontinuity inference problem by incorporating local polynomial fitting techniques such as Fan and Gijbels (1996) and Loader (1996). We show that the empirical likelihood ratios for density discontinuities have an asymptotically chi-squared distribution. Therefore, we can still observe the Wilks phenomenon (Fan, Zhang and Zhang, 2001) in this nonparametric density discontinuity inference problem. Furthermore, we study second-order asymptotic properties of the empirical likelihood statistics and show that the empirical likelihood confidence tests admit Bartlett corrections in our setup. Since DiCiccio, Hall and Romano (1991), Bartlett correctability of empirical likelihood is repeatedly observed in the literature for various setups. We extend the Bartlett correctability result to the nonparametric inference problem of density discontinuity.

This paper is organized as follows. Section 2 presents our basic setup and point estimation methods. In Section 3.1 we construct empirical likelihood functions of discontinuities in densities. Sections 3.2 and 3.3 present the first- and second-order asymptotic properties of the empirical likelihood-based inference methods, respectively. The proposed methods are evaluated by Monte Carlo simulations in Section 4. An empirical application to validation of regression discontinuity design is provided in Section 5. Section 6 concludes. Appendix A contains mathematical proofs, lemmas, and derivations.

2 Setup and Estimation

We first introduce our basic setup. Let $\{X_i\}_{i=1}^n$ be an iid sample of $X \in \mathbb{X} \subseteq \mathbb{R}$ with the probability density function $f(x)$. Suppose that we are interested in (dis)continuity of the density $f(x)$ at some given point $c \in \mathbb{X}$. Let $f_l = \lim_{x\uparrow c} f(x) > 0$ and $f_r = \lim_{x\downarrow c} f(x) > 0$ be the left and right limits of $f(x)$ at $x = c$, respectively. Our object of interest is the difference of the left and right limits:

$$\theta_0 = f_r - f_l. \tag{1}$$

If $\theta_0 = 0$, then the density function $f(x)$ is continuous at $x = c$. We wish to estimate, construct a confidence set, and conduct a hypothesis test for the parameter θ_0.

First, let us consider the point estimation problem of θ_0. If the density is discontinuous at c (i.e., $\theta_0 \neq 0$), we can regard the estimation problems for the limits f_l and f_r as the ones for nonparametric densities at the boundary point c using sub-samples with $X_i < c$ and $X_i \geq c$. To reduce boundary bias in nonparametric density estimation, it is reasonable to apply a local polynomial fitting technique, which has favorable properties on boundaries (see, e.g., Fan and Gijbels, 1996). In density estimation we do not have any regressands or regressors. However, there are at least two ways to adapt the local polynomial fitting method to the density estimation problem, the binning and local likelihood methods.

The binning method (e.g., Cheng, 1994, 1997, and Cheng, Fan and Marron, 1997) creates regressands and regressors based on binned data and then implements local polynomial regression. Let $\left\{X_j^G\right\}_{j=1}^J = \left\{\ldots, c - \frac{3}{2}b, c - \frac{1}{2}b, c + \frac{1}{2}b, c + \frac{3}{2}b, \ldots\right\}$, which plays the role of a regressor, be an equi-spaced grid of width b, where the interval $\left[X_1^G, X_J^G\right]$ covers the support $\mathbb{X}$. Let $\mathbb{I}\{\cdot\}$ be the indicator function and $Z_j^G = \frac{1}{nb}\sum_{i=1}^n \mathbb{I}\left\{\left|X_i - X_j^G\right| < \frac{b}{2}\right\}$, which plays the role of a regressand, be the normalized frequency for the j-th bin. The bin-based local linear estimators $\hat{f}_l^G$ and $\hat{f}_r^G$ for f_l and f_r are defined as solutions to the following weighted least square problems with respect to a_l and a_r, respectively,

$$\min_{a_l, b_l} \sum_{j: X_j^G < c} \mathbb{K}\left(\frac{X_j^G - c}{h}\right)\left(Z_j^G - a_l - b_l\left(X_j^G - c\right)\right)^2, \tag{2}$$

$$\min_{a_r, b_r} \sum_{j: X_j^G \geq c} \mathbb{K}\left(\frac{X_j^G - c}{h}\right)\left(Z_j^G - a_r - b_r\left(X_j^G - c\right)\right)^2,$$

where $\mathbb{K}(\cdot)$ is a symmetric kernel function and h is a bandwidth parameter. We may add higher-order polynomials of $\left(X_j^G - c\right)$ in the regressors to further reduce the bias or to estimate higher-order derivatives of f. Based on these regressions, the parameter θ_0 can be estimated by $\hat{\theta}^G = \hat{f}_r^G - \hat{f}_l^G$. Note that we need to choose two tuning parameters, b and h to compute $\hat{\theta}^G$. This estimator is adopted by

McCrary (2008) to conduct the Wald test for the density continuity hypothesis $H_0 : \theta_0 = 0$. See the papers cited above for the properties of the bin-based density estimators.

As an alternative estimation method, we adapt the local likelihood approach (e.g., Copas, 1995, Hjort and Jones, 1996, and Loader, 1996) to our context. The local likelihood method constructs some localized versions of likelihood functions for f_l and f_r using kernel weights and then conducts likelihood maximization. Let $\hat{a}_l$ and $\hat{a}_r$ be solutions to the following maximization problems with respect to a_l and a_r, respectively,

$$\max_{a_l,b_l} \left\{ \frac{1}{n} \sum_{i:X_i<c} \mathbb{K}\left(\frac{X_i-c}{h}\right)(a_l + b_l(X_i - c)) - \int_{u<c} \mathbb{K}\left(\frac{u-c}{h}\right) \exp(a_l + b_l(u-c))\, du \right\}, \tag{3}$$
$$\max_{a_r,b_r} \left\{ \frac{1}{n} \sum_{i:X_i\geq c} \mathbb{K}\left(\frac{X_i-c}{h}\right)(a_r + b_r(X_i - c)) - \int_{u\geq c} \mathbb{K}\left(\frac{u-c}{h}\right) \exp(a_r + b_r(u-c))\, du \right\}.$$

The local (linear) likelihood estimators for the density limits f_l and f_r are defined as $\hat{f}_l = \exp(\hat{a}_l)$ and $\hat{f}_r = \exp(\hat{a}_r)$, respectively. The discontinuity parameter θ_0 is estimated by $\hat{\theta} = \hat{f}_r - \hat{f}_l$. Higher-order polynomials of $(X_i - c)$ and $(u - c)$ may be added to the linear terms.[1] In contrast to the bin-based estimator, the local likelihood estimator for densities is always positive by construction. This feature of the local likelihood estimator is attractive particularly if we are interested in low density regions to avoid negative density estimates, which are logically inconsistent.[2] By adapting the arguments in Loader (1996, Lemma 1 and Theorem 2) to boundary points and applying the delta method, the asymptotic distribution of $\hat{\theta}$ is obtained as follows.

Theorem 2.1. *Suppose that Assumptions 1-3 and 5 in Section 3.2 hold. Also assume $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Then*

$$\sqrt{nh}\left(\hat{\theta} - \theta - B_L\right) \xrightarrow{d} N(0, V_L),$$

where $B_L = h^2 (f_r'' - f_l'') \frac{K_{r21}^2 - K_{r31}K_{r11}}{2\left(K_{r21}K_{r01} - K_{r11}^2\right)}$, $V_L = (f_r + f_l) \frac{K_{r21}^2 K_{r02} - 2K_{r21}K_{r11}K_{r12} + K_{r11}^2 K_{r22}}{\left(K_{r21}K_{r01} - K_{r11}^2\right)^2}$, $f_l'' = \lim_{x\uparrow c} \frac{d^2 \log f(x)}{dx^2}$, $f_r'' = \lim_{x\downarrow c} \frac{d^2 \log f(x)}{dx^2}$, *and $K_{rj_1j_2}$ is defined in (12) in Appendix A.*

Note that the bias term B_L is of order h^2 and cancels out if the density has a continuous second-order derivative (i.e., $f_r'' = f_l''$).

Inference on possibly discontinuous density functions has been considered in the literature of nonparametric statistics (e.g., Cline and Hart, 1991, Marron and Ruppert, 1994, and Cheng, Fan and Marron, 1997). However, interests in the spread of densities at discontinuity points are not motivated

[1] In general, using a parametric function $\psi(\cdot, \pi_l)$, the local likelihood estimator for f_l can be defined by transforming $\hat{\pi}_l = \arg\max_{\pi_l} \left\{ \frac{1}{n} \sum_{i:X_i<c} \mathbb{K}\left(\frac{X_i-c}{h}\right) \log \psi(X_i, \pi_l) - \int_{u<c} \mathbb{K}\left(\frac{u-c}{h}\right) \psi(u, \pi_l)\, du \right\}$. An estimator for f_r is defined in the same manner. It reduces to the estimator defined by (3) when $\psi(x, \pi_l) = \exp(a_l + b_l(x - c))$ with $\pi_l = (a_l, b_l)'$. Note that when ψ is a constant, we obtain the (normalized) kernel density estimator $\frac{2}{nh} \sum_{i:X_i<c} \mathbb{K}\left(\frac{X_i-c}{h}\right)$. In this sense, to implement local likelihood estimation, we need to choose the bandwidth h and parametric function ψ. See Hjort and Jones (1996) and Park, Kim and Jones (2002) for comparisons of difference choices of ψ.

[2] See, Hjort and Jones (1996, Section 6) and Loader (1996, Section 5) for asymptotic analyses of the local likelihood estimators at boundaries and in tails, respectively.

until recently. McCrary (2008) considered the testing problem for the density continuity in the context of regression discontinuity designs.[3] McCrary (2008) formulated the density continuity testing problem as

$$H_0 : \log f_l = \log f_r, \quad H_1 : \log f_l \neq \log f_r,$$

and suggested the t-test statistic based on the binning estimator:

$$t^G = \frac{\sqrt{nh}\left(\log \hat{f}_r^G - \log \hat{f}_l^G\right)}{\hat{\sigma}_{\mathbb{K}}}, \tag{4}$$

where $\hat{\sigma}_{\mathbb{K}}^2$ is a consistent estimator for the asymptotic variance of the numerator $\sqrt{nh}\left(\log \hat{f}_r^G - \log \hat{f}_l^G\right)$. Using the triangle kernel function $\mathbb{K}(a) = \max\{0, 1 - |a|\}$, McCrary (2008) showed that the numerator $\sqrt{nh}\left(\log \hat{f}_r^G - \log \hat{f}_l^G - B\right)$ converges in distribution to $N\left(0, \sigma_{\mathbb{K}}^2\right)$ with $B = h^2 \frac{1}{20}\left(\frac{f_l''}{f_l} - \frac{f_r''}{f_r}\right)$ and $\sigma_{\mathbb{K}}^2 = \frac{24}{5}\left(\frac{1}{f_l} + \frac{1}{f_r}\right)$. Thus, by undersmoothing (i.e., $\lim_{n\to\infty} h^2\sqrt{nh} = 0$) to neglect the bias term and estimating the standard error $\sigma_{\mathbb{K}}$ by $\hat{\sigma}_{\mathbb{K}} = \sqrt{\frac{24}{5}\left(\frac{1}{\hat{f}_l^G} + \frac{1}{\hat{f}_r^G}\right)}$, we can test H_0 by the t-statistic t^G using the standard normal critical values.

There are at least four issues with McCrary's (2008) Wald-type approach. First, since the asymptotic variance $\sigma_{\mathbb{K}}^2$ and its estimator $\hat{\sigma}_{\mathbb{K}}^2$ depend on the form of the kernel function $\mathbb{K}$, we need to find the formula and estimator of $\sigma_{\mathbb{K}}^2$ for each choice of $\mathbb{K}$. Second, the local linear estimator based on a non-negative sample may produce negative estimates at some design points (Xu and Phillips, 2011). When this happens to either $\hat{f}_l^G$ or $\hat{f}_r^G$, McCrary's (2008) statistic t^G cannot be used. Third, since the test statistic t^G is constructed essentially to test the log difference $\log f_l - \log f_r$, it does not automatically generate a confidence set for $\theta_0 = f_r - f_l$. Finally, although the above Wald or t-test can be modified to test the null hypothesis $\tilde{H}_0 : f_l = f_r$, the Wald test statistic for H_0 and $\tilde{H}_0$ will take different values in finite samples (i.e., lack of invariance to nonlinear hypotheses, see, e.g., Gregory and Veal, 1985). Similar comments apply to the Wald test based on the local likelihood estimator $\hat{\theta}$. To address these issues we propose a new framework for inference of θ_0 in the following section.

3 Empirical Likelihood Inference

3.1 Construction of Test Statistics

In this subsection, we construct empirical likelihood functions for the parameter of interest $\theta_0 = f_r - f_l$ based on the estimation approaches presented in the last section. We first consider the binning approach.

[3] As McCrary (2008, p. 701) argued, continuous density of a running variable is neither necessary nor sufficient to identify a causal parameter of interest without auxiliary assumptions. As a specific example, let us consider the framework of Lee (2008, Propositions 2 and 3), where the running variable is under agent's control but may contain some idiosyncratic component. In Lee's (2008) setup, if the density of the running variable is discontinuous, then we cannot identify even the signs of some weighted average treatment effects (i.e., "ATE^*" and "ATE^{**}" in Lee, 2008) without imposing additional assumptions.

Let $I_j^G = \mathbb{I}\left\{X_j^G \geq c\right\}$ and $X_{j,h}^G = \frac{X_j^G - c}{h}$. The bin-based local linear estimators $\hat{f}_l^G$ and $\hat{f}_r^G$ defined in (2) satisfy the first-order conditions (see, p. 20 of Fan and Gijbels, 1996)

$$\sum_{j=1}^{J} \left(1 - I_j^G\right) K_{lj}^G \left(Z_j^G - \hat{f}_l^G\right) = 0, \quad \sum_{j=1}^{J} I_j^G K_{rj}^G \left(Z_j^G - \hat{f}_r^G\right) = 0, \tag{5}$$

where

$$\begin{aligned} K_{lj}^G &= \mathbb{K}\left(X_{j,h}^G\right) \left\{ \sum_{k=1}^{J} \left(1 - I_k^G\right) \mathbb{K}\left(X_{k,h}^G\right) \left(X_{k,h}^G\right)^2 - \left(X_{j,h}^G\right) \sum_{k=1}^{J} \left(1 - I_k^G\right) \mathbb{K}\left(X_{k,h}^G\right) X_{k,h}^G \right\}, \\ K_{rj}^G &= \mathbb{K}\left(X_{j,h}^G\right) \left\{ \sum_{k=1}^{J} I_j^G \mathbb{K}\left(X_{k,h}^G\right) \left(X_{k,h}^G\right)^2 - X_{j,h}^G \sum_{k=1}^{J} I_k^G \mathbb{K}\left(X_{k,h}^G\right) X_{k,h}^G \right\}. \end{aligned}$$

If we regard (5) as estimating equations or sample moment conditions for $\left(E\left[\hat{f}_l^G\right], E\left[\hat{f}_r^G\right]\right)$, the bin-based empirical likelihood function for $\left(E\left[\hat{f}_l^G\right], E\left[\hat{f}_r^G\right]\right)$ is constructed as

$$L^G\left(a_l, a_r\right) = \sup_{\{p_j\}_{j=1}^{J}} \prod_{j=1}^{J} p_j, \tag{6}$$

$$\text{s.t.} \quad 0 \leq p_j \leq 1, \ \sum_{j=1}^{J} p_j = 1, \ \sum_{j=1}^{J} p_j \left(1 - I_j^G\right) K_{lj}^G \left(Z_j^G - a_l\right) = 0, \ \sum_{j=1}^{J} p_j I_j^G K_{rj}^G \left(Z_j^G - a_r\right) = 0.$$

The weight p_j can be interpreted as probability mass allocated to the observed value of Z_j^G. By applying the Lagrange multiplier method, under certain regularity conditions (see, Theorem 2.2 of Newey and Smith, 2004), we can obtain the dual representation of the maximization problem in (6), that is

$$\ell^G\left(a_l, a_r\right) = -2\left\{\log L^G\left(a_l, a_r\right) + n \log n\right\} = 2 \sup_{\lambda^G \in \Lambda_J^G(a_l, a_r)} \sum_{j=1}^{J} \log\left(1 + \lambda^{G\prime} g_j^G\left(a_l, a_r\right)\right), \tag{7}$$

where $\Lambda_J^G\left(a_l, a_r\right) = \left\{\lambda^G \in \mathbb{R}^2 : \lambda^{G\prime} g_j^G\left(a_l, a_r\right) \in \mathcal{V}^G \text{ for } j = 1, \ldots, J\right\}$, $\mathcal{V}^G$ is an open interval containing 0, and $g_j^G\left(a_l, a_r\right) = \left[\left(1 - I_j^G\right) K_{lj}^G \left(Z_j^G - a_l\right), I_j^G K_{rj}^G \left(Z_j^G - a_r\right)\right]'$. Note that the J-variable maximization problem in (6) with respect to $\{p_j\}_{j=1}^{J}$ reduces to the two-variable convex maximization problem in (7) with respect to λ^G, which is easily implemented by a Newton-type optimization algorithm. Therefore, in practice we use the dual formulation (7) to compute the (log) empirical likelihood function. Based on the empirical likelihood function $\ell^G\left(a_l, a_r\right)$, the concentrated likelihood function for the parameter of interest $\theta_0 = f_r - f_l$ is defined as

$$\ell^G(\theta) = \min_{\{(a_l, a_r) \in \mathcal{A}_l \times \mathcal{A}_r : \theta = a_r - a_l\}} \ell^G\left(a_l, a_r\right), \tag{8}$$

for the parameter space $\mathcal{A}_l \times \mathcal{A}_r$ of (f_l, f_r).

We now define the empirical likelihood function based on the local likelihood approach. Let $I_i = \mathbb{I}\{X_i \geq c\}$ and $X_{i,h} = \frac{X_i - c}{h}$. The first-order conditions for the local likelihood maximization problems in (3) are written as

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}(1, X_{i,h})(1-I_i)\,\mathbb{K}(X_{i,h}) - \int_{x<c}\left(1, \frac{x-c}{h}\right)\mathbb{K}\left(\frac{x-c}{h}\right)\exp\left(a_l + b_l(x-c)\right)dx, \\
0 &= \frac{1}{n}\sum_{i=1}^{n}(1, X_{i,h})\,I_i\mathbb{K}(X_{i,h}) - \int_{x\geq c}\left(1, \frac{x-c}{h}\right)\mathbb{K}\left(\frac{x-c}{h}\right)\exp\left(a_r + b_r(x-c)\right)dx.
\end{aligned}
$$

Based on these estimating equations, the empirical likelihood function is constructed as

$$
L(a_l, a_r, b_l, b_r) = \sup_{\{p_i\}_{i=1}^{n}} \prod_{i=1}^{n} p_i, \tag{9}
$$

$$
\text{s.t.} \quad 0 \leq p_i \leq 1, \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i g_i(a_l, a_r, b_l, b_r) = 0,
$$

where $g_i(a_l, a_r, b_l, b_r) = \begin{pmatrix} (1, X_{i,h})'(1-I_i)\,\mathbb{K}(X_{i,h}) - \int_{x<c}\left(1, \frac{x-c}{h}\right)'\mathbb{K}\left(\frac{x-c}{h}\right)\exp\left(a_l + b_l(x-c)\right)dx, \\ (1, X_{i,h})'\,I_i\mathbb{K}(X_{i,h}) - \int_{x\geq c}\left(1, \frac{x-c}{h}\right)'\mathbb{K}\left(\frac{x-c}{h}\right)\exp\left(a_r + b_r(x-c)\right)dx \end{pmatrix}$.
The weight p_i can be interpreted as probability mass allocated to the observed value of X_i. The dual form of the empirical likelihood function (9) is

$$
\ell(a_l, a_r, b_l, b_r) = 2 \sup_{\lambda \in \Lambda_n(a_l, a_r, b_l, b_r)} \sum_{i=1}^{n} \log\left(1 + \lambda' g_i(a_l, a_r, b_l, b_r)\right), \tag{10}
$$

where $\Lambda_n(a_l, a_r, b_l, b_r) = \left\{\lambda \in \mathbb{R}^4 : \lambda' g_i(a_l, a_r, b_l, b_r) \in \mathcal{V} \text{ for } i = 1, \ldots, n\right\}$ and $\mathcal{V}$ is an open interval containing 0. Also, the concentrated likelihood function for the parameter of interest $\theta_0 = f_r - f_l$ is defined as

$$
\ell(\theta) = \min_{\{(a_l, a_r, b_l, b_r) \in \mathcal{A}_l \times \mathcal{A}_r \times \mathcal{B}_l \times \mathcal{B}_r : \theta = \exp(a_r) - \exp(a_l)\}} \ell(a_l, a_r, b_l, b_r), \tag{11}
$$

for some space $\mathcal{A}_l \times \mathcal{A}_r \times \mathcal{B}_l \times \mathcal{B}_r$ of (a_l, a_r, b_l, b_r).

Note that the constructions of the empirical likelihood functions $\ell^G(\theta)$ in (8) and $\ell(\theta)$ in (11) do not require any parametric functional form of the density function. Precisely, the above constructions give us the empirical likelihood functions for $E\left[\hat{f}_r\right] - E\left[\hat{f}_l\right]$, rather than for $\theta_0 = f_r - f_l$. However, by introducing undersmoothed bandwidths (specifically, letting $nh^5 \to 0$), we can asymptotically neglect the bias component $(f_r - f_l) - \left(E\left[\hat{f}_r\right] - E\left[\hat{f}_l\right]\right)$, and employ the functions $\ell^G(\theta)$ and $\ell(\theta)$ as valid empirical likelihood functions for the parameter θ_0.

One useful feature of the empirical likelihood approach is that it can easily incorporate additional information (see, Chen, 1997, in the context of density estimation). Suppose we have prior information about X specified in the form of $E[m(X)] = 0$, such as the mean, variance, or quantiles. Using the weights $\{p_i\}_{i=1}^{n}$, this information can be incorporated into the likelihood maximization problem (9) by adding the restriction $\sum_{i=1}^{n} p_i m(X_i) = 0$, and the dual form (10) is re-defined by adding $m(X_i)$ to the moment function $g_i(a_l, a_r, b_l, b_r)$. The resulting empirical likelihood inference is more efficient if additional information is valid.

3.2 First-order Asymptotic Properties

We now investigate the first-order asymptotic properties of the proposed empirical likelihood statistics. We impose the following assumptions.

Assumption.

1. *$\{X_i\}_{i=1}^n$ is i.i.d.*

2. *There exists a neighborhood $\mathcal{N}$ around c such that f is continuously second-order differentiable on $\mathcal{N} \setminus \{c\}$. For the matrices V and G defined in (12), V is positive definite and G has full column rank.*

3. *$\mathbb{K}$ is a symmetric and bounded density function with support $[-k, k]$ for some $k > 0$.*

4. *As $n \to \infty$, $h \to 0$, $nh \to \infty$, and $nh^5 \to 0$. Additionally, $b/h \to 0$ for $\ell^G(\theta)$ and $nh^3 \to \infty$ for $\ell(\theta)$.*

5. *For $\ell^G(\theta)$, $\mathcal{A}_l$ and $\mathcal{A}_r$ are compact, $f_l \in int(\mathcal{A}_l)$, and $f_r \in int(\mathcal{A}_r)$. For $\ell(\theta)$, $\mathcal{A}_l$, $\mathcal{A}_r$, $\mathcal{B}_l$, and $\mathcal{B}_r$ are compact, $\log f_l \in int(\mathcal{A}_l)$, $\log f_r \in int(\mathcal{A}_r)$, $f_l'/f_l \in int(\mathcal{B}_l)$, and $f_r'/f_r \in int(\mathcal{B}_r)$.*

Assumption 1 is on the data structure. Although it is beyond the scope of this paper, it would be interesting to extend the proposed method to weakly dependent data, where we are interested in the discontinuity of the stationary distribution. For this extension, we would need to introduce a blocking technique to handle the time series dependence in the moment functions (see, Kitamura, 1997). Assumption 2 restricts the local shape of the density function around $x = c$. This assumption allows discontinuity of the density function at $x = c$. Assumption 3 is on the kernel function $\mathbb{K}$ and implies the second-order kernel. This assumption is satisfied by e.g., the triangle kernel $\mathbb{K}(a) = \max\{0, 1 - |a|\}$ and Epanechnikov kernel $\mathbb{K}(a) = \frac{3}{4}\left(1 - a^2\right)\mathbb{I}\{|a| \leq 1\}$. Assumption 4 is on the bandwidth parameter h. The requirement $nh^5 \to 0$ corresponds to an undersmoothing condition to remove the bias component $(f_r - f_l) - \left(E\left[\hat{f}_r\right] - E\left[\hat{f}_l\right]\right)$ in the construction of empirical likelihood. The requirement $b/h \to 0$ is on the bin width b used for the binning estimator. The condition $nh^3 \to \infty$ is required to obtain the consistency of the minimizers to solve (11). If we do not have the local linear terms $(X_i - c)$ and $(u - c)$ in (3), this condition is unnecessary. Assumption 5 is similar to an assumption that would be used in a parametric estimation problem.

Under these assumptions, the first-order asymptotic distributions of the empirical likelihood functions $\ell^G(\theta)$ and $\ell(\theta)$ evaluated at the true parameter value $\theta = \theta_0$ are obtained as follows.

Theorem 3.1. *Under Assumptions 1-5, it holds*

$$\ell^G(\theta_0) \xrightarrow{d} \chi^2(1), \quad \ell(\theta_0) \xrightarrow{d} \chi^2(1).$$

Note that even in this nonparametric hypothesis testing problem, we observe the convergence of the empirical likelihood statistics to the pivotal χ^2 distribution, i.e., the Wilks phenomenon emerges. The null hypothesis $H_0 : \theta_0 = \theta$ for some given θ can be tested by the test statistic $\ell^G(\theta)$ or $\ell(\theta)$ with the $\chi^2(1)$ critical value. For example, the null of density continuity is tested by $\ell^G(0)$ or $\ell(0)$. Also, by inverting these test statistics, the $100(1-\xi)\%$ empirical likelihood asymptotic confidence sets are obtained as $CS^G = \{\theta : \ell^G(\theta) \leq c_\xi\}$ and $CS = \{\theta : \ell(\theta) \leq c_\xi\}$, where c_ξ is the $(1-\xi)$-th quantile of the $\chi^2(1)$ distribution.

We now compare our empirical likelihood approach with the Wald approach proposed by McCrary (2008). First, in contrast to the t-test based on (4), the empirical likelihood test based on $\ell^G(0)$ or $\ell(0)$ does not require any asymptotic variance estimation, which is automatically incorporated in the construction of the empirical likelihood function. Also, while the Wald test requires the derivation of the asymptotic variance $\sigma^2_{\mathbb{K}}$ for each kernel function, the empirical likelihood tests do not require such derivations. Second, the empirical likelihood confidence set CS^G by the binning method does not require the local linear (or polynomial) estimators of f_l and f_r. Thus, even if the binning estimate of f_l or f_r is negative in finite samples, the confidence set CS^G is still well defined. Third, the empirical likelihood test statistics are invariant to the formulation of the nonlinear null hypotheses. For example, to test the density continuity, we may specify the null hypothesis as $H_0 : \log f_l = \log f_r$, $\tilde{H}_0 : f_l = f_r$, $\bar{H}_0 : \frac{f_l}{f_r} = 1$, etc. For these hypotheses, the empirical likelihood test statistics are identical (i.e., $\ell^G(0)$ or $\ell(0)$). On the other hand, the Wald test statistic is not invariant to the formulation of the null hypotheses and may yield opposite conclusions in finite samples (see, e.g., Gregory and Veal, 1985).[4] [5]

3.3 Second-order Asymptotic Properties

In this subsection, we study the second-order asymptotic properties of the empirical likelihood statistics and confidence sets. For brevity, we only present the result for the empirical likelihood statistic based on the local likelihood approach. Similar results are available for the binning approach. To study the second-order properties of the empirical likelihood statistic $\ell(\theta)$ in (11), we adopt a similar approach

[4] An alternative inference approach is to employ some bootstrap method. The method can be applied to both the binning and local likelihood estimators. Our preliminary simulations show that a bootstrap-based test, where we estimate the standard error $\sigma^2_{\mathbb{K}}$ by bootstrapping, improves McCrary's (2008) t-test and has similar finite-sample size properties with the bin-based empirical likelihood test, however, it is much more numerically expensive to implement. In our experiments, when the bootstrap is applied to the binning estimator with 399 bootstrap replications, the test takes about ten times longer than the bin-based empirical likelihood test. To our best knowledge, theoretical properties of the bootstrap method are still unknown in the context of density discontinuity testing. Although it is beyond the scope of this paper, further investigation into the bootstrap method and comparisons with the empirical likelihood approach are definitely worthwhile.

[5] In addition to the Wald-type approach by McCrary (2008) and likelihood ratio-type approach by this paper, we can adopt the Lagrange multiplier- or score-type approach to test H_0 (see, Smith, 1997). In our context, the maximizers in (7) and (10) with respect to λ^G and λ, respectively, play the roles of the Lagrange multipliers. By adapting the result of Smith (1997) to our setup, the Lagrange multiplier test statistics can be constructed as quadratic forms of these maximizers, and are invariant to the formulation of the nonlinear null hypotheses.

to Chen and Cui (2006), which studied Bartlett correctability of the empirical likelihood statistic in the presence of nuisance parameters (a_l, b_l and b_r in our case). Note that while Chen and Cui (2006) considered moment functions for finite-dimensional parameters, our moment functions are used for the nonparametric object $\theta_0 = f_r - f_l$ and contain the bandwidth parameter h. Thus, although the basic idea of the second-order analysis follows from Chen and Cui (2006), technical details are different from theirs. Let c_ξ and $f_1(\cdot)$ be the $(1-\xi)$-th quantile and probability density function of the $\chi^2(1)$ distribution, respectively. The main results are summarized as follows. See Appendices A.3 and A.4 for technical details.

Theorem 3.2. *Suppose that Assumptions 1-6 hold. Furthermore, assume that $nh/\log n \to \infty$ as $n \to \infty$, and there exists a partition $-k = u_0 < u_1 < \cdots < u_m = k$ such that for each $j = 1, \ldots, m$, the derivative $\frac{d\mathbb{K}(u)}{du}$ is bounded and either strictly positive or strictly negative on (u_{j-1}, u_j). Then it holds*

$$(i) \quad \Pr\{\ell(\theta_0) \le c_\xi\} = 1 - \xi - c_\xi f_1(c_\xi) B_c + O\left((nh)^{-3/2} + (nh)^{-1} h^2 + h^4\right),$$
$$(ii) \quad \Pr\{\ell(\theta_0) \le c_\xi(1 + B_c)\} = 1 - \xi + O\left(n^2 h^{10} + (nh)^{-3/2} + (nh)^{-1} h^2 + h^4\right),$$

where the Bartlett factor B_c is defined in (25) of Appendix A.3.

Additional assumptions are required to establish an Edgeworth expansion (see, Section 5.5 of Hall, 1992).[6] Theorem 3.2 (i) says that the error in the null rejection probability of the empirical likelihood test for θ_0 using the critical value c_ξ based on the first-order $\chi^2(1)$ asymptotic distribution is of order $B_c = O\left(nh^5 + (nh)^{-1}\right)$. Theorem 3.2 (ii) says that the error in the null rejection probability can be reduced by modifying the critical value to $c_\xi(1 + B_c)$, so-called the Bartlett correction. For example, if $h \propto n^{-1/4}$, we have $\Pr\{\ell(\theta_0) \le c_\xi\} = 1 - \xi + O\left(n^{-1/4}\right)$ and $\Pr\{\ell(\theta_0) \le c_\xi(1 + B_c)\} = 1 - \xi + O\left(n^{-1/2}\right)$. In practice, the Bartlett factor B_c has to be estimated. The method of moments estimator of B_c can be obtained by substituting all the population moments involved by their corresponding sample moments. Chen and Cui (2006) suggested a bootstrap estimator for B_c.

Similar results are available for the bin-based empirical likelihood test statistic $\ell^G(\theta_0)$. In particular, the same statements in Theorem 3.2 hold with a different Bartlett factor.[7]

4 Simulations

In this section we study the finite-sample behaviors of the aforementioned methods using simulations. First we focus on the point estimators, i.e. the local linear binning estimator $\hat{\theta}^G$ and the local (log

[6] The additional assumption on the kernel function is introduced to guarantee a version of the Cramér condition. This assumption excludes the uniform kernel, for example. As discussed in Hall (1992, Lemma 5.6), if there is an interval where the kernel function becomes flat, then the current approach of the proof cannot guarantee the boundedness of some characteristic functions. As conjectured by Hall (1992, p. 270), it may be possible to relax this assumption by the method for lattice-valued random variables.

[7] Details are available from the authors upon request.

linear) likelihood estimator $\hat{\theta}$ for θ_0. For comparisons, we also consider the local constant binning estimator $\tilde{\theta}^G$ and the local (log constant) likelihood estimator $\tilde{\theta}$. For the kernel function $\mathbb{K}$, we use the triangle kernel function $\mathbb{K}(a) = \max\{0, 1 - |a|\}$. For the bandwidth h, we consider both fixed bandwidths $h = 1, 2, 3, 4$ and data-dependent bandwidths $h = \alpha h_{dd}$, where h_{dd} is the data-dependent bandwidth used by McCrary (2008) and $\alpha = 1.5^k$ for $k = -1, 0, 1, 2$.[8] For the bin size b to implement $\hat{\theta}^G$ and $\tilde{\theta}^G$, we employ a data-dependent method suggested by McCrary (2008). The data are generated from normal distribution $N(12, 3)$ (following McCrary, 2008) and Student's t distribution $12 + \frac{3}{\sqrt{5}} t(5)$. Both distributions have the same mean and variance. The sample size is $n = 1000$ and the suspected discontinuity point is $c = 13$. Since the above densities are continuous, the true value is $\theta_0 = 0$. The biases, variances and mean square errors (MSEs) of the above estimators are reported in Tables 1 and 2.

Among four estimators, $\hat{\theta}$ performs best in terms of MSEs. Its MSE is slightly smaller than that of its competitor, $\hat{\theta}^G$, when a small bandwidth is used, but is significantly smaller when the bandwidth is relatively large. The dominance of $\hat{\theta}$ mainly comes from its superior bias performance on boundaries, while its variance is comparable with that of $\hat{\theta}^G$. The local constant estimators $\tilde{\theta}^G$ and $\tilde{\theta}$ generally have smaller variances than $\hat{\theta}^G$ and $\hat{\theta}$, but have much larger biases and thus larger MSEs.[9] All four estimators are generally biased downwards.[10] On the other hand, a preliminary simulation indicates that these estimators are generally biased upwards if the discontinuity point suspected is on the left side of the peak, e.g. $c = 11$. Typical bias-variance trade-offs for the bandwidth selection is also observed: the biases are larger and the variances are smaller when the bandwidth increases. Compared to the case of the normal distribution, the four estimators have significantly larger biases and slightly larger variances in the case of the t-distribution. Again, $\hat{\theta}$ appears to have smaller MSEs than other estimators.

Next we look at the tests for (dis)continuity in the density function. We consider a general set-up of mixture of normal distributions. Suppose that the random variable X is drawn from truncated $N(\mu, \sigma^2)$ on $(-\infty, c)$ with probability γ, and from truncated $N(\mu, \sigma^2)$ on $(c, +\infty)$ with probability $1 - \gamma$. Note that X is $N(\mu, \sigma^2)$ distributed when $\gamma = \Phi(c)$, where Φ is the cumulative distribution function of $N(\mu, \sigma^2)$. If $\gamma \neq \Phi(c)$, the density function of X is discontinuous at c, e.g., if $\gamma < \Phi(c)$,

[8] The Monte Carlo average of h_{dd} is around 1.7. Thus, the cases of $h = 1$ and $1.5^{-1} h_{dd}$ can be regarded as undersmoothed bandwidths.

[9] These results can be explained by the asymptotic theory. Under the same assumptions in Theorem 2.1, we can derive $\sqrt{nh}\left(\tilde{\theta} - \theta - B_C\right) \xrightarrow{d} N(0, V_C)$, where $B_C = 2h\left(f_r' K_{r11} - f_l' K_{l11}\right)$ and $V_C = 4\left(f_r + f_l\right) K_{r02}$. For the triangle kernel, $B_C = \frac{h}{3}\left(f_r' + f_l'\right)$ and $V_C = \frac{4}{3}\left(f_r + f_l\right)$. On the other hand, from Theorem 2.1, the asymptotic bias and variance of $\hat{\theta}$ are $B_L = \frac{h^2}{20}\left(f_l'' - f_r''\right)$ and $V_L = \frac{24}{5}\left(f_r + f_l\right)$. The bias of $\tilde{\theta}$ tends to be larger than that of $\hat{\theta}$ because (i) B_C and B_L are of order h and h^2, respectively, and (ii) when the density function is continuously second-order differentiable, B_L vanishes but B_C does not in general. Also, we can see that the asymptotic variance of $\tilde{\theta}$ is smaller than that of $\hat{\theta}$ (note that $V_C/V_L = 5/18$). Similar comments apply to the comparison of $\tilde{\theta}^G$ and $\hat{\theta}^G$.

[10] From the previous footnote, the asymptotic bias of $\tilde{\theta}$ is written as $B_C \approx -0.043h$ for the case of $N(12, 3)$, and $B_C \approx -0.083h$ for the case of $t(5)$. Similar results apply to $\tilde{\theta}^G$. Since $B_L = 0$ in both simulation designs, we need to analyze higher-order bias terms to characterize the downward biases in $\hat{\theta}$ and $\hat{\theta}^G$.

the density of X has an upward jump at c. As above, we set $\mu = 12$, $\sigma^2 = 3$, and $c = 13$. For sample size $n = 1000, 2000, 5000$, we generate random samples of X when $d = 0, 0.02, 0.04, 0.06, 0.08, 0.1$, where $d = \Phi(c) - \gamma$ measures the size of discontinuity. When $d = 0$, the rejection rate (over replications) becomes the finite-sample size of the test.

We consider the Wald statistic W^G using $\hat{\theta}^G$ and empirical likelihood statistics ℓ^G and ℓ. All these statistics have the $\chi^2(1)$ null limiting distribution. The bin size b and the bandwidth h are selected data-dependently following McCrary (2008).

The simulation results are summarized in Table 3. It shows that all three tests have finite-sample sizes that are reasonably close to the nominal ones (5% or 10% under consideration), with mild over-rejection observed for the W^G and ℓ tests and sometime mild under-rejection for the ℓ^G test. Finite-sample quantiles of the three test statistics are also reported. Comparing with the theoretical quantiles of $\chi^2(1)$ distribution, we can see that the distributions of the empirical likelihood statistics ℓ^G and ℓ are better approximated by their limit distribution than that of the Wald statistic W^G is. The p-value plots and p-value discrepancy plots (Davidson and MacKinnon, 1998) for the three tests when $n = 2000$ are displayed in Figure 1. Table 3 also shows that all three tests have monotonic power, and appear to be consistent with power approaching one as sample size increases. The test based on ℓ has uniformly significantly higher power than the other two tests. The ℓ^G test is generally more powerful than W^G especially for small deviations from the null hypothesis.[11]

To summarize, we recommend to use the local likelihood method for point estimation and form tests or confidence sets via empirical likelihood. They suffer from relatively less boundary biases and the associated tests are more powerful than other existent procedures. For the binning estimator, the empirical likelihood test appears to be more conservative (i.e. being very careful to report a discontinuity) than the Wald test while not sacrificing power. These points are reinforced in the empirical example analyzed below.

5 Empirical illustration

Class size is one of the main determinants of the economic cost of education and its effects on children's test scores and on adult earnings have attracted substantial interest. In a recent study, Angrist and Lavy (1999) approached the problem for Israeli public schools and exploited the fact that, the so-called Maimonides' rule, which stipulates that a class with more than 40 pupils should be split into two, is used to determine the division of enrollment cohorts into classes. This rule introduces a nonlinear and non-monotonic relationship between class size and grade enrollment; there is significant drop of class size at the values of enrollments that are just above multiples of 40, e.g., 41-45, 81-85, etc. Angrist and Lavy (1999) used this rule as an exogenous source of variation in class size to identify the effects of class

[11]Note that the power comparison reported here is most favorable for the W^G test as it over-rejects most under the null hypothesis.

size on the scholastic achievement of Israeli pupils.

An important identifying assumption of Angrist and Lavy (1999) is no manipulation of class size by parents, which is the testing focus of this section. Precisely, there could be two kinds of manipulation. The first one is that parents may selectively exploit the Maimonides' rule by registering their children in the schools with enrollments just above multiples of 40 so that their children are placed in classes with smaller sizes. Following McCrary's (2008) arguments, this would lead to an increase in the density of enrollment counts around the point that is just above a multiple of 40. Angrist and Lavy (1999) argued that this kind of manipulation is unlikely to happen for two reasons. First, Israeli pupils are required to attend school in their local registration area. Also, principals are required to grant enrollment to students within their district and are not permitted to register students from outside their district. Second, even in exceptional cases that parents intentionally move to another school district hoping to get a better draw in the enrollment lottery (e.g. 41-45 instead of 38), "there is no way to know (exactly) whether a predicted enrollment of 41 will not decline to 38 by the time school starts, obviating the need for two small classes" (Angrist and Lavy, 1999).

The second kind of manipulation of class size is that parents may extract their kids from the public school system when they find the enrollments of the schools where their kids are registered are just below multiples of 40. This would lead to a decrease of the enrollment density on the left side of the multiples of 40. However, as argued by Angrist and Lavy, unlike in the United States private elementary schooling is rare in Israel.

To assess the validity of the assumption of no manipulation of class size we test continuity of the density function of enrollment counts. We consider fifth graders. In the end, our data contain 2029 schools (Angrist and Lavy, 1999). The histogram is displayed in Figure 2. It shows a sharp increase of densities at the enrollment of 40 but such increase is not clearly observed for other multiples of 40. This observation is reinforced in graphical analysis displayed in Figures 3 and 4, which show the estimated enrollment density function using the data on the either side of 40 and 120 respectively.

We perform the binning and local likelihood estimation ($\hat{\theta}^G$ and $\hat{\theta}$) and the associated tests (W^G, ℓ^G, and ℓ) of the discontinuities in enrollment densities that are suspected at the multiples of 40 over a range of smoothing bandwidths. The results are summarized in Table 4.

The local likelihood method finds upward jumps of the enrollment density at c =40, 80, and 120 and a downward jump at $c = 160$. The associated empirical likelihood tests show that the discontinuity at an enrollment of 40 is very significant with test statistics all valued larger than 20 for different bandwidths. The evidence of discontinuity at 80 is relatively weak and significance depends on the bandwidth used, while no evidence of discontinuity is found at 120 and 160. The progressively weaker evidence of discontinuity coincides with the extent of decrement of class sizes at different multiples of 40. For example, according to Maimonides' rule, the class size drops faster at the enrollment of 40 than it does at 80. It in turns drops faster at 80 than it does at 120 and so on. Thus parents are more likely to selectively manipulate class size as just above 40 because they could place their children in schools with smaller class sizes if the manipulation is successful, than they do as just above 80, 120, or 160.

The binning method generally produces smaller estimates of the discontinuities than the local likelihood method. It estimates f_l larger and estimates f_r smaller, compared with the corresponding local likelihood results.[12] The binning estimates find a positive jump of the enrollment density only at $c = 40$ and negative jumps at $c = 80, 120, 160$. McCrary's (2008) Wald test shows somewhat strong significance of discontinuity at 40 but the significance disappears at even 10% level when a small bandwidth $h = 15$ is used. While no significance is found at $c = 80$, significance with at least 5% level is present at $c = 120$ and 160. Note that it does not support the existence of manipulation at enrollment of 120 and 160 since the point estimates of discontinuities are negative at these two points. The empirical likelihood tests based on the binning estimators are more conservative than the Wald tests and they do not find any significant evidence of manipulation even at the enrollment of 40.[13] Table 5 gives the empirical likelihood confidence sets of the discontinuity at 40 for both binning and local likelihood estimators. It is noteworthy that McCrary's (2008) Wald test cannot generate such interval estimates.

The analysis above provides a nonparametric data-based re-examination of the identifying assumption in the regression discontinuity design used by Angrist and Lavy (1999). It is achieved via testing density continuity of the running variable. Our statistical results show that validation of the no manipulation assumption hinges on the inference methods used and also the amount of smoothing the practitioners decide on. Caution should be used when manipulation is detected, since it casts doubt on nearly randomized assignment of treatment in the neighborhood of the cutoff point and thus makes interpretation of the regression discontinuity application questionable.

6 Conclusion

This paper is concerned with estimation and inference of (dis)continuities of density functions, which often play fundamental roles in empirical economic analysis. Several issues with existing inference methods are addressed and competitive alternatives are suggested. In particular, we consider both the binning and local likelihood estimators of the discontinuities. A novel framework for inference based on the idea of empirical likelihood is introduced. We study the first- and second-order asymptotic properties of the proposed test statistics. The benefits of the proposed methods are illustrated by a simulation study and an empirical application involving the popular regression discontinuity design. It is interesting to conduct higher-order analysis for the power properties of the empirical likelihood statistics particularly to understand the desirable power properties of the local likelihood method reported in our simulation study.

[12] Since both the binning and local likelihood estimators have the same dominant terms for the asymptotic biases, we need to investigate the higher-order bias terms of these estimators to explain those finite sample phenomena.

[13] Although the test ℓ^G is significant at $c = 120$ at the 5% level when the bandwidth $h = 15$ or 20 is used, the point estimate of θ is negative so the hypothesis of manipulation is not supported.

A Mathematical Appendix

Hereafter "w.p.a.1" means "with probability approaching one". Define $f = f(c)$, $\gamma = (a_l, b_l, b_r)'$, $\Gamma = \mathcal{A}_l \times \mathcal{B}_l \times \mathcal{B}_r$,

$$\begin{aligned}
\gamma_0 &= (\alpha_0, \beta_{l0}, \beta_{r0})' = \left(\log f_l, \lim_{x \uparrow c} \frac{d \log f(x)}{dx}, \lim_{x \downarrow c} \frac{d \log f(x)}{dx} \right)', \\
\hat{\gamma} &= \arg\min_{\gamma \in \Gamma} \ell\left(a_l, \log\left(\theta_0 + e^{a_l}\right), b_l, b_r\right), \\
g_i(\gamma) &= g_i\left(a_l, \log\left(\theta_0 + e^{a_l}\right), b_l, b_r\right), \quad A_i = \left((1 - I_i), (1 - I_i)\left(\frac{X_i - c}{h}\right), I_i, I_i\left(\frac{X_i - c}{h}\right) \right)', \\
K_{lj_1j_2} &= \int_{-k \le u < 0} u^{j_1} \mathbb{K}^{j_2}(u)\, du, \quad K_{rj_1j_2} = \int_{0 < u \le k} u^{j_1} \mathbb{K}^{j_2}(u)\, du, \\
V &= \begin{bmatrix} V_l & 0 \\ 0 & V_r \end{bmatrix}, \quad V_l = f \begin{bmatrix} K_{l02} & K_{l12} \\ K_{l12} & K_{l22} \end{bmatrix}, \quad V_r = f \begin{bmatrix} K_{r02} & K_{r12} \\ K_{r12} & K_{r22} \end{bmatrix}, \\
G &= f \begin{bmatrix} K_{l01} & K_{l11} & 0 \\ K_{l11} & K_{l21} & 0 \\ K_{r01} & 0 & K_{r11} \\ K_{r11} & 0 & K_{r21} \end{bmatrix}, \quad \hat{P}(\gamma, \lambda) = \frac{1}{nh} \sum_{i=1}^{n} \log\left(1 + \lambda' g_i(\gamma)\right).
\end{aligned} \tag{12}$$

A.1 Proof of Theorem 3.1

Since the proof is similar, we only show the second statement, $\ell(\theta_0) = \min_{\gamma \in \Gamma} \ell\left(a_l, \log\left(\theta_0 + e^{a_l}\right), b_l, b_r\right) \xrightarrow{d} \chi^2(1)$. The proof of the first part is available from the authors upon request.

First, we show the consistency of $\hat{\gamma}$ to γ_0. By the change of variables and one-sided Taylor expansions,

$$\left| \frac{1}{h} E\left[A_i \mathbb{K}\left(\frac{X_i - c}{h}\right) \right] \right| = O(1), \quad \left| \frac{1}{h} E\left[A_i A_i' \mathbb{K}^2\left(\frac{X_i - c}{h}\right) \right] \right| = O(1).$$

Thus, the Chebyshev inequality implies

$$\sup_{\gamma \in \Gamma} \left| \frac{1}{nh} \sum_{i=1}^{n} g_i(\gamma) - \frac{1}{h} E[g_i(\gamma)] \right| = \left| \frac{1}{nh} \sum_{i=1}^{n} A_i \mathbb{K}\left(\frac{X_i - c}{h}\right) - \frac{1}{h} E\left[A_i \mathbb{K}\left(\frac{X_i - c}{h}\right) \right] \right| = O_p\left((nh)^{-1/2} \right). \tag{13}$$

By the triangle inequality, (13), Lemma 4, and $h^{-1}(nh)^{-1/2} \to 0$ (by Assumption 4),

$$\left| \frac{1}{h^2} E[g_i(\hat{\gamma})] \right| \le \frac{1}{h} \left| \frac{1}{h} E[g_i(\hat{\gamma})] - \frac{1}{nh} \sum_{i=1}^{n} g_i(\hat{\gamma}) \right| + \frac{1}{h} \left| \frac{1}{nh} \sum_{i=1}^{n} g_i(\hat{\gamma}) \right| \xrightarrow{p} 0.$$

Also, by the change of variables,

$$\begin{aligned}
\frac{1}{h^2} E[g_i(\gamma)] = &\left(\frac{1}{h} \int_{u<0} (1, u)\, \mathbb{K}(u) \left\{ f(c + uh) - \exp(a_l + b_l uh) \right\} du, \right. \\
&\left. \frac{1}{h} \int_{u \ge 0} (1, u)\, \mathbb{K}(u) \left\{ f(c + uh) - \exp\left(\log\left(\theta_0 + e^{a_l}\right) + b_r uh\right) \right\} du \right)',
\end{aligned}$$

and thus γ_0 uniquely solves $0 = \lim_{n\to\infty} \frac{1}{h^2} E\left[g_i(\gamma)\right]$ with respect to γ (which can be seen by a second-order expansion of $\log f(c+uh)$ around $u=0$). Therefore, the convergence $\frac{1}{h^2} E\left[g_i(\hat\gamma)\right] \xrightarrow{p} 0$ implies the consistency $\hat\gamma \xrightarrow{p} \gamma_0$.

Second, we derive an asymptotic expansion for the empirical likelihood function $\ell(\hat\gamma)$. From Lemma 3, the Lagrange multiplier $\hat\lambda(\hat\gamma)$ satisfies the first-order condition

$$0 = \frac{1}{nh}\sum_{i=1}^{n} \frac{g_i(\hat\gamma)}{1+\hat\lambda(\hat\gamma)' g_i(\hat\gamma)} = \frac{1}{nh}\sum_{i=1}^{n} g_i(\hat\gamma) - \hat V_1 \hat\lambda(\hat\gamma), \tag{14}$$

w.p.a.1, where $\hat V_1 = \frac{1}{nh}\sum_{i=1}^{n} \frac{g_i(\hat\gamma) g_i(\hat\gamma)'}{\left(1+\tilde\lambda' g_i(\hat\gamma)\right)^2}$ with $\tilde\lambda$ on the line joining $\hat\lambda(\hat\gamma)$ and 0, and the second equality follows from an expansion around $\hat\lambda(\hat\gamma) = 0$. From Lemma 1 and 2 and the consistency of $\hat\gamma$, we have $\hat V_1 \xrightarrow{p} V$. Since V is invertible (Assumption 2), $\hat V_1$ is invertible w.p.a.1. Thus, solving (14) for $\hat\lambda(\hat\gamma)$,

$$\hat\lambda(\hat\gamma) = \hat V_1^{-1} \frac{1}{nh}\sum_{i=1}^{n} g_i(\hat\gamma),$$

w.p.a.1. From this and the second-order expansion of $2\sum_{i=1}^{n} \log\left(1+\hat\lambda(\hat\gamma)' g_i(\hat\gamma)\right)$ yields

$$\ell(\hat\gamma) = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} g_i(\hat\gamma)' \left[2\hat V_1^{-1} - \hat V_1^{-1}\hat V_2 \hat V_1^{-1}\right] \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} g_i(\hat\gamma), \tag{15}$$

where $\hat V_2 = \frac{1}{nh}\sum_{i=1}^{n} \frac{g_i(\hat\gamma) g_i(\hat\gamma)'}{\left(1+\bar\lambda' g_i(\hat\gamma)\right)^2}$ with $\bar\lambda$ on the line joining $\hat\lambda(\hat\gamma)$ and 0.

Third, we derive the asymptotic distribution of $\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} g_i(\hat\gamma)$. Since the derivative of the first-order condition (14) with respect to $\hat\lambda(\hat\gamma)$ converges in probability to the positive definite matrix V, we can apply the implicit function theorem, i.e., $\hat\lambda(\gamma)$ is continuously differentiable with respect to γ in a neighborhood of $\hat\gamma$ w.p.a.1. The envelope theorem implies

$$0 = \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{1+\hat\lambda(\hat\gamma)' g_i(\hat\gamma)} \left(\frac{\partial g_i(\hat\gamma)}{\partial\gamma'}\right)' \hat\lambda(\hat\gamma), \tag{16}$$

w.p.a.1. On the other hand, an expansion of (14) around $\left(\hat\gamma, \hat\lambda(\hat\gamma)\right) = (\gamma_0, 0)$ yields

$$0 = \frac{1}{nh}\sum_{i=1}^{n} g_i(\gamma_0) + \left(\frac{1}{nh}\sum_{i=1}^{n} \frac{1}{1+\tilde\lambda' g_i(\tilde\gamma)} \frac{\partial g_i(\tilde\gamma)}{\partial\gamma'} H^{-1}\right) H(\hat\gamma - \gamma_0) - \hat V_3 \hat\lambda(\hat\gamma), \tag{17}$$

where $H = \operatorname{diag}(1, h, h)$, $\left(\tilde\gamma, \tilde\lambda\right)$ is a point on the line joining $\left(\hat\gamma, \hat\lambda(\hat\gamma)\right)$ and $(\gamma_0, 0)$, and $\hat V_3 = \frac{1}{nh}\sum_{i=1}^{n} \frac{g_i(\tilde\gamma) g_i(\tilde\gamma)'}{\left(1+\tilde\lambda' g_i(\tilde\gamma)\right)^2}$ is implicitly defined. Combining (16) multiplied H^{-1} from left and (17),

$$0 = \begin{pmatrix} 0 \\ \frac{1}{nh}\sum_{i=1}^{n} g_i(\gamma_0) \end{pmatrix} + \hat M \begin{pmatrix} H(\hat\gamma - \gamma_0) \\ \hat\lambda(\hat\gamma) \end{pmatrix}, \quad \text{where } \hat M = \begin{pmatrix} 0 & -\hat G_1' \\ -\hat G_2 & -\hat V_3 \end{pmatrix}, \tag{18}$$

where

$$\hat G_1 = \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{1+\hat\lambda(\hat\gamma)' g_i(\hat\gamma)} \frac{\partial g_i(\hat\gamma)}{\partial\gamma'} H^{-1}, \quad , \hat G_2 = -\frac{1}{nh}\sum_{i=1}^{n} \frac{1}{1+\tilde\lambda' g_i(\tilde\gamma)} \frac{\partial g_i(\tilde\gamma)}{\partial\gamma'} H^{-1},$$

which satisfy $\hat{G}_1, \hat{G}_2 \xrightarrow{p} G$ by a similar argument to Lemma 1 and the consistency of $\hat{\gamma}$. Also note that $\hat{V}_3 \xrightarrow{p} V$. Since G is full column rank and V is positive definite (Assumption 2), $\hat{M}$ is invertible w.p.a.1. By solving (18) for $\sqrt{nh}H(\hat{\gamma}-\gamma_0)$, we have

$$\sqrt{nh}H(\hat{\gamma}-\gamma_0) = \left(G'V^{-1}G\right)^{-1} G'V^{-1} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} g_i(\gamma_0) + o_p(1).$$

From this and an expansion of $\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} g_i(\hat{\gamma})$ around $\hat{\gamma}=\gamma_0$,

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} g_i(\hat{\gamma}) = \left[I - G\left(G'V^{-1}G\right)^{-1} G'V^{-1}\right] \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} g_i(\gamma_0) + o_p(1). \tag{19}$$

Combining (15), (19), $\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} g_i(\gamma_0) \xrightarrow{d} N(0,V)$ (Lemma 1), and $2\hat{V}_1^{-1} - \hat{V}_1^{-1}\hat{V}_2\hat{V}_1^{-1} \xrightarrow{p} V^{-1}$ (by Lemma 1 and 2 with the consistency of $\hat{\gamma}$), we have

$$\begin{aligned} \ell(\hat{\gamma}) &\xrightarrow{d} \phi' V^{1/2}\left[I - G\left(G'V^{-1}G\right)^{-1} G'V^{-1}\right]' V^{-1}\left[I - G\left(G'V^{-1}G\right)^{-1} G'V^{-1}\right] V^{1/2}\phi \\ &= \phi'\left[I - A\left(A'A\right)^{-1} A'\right]\phi = \chi^2(1), \end{aligned}$$

where $\phi \sim N(0,I)$ and $A = V^{-1/2}G$. Therefore, the conclusion is obtained.

A.2 Lemma for Theorem 3.1

Lemma. *Under Assumptions 1-5,*

1. $\frac{1}{nh}\sum_{i=1}^{n} g_i(\gamma_0) g_i(\gamma_0)' \xrightarrow{p} V$, $\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} g_i(\gamma_0) \xrightarrow{d} N(0,V)$;
2. *For each* $\zeta \in (0,\infty)$ *and* $\bar{\Lambda}_n = \left\{\lambda : |\lambda| \le n^{-\zeta}\right\}$, $\sup_{\gamma\in\Gamma,\lambda\in\bar{\Lambda}_n,1\le i\le n} |\lambda' g_i(\gamma)| \xrightarrow{p} 0$ *and for each* $\gamma \in \Gamma$, $\bar{\Lambda}_n \subseteq \Lambda_n(\gamma) = \Lambda_n(a_l, \log(\theta_0 + e^{a_l}), b_l, b_r)$ *w.p.a.1;*
3. *For any* $\bar{\gamma}$ *satisfying* $\bar{\gamma} \xrightarrow{p} \gamma_0$ *and* $\frac{1}{nh}\sum_{i=1}^{n} g_i(\bar{\gamma}) = O_p\left((nh)^{-1/2}\right)$, *there exists* $\hat{\lambda}(\bar{\gamma}) = \arg\max_{\lambda\in\Lambda_n(\bar{\gamma})} \hat{P}(\bar{\gamma},\lambda)$ *w.p.a.1.,* $\left|\hat{\lambda}(\bar{\gamma})\right| = O_p\left((nh)^{-1/2}\right)$, *and* $\sup_{\lambda\in\Lambda_n(\bar{\gamma})} \hat{P}(\bar{\gamma},\lambda) = O_p\left((nh)^{-1}\right)$;
4. $\frac{1}{nh}\sum_{i=1}^{n} g_i(\hat{\gamma}) = O_p\left((nh)^{-1/2}\right)$.

Proof of 1. Proof of the first statement. Let $\hat{V} = \left[\hat{V}_{ab}\right] = \frac{1}{nh}\sum_{i=1}^{n} g_i(\gamma_0) g_i(\gamma_0)'$ for $a,b = 1,\ldots,4$. By the change of variables,

$$\begin{aligned} \int_{x<c} \left(1, \frac{x-c}{h}\right) \mathbb{K}\left(\frac{x-c}{h}\right) \exp\left(\alpha_0 + \beta_{l0}(x-c)\right) dx &= h\int_{u<0} (1,u)\,\mathbb{K}(u) \exp\left(\alpha_0 + \beta_{l0}uh\right) du = O(h), \\ \int_{x\ge c} \left(1, \frac{x-c}{h}\right) \mathbb{K}\left(\frac{x-c}{h}\right) \exp\left(\alpha_0 + \beta_{l0}(x-c)\right) dx &= h\int_{u\ge 0} (1,u)\,\mathbb{K}(u) \exp\left(\alpha_0 + \beta_{l0}uh\right) du = O(h). \end{aligned}$$

Thus, we have

$$\begin{aligned} \hat{V}_{11} &= \frac{1}{nh}\sum_{i=1}^{n} (1-I_i)\,\mathbb{K}^2\left(\frac{X_i-c}{h}\right) - \frac{2}{nh}\sum_{i=1}^{n} (1-I_i)\,\mathbb{K}\left(\frac{X_i-c}{h}\right) O(h) + O(h) \\ &= \frac{1}{h}E\left[(1-I_i)\,\mathbb{K}^2\left(\frac{X_i-c}{h}\right)\right] - \frac{2}{h}E\left[(1-I_i)\,\mathbb{K}\left(\frac{X_i-c}{h}\right)\right] O(h) + o_p(1) \\ &\xrightarrow{p} fK_{l02}, \end{aligned}$$

where the second equality follows from the weak law of large numbers and the convergence follows from the change of variables and Taylor expansions. The similar argument yields

$$\begin{aligned}\hat{V}_{22} &\xrightarrow{p} fK_{l22}, \quad \hat{V}_{33} \xrightarrow{p} fK_{r02}, \quad \hat{V}_{44} \xrightarrow{p} fK_{r22}, \\ \hat{V}_{12} &= \hat{V}_{21} \xrightarrow{p} fK_{l12}, \quad \hat{V}_{13} = \hat{V}_{31} \xrightarrow{p} 0, \quad \hat{V}_{14} = \hat{V}_{41} \xrightarrow{p} 0, \\ \hat{V}_{23} &= \hat{V}_{32} \xrightarrow{p} 0, \quad \hat{V}_{34} = \hat{V}_{43} \xrightarrow{p} fK_{r12}.\end{aligned}$$

The conclusion is obtained.

Proof of the second statement. Observe that

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} g_i(\gamma_0) = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \{g_i(\gamma_0) - E[g_i(\gamma_0)]\} + \sqrt{\frac{n}{h}} E[g_i(\gamma_0)]. \tag{20}$$

Let $g_{i,1}(\gamma_0)$ be the first element of $g_i(\gamma_0)$. By the change of variables,

$$\begin{aligned}\sqrt{\frac{n}{h}} E[g_{i,1}(\gamma_0)] &= \sqrt{nh} \int_{u<0} \mathbb{K}(u) \{f(c+uh) - \exp(\alpha_0 + \beta_{l0} uh)\} du \\ &= \sqrt{nh} K_{l01} \{f - \exp(\alpha_0)\} + \sqrt{nh^3} K_{l11} \{f_l' - \exp(\alpha_0)\beta_{l0}\} + O\left(\sqrt{nh^5}\right), \\ &\to 0,\end{aligned}$$

where the second equality follows from one-sided Taylor expansions and the convergence follows from the definitions of α_0 and β_{l0} and $nh^5 \to 0$ (Assumption 4). By applying the same argument, the second term of (20) satisfies $\left|\sqrt{\frac{n}{h}} E[g_i(\gamma_0)]\right| \to 0$. For the first term of (20), the Lyapunov central limit theorem implies $\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \{g_i(\gamma_0) - E[g_i(\gamma_0)]\} \xrightarrow{d} N(0, V)$, where the asymptotic variance is obtained from the first statement of Lemma 1 and $\left|\sqrt{\frac{n}{h}} E[g_i(\gamma_0)]\right| \to 0$. Therefore, the conclusion is obtained.

Proof of 2. Pick any $\zeta \in (0, \infty)$ and $\bar{\zeta} \in (0, \zeta)$. Since $E\left[\left(\sup_{\gamma \in \Gamma} |g_i(\gamma)|\right)^{1/\bar{\zeta}}\right] < \infty$ (because $\mathbb{K}$ has bounded support), the Markov inequality implies $\Pr\left\{\sup_{\gamma \in \Gamma} |g_i(\gamma)| \geq n^{\bar{\zeta}}\right\} \to 0$, i.e., $\sup_{\gamma \in \Gamma} |g_i(\gamma)| = O_p\left(n^{\bar{\zeta}}\right)$. Thus, $\sup_{\gamma \in \Gamma, \lambda \in \bar{\Lambda}_n, 1 \leq i \leq n} |\lambda' g_i(\gamma)| \leq O_p\left(n^{-\zeta + \bar{\zeta}}\right)$ and the first statement is obtained. Also, this implies that for each $i = 1, \ldots, n$, $\gamma \in \Gamma$, and $\lambda \in \bar{\Lambda}_n$, $\lambda' g_i(\gamma) \in \mathcal{V}$, w.p.a.1. Thus, the second statement follows.

Proof of 3. The basic steps are similar to Newey and Smith (2004, Lemma A2). Pick any $\zeta \in (0, \infty)$ satisfying $(nh)^{-1/2} n^{\zeta} \to 0$. Since $\bar{\Lambda}_n$ is compact and $\hat{P}(\bar{\gamma}, \lambda)$ is continuous in λ, there exists $\bar{\lambda} = \arg\max_{\lambda \in \bar{\Lambda}_n} \hat{P}(\bar{\gamma}, \lambda)$ w.p.a.1. Let $\bar{g} = \frac{1}{nh} \sum_{i=1}^{n} g_i(\bar{\gamma})$. Observe that for some $C > 0$,

$$0 = \hat{P}(\bar{\gamma}, 0) \leq \hat{P}(\bar{\gamma}, \bar{\lambda}) = \bar{\lambda}' \bar{g} - \frac{1}{2} \bar{\lambda}' \left(\frac{1}{nh} \sum_{i=1}^{n} \frac{g_i(\bar{\gamma}) g_i(\bar{\gamma})'}{\left(1 + \dot{\lambda}' g_i(\bar{\gamma})\right)^2} \right) \bar{\lambda} \leq |\bar{\lambda}| |\bar{g}| - C |\bar{\lambda}|^2, \tag{21}$$

w.p.a.1., where the first inequality follows from the definition of $\bar{\lambda}$, the second equality follows from a second-order expansion with $\dot{\lambda}$ on the line joining $\bar{\lambda}$ and 0, and the second inequality follows from $\frac{1}{nh} \sum_{i=1}^{n} g_i(\bar{\gamma}) g_i(\bar{\gamma})' \xrightarrow{p} V$, positive definiteness of V, and Lemma 2. Since $|\bar{\lambda}| \leq C |\bar{g}|$ and $|\bar{g}| =$

$O_p\left((nh)^{-1/2}\right)$ by the assumption, we have $\left|\bar{\lambda}\right| = O_p\left((nh)^{-1/2}\right)$. Since ζ is chosen to satisfy $(nh)^{-1/2} n^{\zeta} \to 0$, we have $\bar{\lambda} \in int\left(\bar{\Lambda}_n\right)$, i.e., $\bar{\lambda}$ is an interior solution. Thus, from concavity of $\hat{P}(\bar{\gamma}, \lambda)$ in λ, convexity of $\Lambda_n(\bar{\gamma})$, and $\bar{\Lambda}_n \subseteq \Lambda_n(\bar{\gamma})$ (by Lemma 2), $\bar{\lambda} = \hat{\lambda}(\bar{\gamma}) = \arg\max_{\lambda \in \Lambda_n(\bar{\gamma})} \hat{P}(\bar{\gamma}, \lambda)$ w.p.a.1., i.e., the first statement is obtained. Since $\left|\bar{\lambda}\right| = O_p\left((nh)^{-1/2}\right)$, the second statement is also obtained. The third statement is obtained from (21) with $\bar{\lambda} = \hat{\lambda}(\bar{\gamma})$.

Proof of 4. The basic steps are similar to Newey and Smith (2004, Lemma A3). Pick any $\zeta \in (0, \infty)$ satisfying $(nh)^{-1/2} n^{\zeta} \to 0$. Let $\hat{g} = \frac{1}{nh}\sum_{i=1}^{n} g_i(\hat{\gamma})$ and $\tilde{\lambda} = n^{-\zeta}\hat{g}/\left|\hat{g}\right|$ for ζ. Observe that for some $C > 0$,

$$\hat{P}\left(\hat{\gamma}, \tilde{\lambda}\right) = \tilde{\lambda}'\hat{g} - \frac{1}{2}\tilde{\lambda}'\left(\frac{1}{nh}\sum_{i=1}^{n}\frac{g_i(\hat{\gamma})\, g_i(\hat{\gamma})'}{\left(1 + \dot{\lambda}' g_i(\hat{\gamma})\right)^2}\right)\tilde{\lambda} \geq n^{-\zeta}\left|\hat{g}\right| - Cn^{-2\zeta}, \tag{22}$$

w.p.a.1., where the equality follows from a second-order expansion with $\dot{\lambda}$ on the line joining $\bar{\lambda}$ and 0, and the inequality follows from $\frac{1}{nh}\sum_{i=1}^{n} g_i(\hat{\gamma})\, g_i(\hat{\gamma})' \xrightarrow{p} V$, boundedness of V, and Lemma 2. Also note that

$$\sup_{\lambda \in \Lambda_n(\hat{\gamma})} \hat{P}(\hat{\gamma}, \lambda) \leq \sup_{\lambda \in \Lambda_n(\gamma_0)} \hat{P}(\gamma_0, \lambda) = O_p\left((nh)^{-1}\right), \tag{23}$$

w.p.a.1., where the inequality follows from the definition of $\hat{\gamma}$, and the equality follows from Lemma 3 with $\bar{\gamma} = \gamma_0$ and $\frac{1}{nh}\sum_{i=1}^{n} g_i(\gamma_0) = O_p\left((nh)^{-1/2}\right)$ (by Lemma 1). Since $\tilde{\lambda} \in \bar{\Lambda}_n$, Lemma 2 guarantees $\tilde{\lambda} \in \Lambda_n(\hat{\gamma})$, w.p.a.1., which implies $\hat{P}\left(\hat{\gamma}, \tilde{\lambda}\right) \leq \sup_{\lambda \in \Lambda_n(\hat{\gamma})} \hat{P}(\hat{\gamma}, \lambda)$. Thus, combining (22) and (23),

$$n^{-\zeta}\left|\hat{g}\right| - Cn^{-2\zeta} \leq O_p\left((nh)^{-1}\right), \tag{24}$$

w.p.a.1. Since we chose ζ to satisfy $(nh)^{-1/2} n^{\zeta} \to 0$, we have $\left|\hat{g}\right| = O_p\left(n^{-\zeta}\right)$. Now, pick any $\epsilon_n \to 0$ and define $\check{\lambda} = \epsilon_n \hat{g}$. From $\left|\hat{g}\right| = O_p\left(n^{-\zeta}\right)$, we have $\check{\lambda} = o_p\left(n^{-\zeta}\right)$ and $\check{\lambda} \in \bar{\Lambda}_n \subseteq \Lambda_n(\hat{\gamma})$. Thus, we apply the same argument to (22)-(24) after replacing $\tilde{\lambda}$ with $\check{\lambda}$. Then we obtain

$$\epsilon_n\left|\hat{g}\right|^2 - C\epsilon_n^2\left|\hat{g}\right|^2 \leq \hat{P}\left(\hat{\gamma}, \check{\lambda}\right) \leq \sup_{\lambda \in \Lambda_n(\hat{\gamma})} \hat{P}(\hat{\gamma}, \lambda) = O_p\left((nh)^{-1}\right),$$

which implies $\epsilon_n\left|\hat{g}\right|^2 = O_p\left((nh)^{-1}\right)$. Since this results holds for any $\epsilon_n \to 0$, we obtain the conclusion.

A.3 Proof of Theorem 3.2

We introduce some notation. Let us denote the moment functions evaluated at $\theta = \theta_0$ as $g_i(\gamma) = g_i\left(a_l, \log\left(\theta_0 + e^{a_l}\right), b_l, b_r\right)$, where $\gamma = (a_l, b_l, b_r)'$. Let $\gamma_0 = \left(\log f_l, \lim_{x\uparrow c}\frac{d\log f(x)}{dx}, \lim_{x\downarrow c}\frac{d\log f(x)}{dx}\right)'$, $\Omega = \frac{1}{h}E\left[g_i(\gamma_0)\, g_i(\gamma_0)'\right]$, and T be a 4×4 orthogonal matrix satisfying $T\Omega^{-1/2}\frac{1}{h}E\left[\frac{\partial g_i(\gamma_0)}{\partial \gamma'}\right]H^{-1}U = (\Lambda, 0_{3\times 1})'$, where $H = \mathrm{diag}(1, h, h)$, U is a 3×3 orthogonal matrix and Λ is a 3×3 nonsingular diagonal matrix. We orthogonalize the moment functions (evaluated at θ_0) as $w_i(\gamma) = T\Omega^{-1/2} g_i(\gamma)$ so that $\frac{1}{h}E\left[w_i(\gamma_0)\, w_i(\gamma_0)'\right] = I$. The (profile) empirical likelihood ratio in (11) can be rewritten as

$$\ell(\theta_0) = \min_{\gamma \in \mathcal{A}_l \times \mathcal{B}_l \times \mathcal{B}_r} 2\sum_{i=1}^{n}\log\left(1 + \tilde{\lambda}(\gamma)' w_i(\gamma)\right),$$

where $\tilde{\lambda}(\gamma)$ solves $\sum_{i=1}^{n} \frac{w_i(\gamma)}{1+\lambda' w_i(\gamma)} = 0$ with respect to λ for given values of γ. Let $\tilde{\gamma} = \arg\min_{\gamma \in \mathcal{A}_l \times \mathcal{B}_l \times \mathcal{B}_r} 2\sum_{i=1}^{n} \log\left(1 + \tilde{\lambda}(\gamma)' w_i(\gamma)\right)$ and $\tilde{\lambda} = \tilde{\lambda}(\tilde{\gamma})$. The first-order conditions for $\tilde{\lambda}$ and $\tilde{\gamma}$ are written as $\left(\frac{1}{nh}\sum_{i=1}^{n} \frac{w_i(\tilde{\gamma})'}{1+\tilde{\lambda}' w_i(\tilde{\gamma})}, H^{-1}\frac{1}{nh}\sum_{i=1}^{n} \frac{\tilde{\lambda}' \partial w_i(\tilde{\gamma})/\partial \gamma'}{1+\tilde{\lambda}' w_i(\tilde{\gamma})}\right)' = 0$. The fourth-order Taylor expansion of this condition around $\left(\tilde{\lambda}, \tilde{\gamma}\right) = (0_{4\times 1}, \gamma_0)$ and inversions yield expansion formulae for $\tilde{\lambda}$ and $H(\tilde{\gamma} - \gamma_0)$. By inserting those formulae to the fourth-order Taylor expansion of $\ell(\theta_0) = 2\sum_{i=1}^{n} \log\left(1 + \tilde{\lambda}' w_i(\tilde{\gamma})\right)$ around $\tilde{\lambda}' w_i(\tilde{\gamma}) = 0$, we can obtain an expansion formula for $\ell(\theta_0)$. Then based on this expansion formula, further lengthy calculations yield the the signed root expansion formula of $\ell(\theta_0)$, which is defined by the following notation. Let $U\Lambda^{-1} = \left(\omega^{kl}\right)_{3\times 3}$, $a^j = j$-th element of a vector a,

$$
\begin{aligned}
\alpha^{j_1 \cdots j_k} &= E\left[\frac{1}{h} w_i^{j_1}(\gamma_0) \cdots w_i^{j_k}(\gamma_0)\right], \quad A^{j_1 \cdots j_k} = \begin{cases} \frac{1}{nh}\sum_{i=1}^{n} w_i^{j_k}(\gamma_0) & \text{for } k = 1 \\ \frac{1}{nh}\sum_{i=1}^{n} w_i^{j_1}(\gamma_0) \cdots w_i^{j_k}(\gamma_0) - \alpha^{j_1 \cdots j_k} & \text{for } k \geq 2 \end{cases}, \\
\gamma^{j_1, m_1; j_2, m_2} &= \frac{1}{h^d} E\left[\frac{\partial w_i^{j_1}(\gamma_0)}{\partial \gamma_{m_1}} \frac{\partial w_i^{j_2}(\gamma_0)}{\partial \gamma_{m_2}}\right], \quad C^{j_1, m_1; j_2, m_2} = \frac{1}{nh^d}\sum_{i=1}^{n} \frac{\partial w_i^{j_1}(\gamma_0)}{\partial \gamma_{m_1}} \frac{\partial w_i^{j_2}(\gamma_0)}{\partial \gamma_{m_2}} - \gamma^{j_1, m_1; j_2, m_2}, \\
\gamma^{j, m_1 m_2} &= \frac{1}{h^d} E\left[\frac{\partial^2 w_i^j(\gamma_0)}{\partial \gamma_{m_1} \partial \gamma_{m_2}}\right], \quad C^{j, m_1 m_2} = \frac{1}{nh^d}\sum_{i=1}^{n} \frac{\partial^2 w_i^j(\gamma_0)}{\partial \gamma_{m_1} \partial \gamma_{m_2}} - \gamma^{j, m_1 m_2},
\end{aligned}
$$

where

$$
d = \begin{cases} 1 & \text{if } m_1, m_2 = 1 \\ 2 & \text{if } m_1 = 1,\ m_2 = 2, 3,\ \text{or } m_1 = 2, 3,\ m_2 = 1 \\ 3 & \text{if } m_1, m_2 = 2, 3 \end{cases}
$$

Hereafter, the ranges of the superscripts are fixed as $k, l, m, n, o, p, q, r \in \{1, 2, 3\}$. Also, by the convention, repeated superscripts are summed over (e.g.,$\omega^{kl} C^{4,k} A^l = \sum_{k=1}^{3}\sum_{l=1}^{3} \omega^{kl} C^{4,k} A^l$). Based on the above notation, the signed root expansion of $\ell(\theta_0)$ is obtained as

$$
(nh)^{-1} \ell(\theta_0) = (R_1 + R_2 + R_3)^2 + O_p\left(\left\{(nh)^{-1/2} + h^2\right\}^5\right),
$$

where

$$
R_1 = A^4,
$$

$$
R_2 = -\frac{1}{2} A^4 A^{44} + \frac{1}{3}\alpha^{444} A^4 A^4 - \omega^{kl} C^{4,k} A^l + \frac{1}{2}\omega^{km}\omega^{ln}\gamma^{4,kl} A^m A^n + \omega^{lm}\gamma^{4;4,l} A^4 A^m,
$$

$$
\begin{aligned}
R_3 \;=\;& \omega^{kl}\omega^{mn}C^{n,k}C^{4,m}A^{l}+\frac{1}{2}\omega^{kl}C^{4,k}A^{44}A^{l}-\frac{1}{2}\omega^{kl}\omega^{ml}C^{4,k}C^{4,m}A^{4}+\frac{3}{8}A^{44}A^{44}A^{4}+\omega^{kl}C^{4,k}A^{l4}A^{4}\\
&+\left\{\begin{array}{c}\omega^{kl}\gamma^{4;4,k}\alpha^{l44}-\frac{1}{2}\omega^{kl}\omega^{ml}\gamma^{4;4,k}\gamma^{4;4,m}\\ +\frac{4}{9}\alpha^{444}\alpha^{444}-\frac{1}{4}\alpha^{4444}\end{array}\right\}A^{4}A^{4}A^{4}\\
&+\omega^{mp}\left\{\begin{array}{c}\omega^{kl}\left(\gamma^{4;l,m}+\gamma^{l;4,m}\right)\gamma^{4;4,k}+3\omega^{kl}\gamma^{4,km}\alpha^{l44}\\ +\omega^{lo}\gamma^{4;4,l}\left(\gamma^{4;o,m}+\gamma^{o;4,m}-3\omega^{no}\gamma^{4,nm}\right)\\ +\alpha^{444}\left(\frac{2}{3}\gamma^{4;4,m}+\gamma^{4;4,m}\right)-\gamma^{4;4;4,m}\end{array}\right\}A^{p}A^{4}A^{4}\\
&+\omega^{mp}\omega^{nq}\left\{\begin{array}{c}-\frac{1}{2}\omega^{kl}\omega^{ol}\gamma^{4,km}\gamma^{4,on}+\omega^{kl}\gamma^{4,km}\left(\gamma^{l,n;4}+\gamma^{l;4,n}\right)\\ +\frac{1}{2}\omega^{kl}\gamma^{l,mn}\gamma^{4;4,k}+\frac{1}{3}\gamma^{4,mn}\alpha^{444}-\frac{1}{2}\gamma^{4;4,mn}-\frac{1}{2}\gamma^{4,m;4,n}\\ +\frac{1}{2}\gamma^{4;4,m}\left(\gamma^{4;4,n}+\gamma^{4;4,n}\right)+\frac{1}{2}\gamma^{4;4,m}\gamma^{4;4,n}\end{array}\right\}A^{p}A^{q}A^{4}\\
&+\frac{1}{2}\omega^{mp}\omega^{nq}\omega^{or}\left\{\omega^{kl}\omega^{4,kl}\omega^{l,mn}+\gamma^{4,mn}\gamma^{4;4,o}-\frac{1}{3}\gamma^{4;m,no}\right\}A^{p}A^{q}A^{r}\\
&-\frac{1}{2}\omega^{kn}\omega^{lo}\omega^{pm}\gamma^{m,kl}C^{4,p}A^{n}A^{o}-\frac{1}{4}\omega^{kn}\omega^{lo}\gamma^{4,kl}A^{44}A^{n}A^{o}+\omega^{km}\omega^{nm}\omega^{lo}\gamma^{4,kl}C^{4,n}A^{o}A^{4}\\
&-\omega^{km}\omega^{lo}\gamma^{4,kl}A^{m4}A^{o}A^{4}-\omega^{kp}\omega^{ln}\omega^{mo}\gamma^{4,kl}C^{p,n}A^{n}A^{o}+\frac{1}{2}\omega^{kn}\omega^{lo}C^{4,kl}A^{n}A^{o}+\frac{1}{3}A^{444}A^{4}A^{4}\\
&+\omega^{lm}\omega^{nm}\gamma^{4;4,l}C^{4,n}A^{4}A^{4}-\omega^{lm}\gamma^{4;4,l}A^{m4}A^{4}A^{4}-\omega^{ln}\omega^{mo}\gamma^{4;4,l}C^{n,m}A^{o}A^{4}\\
&-\omega^{ln}\omega^{mk}\left(\gamma^{k;4,l}+\gamma^{4;k,l}\right)C^{4,m}A^{n}A^{4}-\omega^{ln}\omega^{mo}\gamma^{4;4,l}C^{4,m}A^{n}A^{o}-\left(\frac{1}{2}\gamma^{4;4,l}+\gamma^{4;4,l}\right)\omega^{ln}A^{44}A^{n}A^{4}\\
&+\omega^{lm}C^{4;4,l}A^{m}A^{4}-\omega^{mk}\alpha^{k44}C^{4,m}A^{4}A^{4}-\frac{2}{3}\omega^{lm}\alpha^{444}C^{4,l}A^{m}A^{4}-\frac{5}{6}\alpha^{444}A^{44}A^{4}A^{4}.
\end{aligned}
$$

Based on this formula, we compute the cumulants of $R = R_1 + R_2 + R_3$. Let κ_j be the j-th cumulant of R. In Appendix A.4, we derive

$$
\begin{aligned}
\kappa_1 &= \alpha^4-(nh)^{-1}\left\{\frac{1}{6}\alpha^{444}-\omega^{kl}\gamma^{l;4,k}+\frac{1}{2}\omega^{km}\omega^{lm}\gamma^{4,kl}\right\}+O\left((nh)^{-2}+(nh)^{-1}h^2+h^4\right),\\
\kappa_2 &= (nh)^{-1}+(nh)^{-1}\left\{\frac{1}{3}\alpha^4\alpha^{444}+2\alpha^4\omega^{kl}\gamma^{l;4,k}-\alpha^4\omega^{km}\omega^{lm}\gamma^{4,kl}\right\}+(nh)^{-2}\Delta+O\left((nh)^{-3}+(nh)^{-2}h^2\right),\\
\kappa_3 &= O\left((nh)^{-3}+(nh)^{-2}h^2\right),\\
\kappa_4 &= O\left((nh)^{-4}+(nh)^{-3}h^2\right),
\end{aligned}
$$

where Δ is defined in (27).

Based on these cumulants, we can apply a conventional argument to derive the Edgeworth expansion and Bartlett correction for the empirical likelihood statistic $\ell(\theta_0)$ (see, Chen and Cui, 2006, and Chen and Qin, 2000). Thus, the conclusions are obtained with the Bartlet factor

$$
B_c=(nh)\left(\alpha^4\right)^2+(nh)^{-1}\left\{\Delta+\left(\frac{1}{6}\alpha^{444}\right)^2+\left(\omega^{kl}\gamma^{l;4,k}\right)^2+\left(\frac{1}{2}\omega^{km}\omega^{lm}\gamma^{4,kl}\right)^2\right\}. \tag{25}
$$

A.4 Computation of Cumulants

A.4.1 1st Cumulant

For $R_1 = A^4$, we have

$$E\left[R_1\right] = E\left[\frac{1}{nh}\sum_{i=1}^{n} w_i^4\left(\gamma_0\right)\right] = \alpha^4.$$

For $R_2 = -\frac{1}{2}A^4A^{44} + \frac{1}{3}\alpha^{444}A^4A^4 - \omega^{kl}C^{4,k}A^l + \frac{1}{2}\omega^{km}\omega^{ln}\gamma^{4,kl}A^mA^n + \omega^{lm}\gamma^{4;4,l}A^4A^m$, the first term satisfies

$$E\left[-\frac{1}{2}A^4A^{44}\right] = -\frac{1}{2}E\left[\left(\frac{1}{nh}\sum_{i=1}^{n} w_i^4\left(\gamma_0\right)\right)\left(\frac{1}{nh}\sum_{i=1}^{n}\left(w_i^4\left(\gamma_0\right)\right)^2 - 1\right)\right] = -\frac{1}{2}\left(nh\right)^{-1}\alpha^{444} + O\left(\left(nh\right)^{-1}h^3\right),$$

the second term satisfies

$$E\left[\frac{1}{3}\alpha^{444}\left(A^4\right)^2\right] = \frac{1}{3}\alpha^{444}E\left[\left(\frac{1}{nh}\sum_{i=1}^{n} w_i^4\left(\gamma_0\right)\right)\left(\frac{1}{nh}\sum_{i=1}^{n} w_i^4\left(\gamma_0\right)\right)\right] = \frac{1}{3}\left(nh\right)^{-1}\alpha^{444} + O\left(h^4\right),$$

the third term satisfies

$$E\left[-\omega^{kl}C^{4,k}A^l\right] = -\omega^{kl}E\left[\left(\frac{1}{nh}\sum_{i=1}^{n}\frac{\partial w_i^4\left(\gamma_0\right)}{\partial\gamma_k} - \gamma^{4,k}\right)\left(\frac{1}{nh}\sum_{i=1}^{n} w_i^l\left(\gamma_0\right)\right)\right] = -\left(nh\right)^{-1}\omega^{kl}\gamma^{l;4,k} + O\left(\left(nh\right)^{-1}h^3\right),$$

the fourth term satisfies

$$\begin{aligned} E\left[\frac{1}{2}\omega^{km}\omega^{ln}\gamma^{4,kl}A^mA^n\right] &= \frac{1}{2}\omega^{km}\omega^{lm}\gamma^{4,kl}E\left[\left(\frac{1}{nh}\sum_{i=1}^{n} w_i^m\left(\gamma_0\right)\right)\left(\frac{1}{nh}\sum_{i=1}^{n} w_i^n\left(\gamma_0\right)\right)\right] \\ &= \left(nh\right)^{-1}\frac{1}{2}\omega^{km}\omega^{lm}\gamma^{4,kl} + O\left(h^4\right), \end{aligned}$$

and the fifth term satisfies

$$E\left[\omega^{lm}\gamma^{4;4,l}A^4A^m\right] = \omega^{lm}\gamma^{4;4,l}E\left[\left(\frac{1}{nh}\sum_{i=1}^{n} w_i^4\left(\gamma_0\right)\right)\left(\frac{1}{nh}\sum_{i=1}^{n} w_i^m\left(\gamma_0\right)\right)\right] = O\left(h^4\right).$$

Combining these results,

$$E\left[R_2\right] = -\frac{1}{6}\left(nh\right)^{-1}\alpha^{444} - \left(nh\right)^{-1}\omega^{kl}\gamma^{l;4,k} + \left(nh\right)^{-1}\frac{1}{2}\omega^{km}\omega^{lm}\gamma^{4,kl} + O\left(\left(nh\right)^{-1}h^3 + h^4\right).$$

Also, similar but more lengthy calculation yields

$$E\left[R_3\right] = O\left(\left(nh\right)^{-2} + \left(nh\right)^{-1}h^2 + h^6\right).$$

Therefore, the 1st cumulant $\kappa_1 = E\left[R_1\right] + E\left[R_2\right] + E\left[R_3\right]$ is written as

$$\kappa_1 = \alpha^4 + \left(nh\right)^{-1}\left\{-\frac{1}{6}\alpha^{444} - \omega^{kl}\gamma^{l;4,k} + \frac{1}{2}\omega^{km}\omega^{lm}\gamma^{4,kl}\right\} + O\left(\left(nh\right)^{-2} + \left(nh\right)^{-1}h^2 + h^4\right).$$

A.4.2 2nd Cumulant

Observe that

$$\begin{aligned}\kappa_2 &= E\left[R^2\right]-(E\left[R\right])^2\\ &= \left\{E\left[R_1^2\right]-(E\left[R_1\right])^2\right\}+E\left[R_2^2\right]+2\left\{E\left[R_2R_1\right]-E\left[R_1\right]E\left[R_2\right]\right\}+2E\left[R_3R_1\right]-(E\left[R_2\right])^2\\ &\quad+O\left((nh)^{-3}+(nh)^{-2}h^2\right).\end{aligned} \tag{26}$$

The first term of (26) satisfies

$$E\left[R_1^2\right]-(E\left[R_1\right])^2=(nh)^{-1}+O\left((nh)^{-1}h^5\right).$$

The second term of (26) satisfies

$$\begin{aligned}&(nh)^2E\left[R_2^2\right]\\ =\ &\frac{1}{4}\alpha^{4444}-\frac{1}{6}\alpha^{444}\alpha^{444}-\frac{1}{4}+\frac{1}{3}\omega^{kl}\gamma^{4,k;l}\alpha^{444}+\omega^{km}\omega^{lm}\left(\gamma^{4,k;4,l}-\gamma^{4,l;4}\gamma^{4,k;4}-\frac{1}{6}\gamma^{4,kl}\alpha^{444}\right)\\ &+\omega^{kl}\omega^{mn}\left(\gamma^{4,k;l}\gamma^{4,m;n}+\gamma^{4,k;n}\gamma^{4,m;l}\right)+\omega^{kl}\omega^{mn}\omega^{pn}\left(-\gamma^{4,k;l}\gamma^{4,mp}-2\gamma^{4,kp}\gamma^{4,m;l}\right)\\ &+\frac{1}{4}\left(\omega^{km}\omega^{lm}\omega^{pr}\omega^{qr}+\omega^{km}\omega^{ln}\omega^{pm}\omega^{qn}+\omega^{km}\omega^{ln}\omega^{pn}\omega^{qm}\right)\gamma^{4,kl}\gamma^{4,pq}+O\left((nh)^{-1}+h^2\right).\end{aligned}$$

The third term of (26) is

$$\begin{aligned}&E\left[R_2R_1\right]-E\left[R_1\right]E\left[R_2\right]\\ =\ &(nh)^{-1}\left\{\frac{1}{6}\alpha^4\alpha^{444}+\alpha^4\omega^{kl}\gamma^{l;4,k}-\frac{1}{2}\alpha^4\omega^{km}\omega^{lm}\gamma^{4,kl}\right\}\\ &+(nh)^{-2}\left\{-\frac{1}{2}\alpha^{4444}+\frac{1}{2}+\frac{1}{3}\alpha^{444}\alpha^{444}-\gamma^{l;4;4,k}\omega^{kl}+\frac{1}{2}\gamma^{4,kl}\omega^{km}\omega^{ln}\alpha^{mn4}+\gamma^{4;4,l}\omega^{lm}\alpha^{m44}\right\}\\ &+O\left((nh)^{-3}+(nh)^{-2}h^2\right).\end{aligned}$$

The fourth term of (26) is

$$\begin{aligned}&(nh)^2E\left[R_3R_1\right]\\ =\ &\omega^{kl}\omega^{mn}\left(\gamma^{n,k;4}\gamma^{4,m;l}+\gamma^{n,k;l}\gamma^{4,m;4}\right)+\frac{1}{2}\left(\gamma^{4,k;l}\alpha^{444}+\gamma^{4,k;4}\alpha^{44l}\right)-\frac{1}{2}\omega^{kl}\omega^{ml}\left(\gamma^{4,k;4,m}+2\gamma^{4,k;4}\gamma^{4,m;4}\right)\\ &+\frac{3}{8}\left(\alpha^{4444}+2\alpha^{444}\alpha^{444}-1\right)+\omega^{kl}\left(\gamma^{4,k;4,l}+2\gamma^{4,k;4}\alpha^{l44}\right)+3\omega^{kl}\gamma^{4;4,k}\alpha^{l44}\\ &-\frac{3}{2}\omega^{kl}\omega^{ml}\gamma^{4;4,k}\gamma^{4;4,m}+\frac{4}{3}\alpha^{444}\alpha^{444}-\frac{3}{4}\alpha^{4444}\\ &+\omega^{mp}\omega^{np}\left\{\begin{array}{c}-\frac{1}{2}\omega^{kl}\omega^{nl}\gamma^{4,km}\gamma^{4,on}+\omega^{kl}\gamma^{4,km}\left(\gamma^{l,m;4}+\gamma^{l;4,m}\right)+\frac{1}{2}\omega^{kl}\gamma^{l,mn}\gamma^{4;4,k}\\ +\frac{1}{3}\alpha^{444}\gamma^{4,mn}-\frac{1}{2}\gamma^{4;4,mn}-\frac{1}{2}\gamma^{4,m;4,n}+\frac{3}{2}\gamma^{4;4,m}\gamma^{4;4,n}\end{array}\right\}\\ &-\frac{1}{2}\omega^{kn}\omega^{ln}\omega^{pm}\gamma^{m,kl}\gamma^{4,p;4}+\frac{1}{4}\omega^{kn}\omega^{ln}\gamma^{4,kl}\alpha^{444}+\omega^{km}\omega^{nm}\omega^{lo}\gamma^{4,kl}\gamma^{4,n;o}-\omega^{kn}\omega^{lo}\gamma^{4,kl}\alpha^{mo4}\\ &-\omega^{kp}\omega^{ln}\omega^{mn}\gamma^{4,kl}\gamma^{p,n;4}+\frac{1}{2}\omega^{kn}\omega^{ln}\gamma^{4,kl;4}+\alpha^{4444}+3\omega^{lm}\omega^{nm}\gamma^{4;4,l}\gamma^{4,n;4}-4\omega^{lm}\gamma^{4;4,l}\alpha^{m44}\\ &-\omega^{ln}\omega^{mo}\gamma^{4,4,l}\gamma^{n,m;o}-\omega^{ln}\omega^{mk}\left(\gamma^{k;4,l}+\gamma^{4;k,l}\right)\gamma^{4,m;n}-\omega^{ln}\omega^{mn}\gamma^{4;4,l}\gamma^{4,m;4}\\ &-\omega^{ln}\gamma^{4;4,l}\alpha^{n44}+\omega^{lm}\gamma^{4;4,l;m}-3\omega^{mk}\alpha^{k44}\gamma^{4,m;4}-\frac{2}{3}\omega^{lm}\gamma^{4,l;m}\alpha^{444}-\frac{5}{2}\alpha^{444}\alpha^{444}+O\left((nh)^{-1}+h^2\right).\end{aligned}$$

Combining these results,

$$\kappa_2 = (nh)^{-1}+(nh)^{-1}\left\{\frac{1}{3}\alpha^4\alpha^{444}+2\alpha^4\omega^{kl}\gamma^{l;4,k}-\alpha^4\omega^{km}\omega^{lm}\gamma^{4,kl}\right\}+(nh)^{-2}\Delta+O\left((nh)^{-3}+(nh)^{-2}h^2\right),$$

where

$$\begin{aligned}\Delta &= \frac{1}{2}\alpha^{4444}-\frac{13}{36}\alpha^{444}\alpha^{444}+2\omega^{kl}\gamma^{l;4;4,k}-\omega^{km}\omega^{ln}\gamma^{4,kl}\alpha^{mn4}-\omega^{lm}\gamma^{4;4,l}\alpha^{m44}\\ &\quad+\omega^{km}\omega^{lm}\left(-\gamma^{4,k;4,l}+\frac{1}{6}\gamma^{4,kl}\alpha^{444}\right)-\frac{1}{3}\omega^{kl}\gamma^{4,k;l}\alpha^{444}+2\omega^{kl}\omega^{mn}\omega^{pn}\gamma^{4,p;l}\gamma^{4,km}\\ &\quad-\frac{1}{2}\omega^{km}\omega^{ln}\omega^{pm}\omega^{qn}\gamma^{4,kl}\gamma^{4,pq}+\omega^{kl}\omega^{ml}\gamma^{4;4,k}\gamma^{4,m;4}. \end{aligned} \tag{27}$$

A.4.3 3rd Cumulant

Using the results to derive the first and second cumulants, the third cumulant is written as

$$\begin{aligned}\kappa_3 &= E\left[R^3\right]-3E[R]E\left[R^2\right]+2(E[R])^3\\ &= E\left[(R_1+R_2)^3\right]-3E[R_1+R_2]E\left[(R_1+R_2)^2\right]+O\left((nh)^{-3}+(nh)^{-2}h^2\right)\\ &= \left\{E\left[R_1^3\right]-3E[R_1]E\left[R_1^2\right]\right\}-3E[R_2]E\left[R_1^2\right]+3E\left[R_2R_1^2\right]+O\left((nh)^{-3}+(nh)^{-2}h^2\right). \end{aligned}\tag{28}$$

The first term of (28) satisfies

$$\left\{E\left[R_1^3\right]-3E[R_1]E\left[R_1^2\right]\right\}=(nh)^{-2}\alpha^{444}+O\left((nh)^{-3}+(nh)^{-2}h^2\right).$$

The second term of (28) satisfies

$$-3E[R_2]E\left[R_1^2\right]=(nh)^{-2}\left(\frac{1}{2}\alpha^{444}+3\omega^{kl}\gamma^{l;4,k}-\frac{3}{2}\gamma^{4,kl}\omega^{km}\omega^{lm}\right)+O\left((nh)^{-3}+(nh)^{-2}h^2\right).$$

The third term of (28) satisfies

$$3E\left[R_2R_1^2\right]=(nh)^{-2}\left(-\frac{3}{2}\alpha^{444}-3\omega^{kl}\gamma^{l;4,k}+\frac{3}{2}\gamma^{4,kl}\omega^{km}\omega^{lm}\right)+O\left((nh)^{-3}+(nh)^{-2}h^2\right).$$

Combining these results, we obtain $\kappa_3 = O\left((nh)^{-3}+(nh)^{-2}h^2\right)$.

A.4.4 4th Cumulant

In this subsection, let $t_1=\alpha^{4444}$, $t_2=3$, $t_3=4\left(\alpha^{444}\right)^2$, and $t_4=3\left(\alpha^{444}\right)^2$. Using the results to obtain the first, second, and third cumulants,

$$\begin{aligned}\kappa_4 &= E\left[R^4\right]-3\left(E\left[R^2\right]\right)^2-4E[R]E\left[R^3\right]+12(E[R])^2E\left[R^2\right]-6(E[R])^4\\ &= \left\{E\left[R_1^4\right]-3\left(E\left[R_1^2\right]\right)^2-4E[R_1]E\left[R_1^3\right]+12(E[R_1])^2E\left[R_1^2\right]-6(E[R_1])^4\right\}\\ &\quad+\left\{4E\left[R_2R_1^3\right]-12E[R_2R_1]E\left[R_1^2\right]-12E\left[R_2R_1^2\right]E[R_1]\right\}\\ &\quad+\left\{6E\left[R_2^2R_1^2\right]-E\left[R_2^2\right]E\left[R_1^2\right]\right\}\\ &\quad+\left\{4E\left[R_3R_1^3\right]-12E[R_3R_1]E\left[R_1^2\right]\right\}-\left\{4E[R_2]E\left[R_1^3\right]-12E[R_2]E\left[R_2R_1^2\right]+12(E[R_2])^2E\left[R_1^2\right]\right\}\\ &\quad+O\left((nh)^{-4}+(nh)^{-3}h^2\right). \end{aligned}\tag{29}$$

The first term of (29) satisfies

$$(nh)^3\left\{E\left[R_1^4\right]-3\left(E\left[R_1^2\right]\right)^2-4E\left[R_1\right]E\left[R_1^3\right]+12\left(E\left[R_1\right]\right)^2E\left[R_1^2\right]-6\left(E\left[R_1\right]\right)^4\right\}=t_1-t_2+O\left((nh)^{-1}+h^2\right).$$

The second term of (29) satisfies

$$\begin{aligned}&(nh)^3\left\{4E\left[R_2R_1^3\right]-12E\left[R_2R_1\right]E\left[R_1^2\right]-12E\left[R_2R_1^2\right]E\left[R_1\right]\right\}\\ =\;&-6t_1+2t_2-\frac{1}{6}t_3+\frac{2}{3}t_4+2\gamma^{4,kl}\omega^{km}\omega^{lm}\alpha^{444}-4\omega^{kl}\gamma^{4,k:l}\alpha^{444}+O\left((nh)^{-1}+h^2\right).\end{aligned}$$

The third term of (29) satisfies

$$(nh)^3\left\{6E\left[R_2^2R_1^2\right]-E\left[R_2^2\right]E\left[R_1^2\right]\right\}=3t_1-t_2+\frac{1}{6}t_3-\frac{5}{9}t_4+4\omega^{kl}\gamma^{4,k:l}\alpha^{444}-2\gamma^{4,kl}\omega^{km}\omega^{lm}\alpha^{444}+O\left((nh)^{-1}+h^2\right).$$

The fourth term of (29) satisfies

$$(nh)^3\left\{4E\left[R_3R_1^3\right]-12E\left[R_3R_1\right]E\left[R_1^2\right]\right\}=2t_1-\frac{1}{9}t_4+O\left((nh)^{-1}+h^2\right).$$

Using the results to derive the first, second, and third cumulants, the fifth term of (29) is of order $O\left((nh)^{-4}+(nh)^{-3}h^2\right)$. Combining these results, we obtain $\kappa_4=O\left((nh)^{-4}+(nh)^{-3}h^2\right)$.

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] Chen, S. X. (1997) Empirical likelihood-based kernel density estimation, *Australian Journal of Statistics*, 39, 47-56.

[3] Chen, S. X. and H. Cui (2006) On Bartlett correction of empirical likelihood in the presence of nuisance parameters, *Biometrika*, 93, 215-220.

[4] Chen, S. X. and Y. S. Qin (2000) Empirical likelihood confidence intervals for local linear smoothers, *Biometrika*, 87, 946-953.

[5] Cheng, M. Y. (1994) On boundary effects of smooth curve estimators, Unpublished manuscript series # 2319, University of North Carolina.

[6] Cheng, M. Y. (1997) A bandwidth selector for local linear density estimators, *Annals of Statistics*, 25, 1001-1013.

[7] Cheng, M. Y., Fan, J. and J. S. Marron (1997) On automatic boundary corrections, *Annals of Statistics*, 25, 1691-1708.

[8] Cline, D. B. and J. D. Hart (1991) Kernel estimation of densities with discontinuities or discontinuous derivatives, *Statistics*, 22, 69-84.

[9] Copas, J. B. (1995) Local likelihood based on kernel censoring, *Journal of the Royal Statistical Society*, B, 57, 221-235.

[10] Davidson, R. and J. G. MacKinnon (1998) Graphical methods for investigating the size and power of test statistics, *Manchester School*, 66, 1-26.

[11] DiCiccio, T. J., Hall, P. and J. Romano (1991) Empirical likelihood is Bartlett-correctable, *Annals of Statistics*, 19, 1053-1061.

[12] Fan, J. and I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*, Chapman & Hall.

[13] Fan, J., Zhang, C. and J. Zhang (2001) Generalized likelihood ratio statistics and Wilks phenomenon, *Annals of Statistics*, 29, 153-193.

[14] Gregory, A. W. and M. R. Veal (1985) Formulating Wald tests of nonlinear restrictions, *Econometrica*, 53, 1465-68.

[15] Hahn, J., Todd, P. and W. van der Klaauw (2001) Identification and estimation of treatment effects with a regression discontinuity design, *Econometrica*, 69, 201-209.

[16] Hjort, N. L. and M. C. Jones (1996) Locally parametric nonparametric density estimation, *Annals of Statistics*, 24, 1619-1647.

[17] Imbens, G. W. and T. Lemieux (2008) Regression discontinuity designs: a guide to practice, *Journal of Econometrics*, 142, 615-635.

[18] Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent data, *Annals of Statistics*, 25, 2084-2102.

[19] Lee, D. S. (2008) Randomized experiments from non-random selection in U.S. House elections, *Journal of Econometrics*, 142, 675-697.

[20] Loader, C. R. (1996) Local likelihood density estimation, *Annals of Statistics*, 24, 1602-1618.

[21] Marron, J. S. and D. Ruppert (1994) Transformations to reduce boundary bias in kernel density estimation, *Journal of the Royal Statistical Society*, B, 56, 653-671.

[22] McCrary, J. (2008) Manipulation of the running variable in the regression discontinuity design: a density test, *Journal of Econometrics*, 142, 698-714.

[23] Newey, W. K. and R. J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica*, 72, 219-255.

[24] Owen, A. (2001) *Empirical Likelihood*, Chapman & Hall, New York.

[25] Park, B. U., Kim, W. C. and M. C. Jones (2002) On local likelihood density estimation, *Annals of Statistics*, 30, 1480-1495.

[26] Porter, J. (2003) Estimation in the regression discontinuity model, Working paper, Department of Economics, University of Wisconsin.

[27] Saez, E. (2010) Do taxpayers bunch at kink points?, *American Economic Journal: Economic Policy*, 2, 180-212.

[28] Smith, R. J. (1997) Alternative semi-parametric likelihood approaches to generalised method of moments estimation, *Economic Journal*, 107, 503-519.

[29] Xu, K.-L. and P. C. B. Phillips (2011) Tilted nonparametric estimation of volatility functions with empirical applications, *Journal of Business and Economic Statistics*, 29, 518-528.

Table 1: Finite-sample biases, standard deviations (Std.'s) and root mean square errors (RMSEs) of the binning and the local likelihood estimators of θ. The data are generated from a $N(12,3)$ density and the sample size $n = 1000$. The discontinuity point is $c = 13$.

		Fixed bandwidth				Data dependent bandwidth			
	Est.	$h=1$	$h=2$	$h=3$	$h=4$	$\alpha=1.5^{-1}$	$\alpha=1$	$\alpha=1.5$	$\alpha=1.5^2$
Bias	$\widetilde{\theta}^G$	$-.0403$	$-.0710$	$-.0886$	$-.0949$	$-.0462$	$-.0644$	$-.0825$	$-.0936$
	$\widehat{\theta}^G$	$-.0020$	$-.0148$	$-.0360$	$-.0624$	$-.0039$	$-.0108$	$-.0266$	$-.0577$
	$\widetilde{\theta}$	$-.0408$	$-.0725$	$-.0907$	$-.0974$	$-.0478$	$-.0662$	$-.0841$	$-.0959$
	$\widehat{\theta}$	.0016	$-.0032$	$-.0094$	$-.0221$	$-.0024$	$-.0036$	$-.0057$	$-.0194$
Std.	$\widetilde{\theta}^G$	.0236	.0157	.0127	.0104	.0227	.0188	.0149	.0107
	$\widehat{\theta}^G$	.0468	.0312	.0254	.0217	.0409	.0345	.0296	.0262
	$\widetilde{\theta}$	.0232	.0155	.0124	.0103	.0228	.0185	.0151	.0105
	$\widehat{\theta}$	.0452	.0326	.0292	.0283	.0425	.0354	.0316	.0291
RMSE	$\widetilde{\theta}^G$	.0467	.0728	.0895	.0955	.0514	.0671	.0838	.0942
	$\widehat{\theta}^G$	.0468	.0345	.0441	.0661	.0411	.0362	.0398	.0633
	$\widetilde{\theta}$	.0469	.0742	.0915	.0979	.0530	.0688	.0854	.0965
	$\widehat{\theta}$	.0453	.0328	.0307	.0359	.0426	.0356	.0321	.0350
						h			
						Mean	1.7257		
						Std.	0.1975		

Table 2: Finite-sample biases, standard deviations (Std.'s) and root mean square errors (RMSEs) of the binning and the local likelihood estimators of θ. The data are generated from a Student's t density (i.e., $12 + t\left(5\right)/\sqrt{5/9}$) and the sample size $n = 1000$. The discontinuity point is $c = 13$.

		Fixed bandwidth				Data dependent bandwidth			
	Est.	$h=1$	$h=2$	$h=3$	$h=4$	$\alpha=1.5^{-1}$	$\alpha=1$	$\alpha=1.5$	$\alpha=1.5^2$
Bias	$\widetilde{\theta}^G$	$-.0759$	$-.1208$	$-.1355$	$-.1315$	$-.0887$	$-.1163$	$-.1319$	$-.1276$
	$\widehat{\theta}^G$	$-.0083$	$-.0394$	$-.0876$	$-.1254$	$-.0120$	$-.0369$	$-.0790$	$-.1274$
	$\widetilde{\theta}$	$-.0762$	$-.1223$	$-.1378$	$-.1343$	$-.0899$	$-.1180$	$-.1343$	$-.1305$
	$\widehat{\theta}$	.0036	$-.0183$	$-.0492$	$-.0788$	$-.0060$	$-.0184$	$-.0442$	$-.0832$
Std.	$\widetilde{\theta}^G$	.0235	.0166	.0130	.0100	.0267	.0209	.0138	.0120
	$\widehat{\theta}^G$	.0455	.0333	.0262	.0219	.0444	.0399	.0381	.0305
	$\widetilde{\theta}$	.0232	.0163	.0130	.0096	.0254	.0199	.0136	.0119
	$\widehat{\theta}$	.0456	.0358	.0320	.0296	.0421	.0374	.0377	.0380
RMSE	$\widetilde{\theta}^G$	.0795	.1220	.1361	.1319	.0926	.1182	.1326	.1282
	$\widehat{\theta}^G$	.0462	.0516	.0915	.1273	.0460	.0543	.0877	.1310
	$\widetilde{\theta}$	.0796	.1234	.1385	.1346	.0934	.1197	.1350	.1311
	$\widehat{\theta}$	.0457	.0402	.0587	.0842	.0425	.0417	.0581	.0914
						h			
						Mean	1.9132		
						Std.	0.3560		

Table 3: Finite-sample sizes, quantiles and powers of the $W^G(\theta)$, $\ell^G(\theta)$, and $\ell(\theta)$ tests of density continuity, i.e., $H_0 : \theta_0 = 0$ (nominal sizes: 5% and 10%). The powers are calculated when the data are generated from a mixture of left and right truncated normal distributions at c with probability γ, where $\gamma = \Phi(c) - d$ with $d \in \{0.02, 0.04, 0.06, 0.08, 0.10\}$

					Power (vs. value of d)				
n	Test	Size	Finite Sample Quantile	Asymptotic Quantile	0.02	0.04	0.06	0.08	0.10
1000	W^G, 5%	.073	4.51	3.84	.058	.104	.202	.368	.545
	ℓ^G, 5%	.067	4.19		.059	.139	.244	.367	.491
	ℓ, 5%	.067	4.29		.082	.190	.366	.578	.754
	W^G, 10%	.138	3.27	2.71	.107	.176	.303	.489	.651
	ℓ^G, 10%	.137	3.19		.131	.211	.353	.505	.648
	ℓ, 10%	.110	2.89		.152	.273	.474	.681	.841
2000	W^G, 5%	.074	4.53	3.84	.071	.191	.394	.642	.856
	ℓ^G, 5%	.047	3.79		.072	.191	.418	.636	.800
	ℓ, 5%	.050	3.82		.112	.306	.604	.845	.973
	W^G, 10%	.134	3.18	2.71	.126	.261	.530	.754	.924
	ℓ^G, 10%	.125	2.98		.144	.280	.541	.753	.909
	ℓ, 10%	.104	2.75		.184	.418	.715	.898	.983
5000	W^G, 5%	.065	4.31	3.84	.108	.427	.796	.960	.995
	ℓ^G, 5%	.038	3.47		.112	.423	.769	.932	.987
	ℓ, 5%	.062	4.54		.193	.593	.900	.990	1.00
	W^G, 10%	.127	3.07	2.71	.185	.548	.864	.983	.998
	ℓ^G, 10%	.099	2.64		.185	.545	.861	.977	.995
	ℓ, 10%	.112	2.85		.282	.690	.952	.996	1.00

Table 4: Estimation and testing of the discontinuity of the density of enrollments at multiples of 40 (according to Maimonides's rule, Angrist and Lavy, 1999) for fifth graders. The binning and local likelihood methods are used with various smoothing bandwidths.

		Binning method					Local likelihood method			
c	h	$\widehat{f_l}$	$\widehat{f_r}$	$\widehat{\theta}^G$	Wald W^G	EL ℓ^G	$\widehat{f_l}$	$\widehat{f_r}$	$\widehat{\theta}$	EL ℓ
40	15	.0056	.0080	.0024	2.666	0.444	.0039	.0114	.0075	20.67
	20	.0052	.0089	.0037	7.888	1.331	.0040	.0114	.0074	26.94
	25	.0050	.0094	.0044	13.57	2.321	.0040	.0114	.0074	33.28
	30	.0054	.0099	.0045	16.23	2.856	.0045	.0116	.0072	36.36
80	15	.0115	.0095	−.0019	1.141	1.014	.0081	.0140	.0059	8.301
	20	.0112	.0089	−.0024	2.389	1.964	.0085	.0116	.0030	3.366
	25	.0108	.0087	−.0021	2.395	1.534	.0087	.0107	.0021	2.092
	30	.0105	.0089	−.0016	1.698	1.001	.0088	.0107	.0020	2.350
120	15	.0092	.0050	−.0043	7.920	3.906	.0064	.0078	.0014	0.577
	20	.0088	.0048	−.0039	9.355	3.977	.0066	.0070	.0003	0.045
	25	.0075	.0047	−.0029	7.005	2.403	.0060	.0063	.0003	0.069
	30	.0066	.0045	−.0021	4.988	1.263	.0055	.0060	.0005	0.178
160	15	.0027	.0010	−.0017	4.670	3.253	.0017	.0013	−.0003	0.142
	20	.0025	.0010	−.0015	5.176	2.397	.0018	.0012	−.0006	0.730
	25	.0023	.0010	−.0012	4.550	1.710	.0017	.0013	−.0005	0.586
	30	.0021	.0011	−.0011	4.316	1.456	.0017	.0013	−.0004	0.473

Table 5: Empirical likelihood confidence sets (EL CSs) of the discontinuity of the density of enrollments at $c = 40$ for fifth graders. The binning and local likelihood methods are used with various smoothing bandwidths.

	Binning method			Local likelihood method		
h	$\widehat{\theta}^G$	EL CS	Length	$\widehat{\theta}$	EL CS	Length
15	.0024	[−.0104, .0152]	.0256	.0075	[.0046, .0106]	.0060
20	.0037	[−.0036, .0120]	.0156	.0074	[.0049, .0101]	.0052
25	.0044	[−.0014, .0108]	.0122	.0074	[.0051, .0097]	.0046
30	.0045	[−.0007, .0106]	.0113	.0072	[.0051, .0096]	.0045

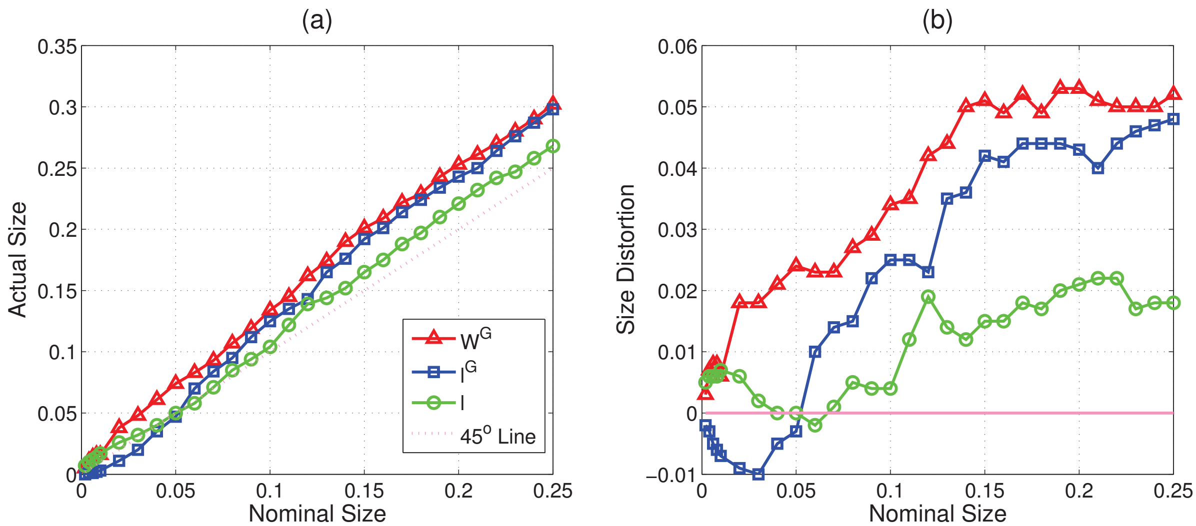

Figure 1: (a) P-value plots and (b) P-value discrepancy plots (Davidson and MacKinnon, 1998) for W^G, ℓ^G and ℓ tests when $n = 2000$.

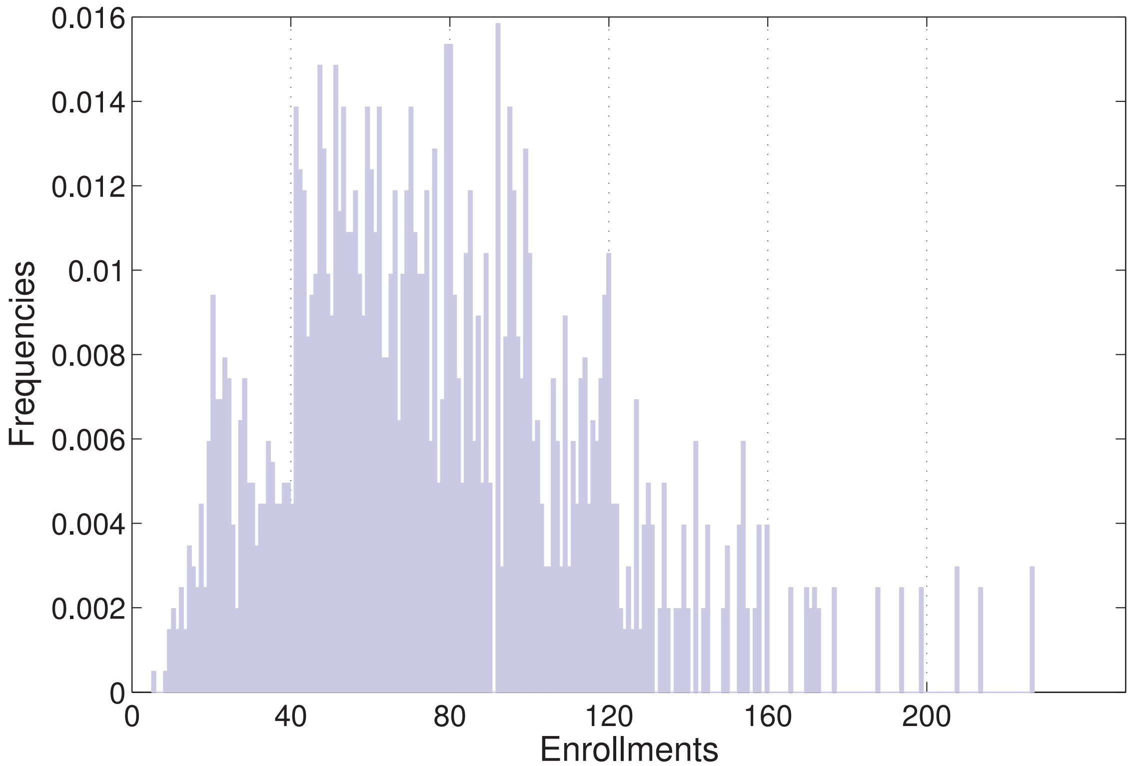

Figure 2: Histogram of the enrollments of 2029 classes in Grade 5 (Data: Angrist and Lavy, 1999).

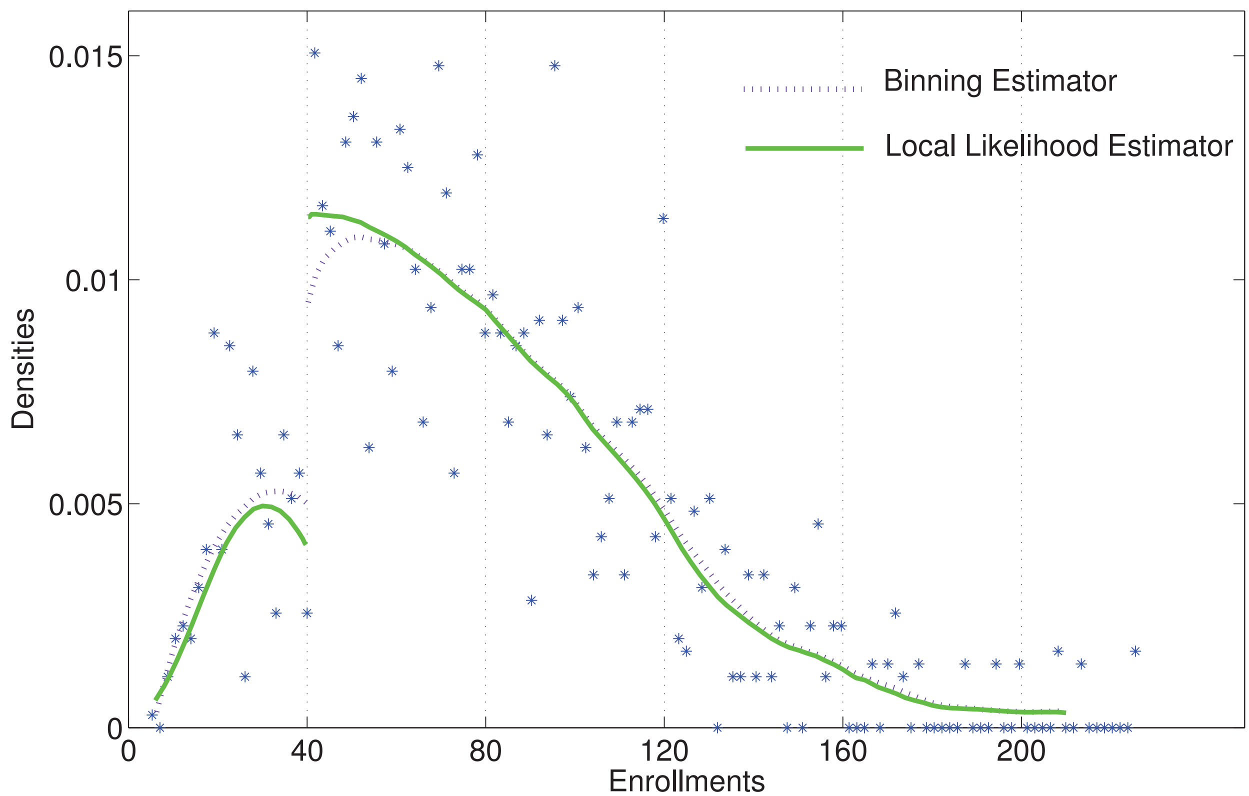

Figure 3: The estimated density function of school enrollments for the fifth graders using the data on left and right sides of $c = 40$. Binned data are also displayed.

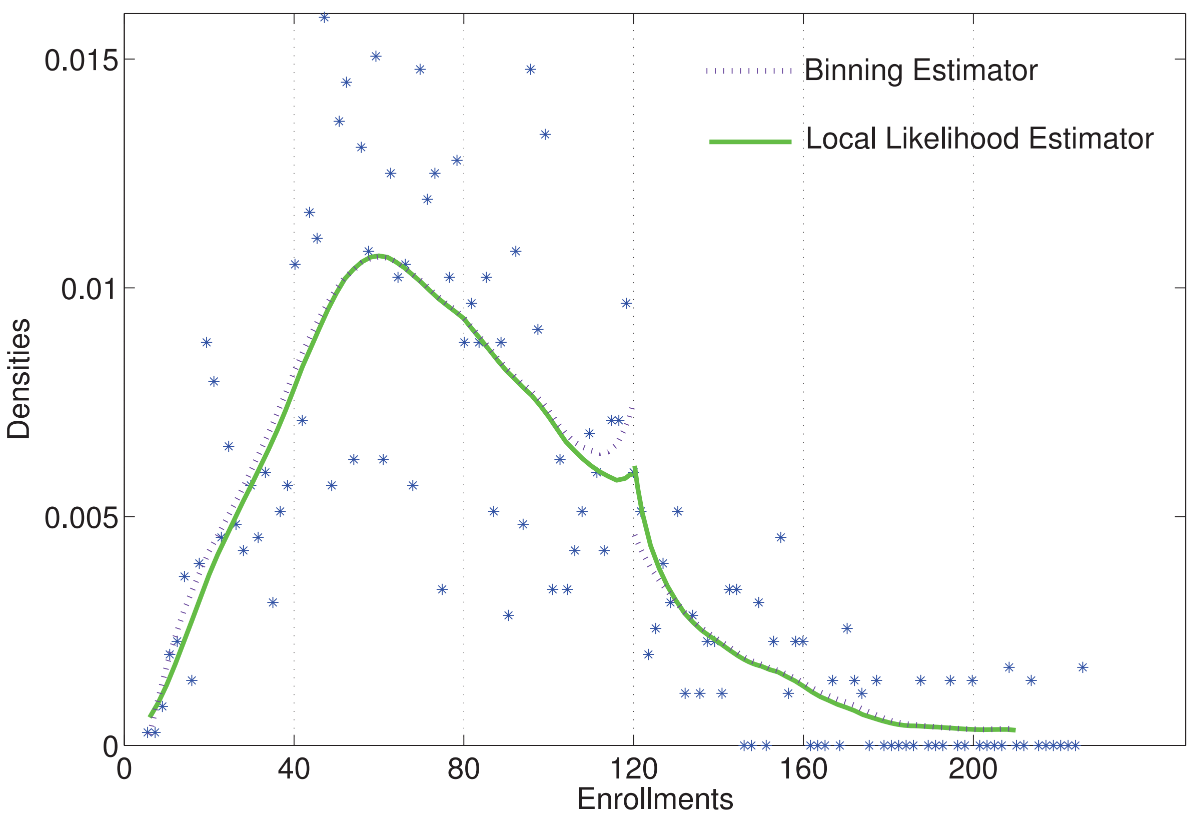

Figure 4: The estimated density function of school enrollments for the fifth graders using the data on left and right sides of $c = 120$. Binned data are also displayed.