Econometric Reviews, 35(4):586–637, 2016

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474938.2013.833831

Understanding Estimators of Treatment Effects in Regression Discontinuity Designs

Ping Yu
Department of Economics, University of Auckland, Auckland, New Zealand

In this paper, we propose two new estimators of treatment effects in regression discontinuity designs. These estimators can aid understanding of the existing estimators such as the local polynomial estimator and the partially linear estimator. The first estimator is the partially polynomial estimator which extends the partially linear estimator by further incorporating derivative differences of the conditional mean of the outcome on the two sides of the discontinuity point. This estimator is related to the local polynomial estimator by a relocalization effect. Unlike the partially linear estimator, this estimator can achieve the optimal rate of convergence even under broader regularity conditions. The second estimator is an instrumental variable estimator in the fuzzy design. This estimator will reduce to the local polynomial estimator if higher order endogeneities are neglected. We study the asymptotic properties of these two estimators and conduct simulation studies to confirm the theoretical analysis.

Keywords Instrumental variable estimator; Local polynomial estimator; Optimal rate of convergence; Partially linear estimator; Partially polynomial estimator; Regression discontinuity design.

JEL Classification C13; C14; C21.

1. INTRODUCTION

The regression discontinuity design (RDD) has got much popularity in applied econometric practice for identifying treatment effects since its introduction by Thistlewaite and Campbell (1960). Classical applications include Angrist and Lavy (1999), Battistin and Rettore (2002), Black (1999), Card et al. (2008), Chay and Greenstone (2005), Chay et al. (2005), Dell (2010), DesJardins and McCall (2008), DiNardo and Lee (2004), Jacob and Lefgren (2004), Lee (2008), Ludwig and Miller (2007), Pence (2006), and Van der Klaauw (2002) among others. See Cook (2008) for a historical

Address correspondence to Ping Yu, Department of Economics, University of Auckland, Owen G Glenn Building, 12 Grafton Road, Auckland 1142, New Zealand; E-mail: p.yu@auckland.ac.nz
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lecr.

introduction of RDDs in three academic disciplines, and see Imbens and Lemieux (2008), Lee and Lemieux (2010), and Van der Klaauw (2008) for excellent reviews on up-to-date theoretical developments and applications.

We know human behaviors always evolve smoothly unless an abrupt change happens exogenously. This observation lies in the heart of RDDs. Suppose a treatment t is given based on an observed forcing variable x by

$$t = \begin{cases} T_1, & \text{if } x \geq \pi, \\ T_0, & \text{if } x < \pi, \end{cases}$$

where the cut-off point π is known, and both T_0 and T_1 follow the Bernoulli distribution with different conditional means especially at $x = \pi$. Let Y_1 and Y_0 be the potential outcomes corresponding to the two treatment assignments, then the observed outcome is $y = tY_1 + (1-t)Y_0$. Trochim (1984) divides RDDs into the sharp design and fuzzy design depending on t being a deterministic function of x or not. In the sharp design, the treatment assignment $T_1 = 1$ and $T_0 = 0$ almost surely. Hahn et al. (2001) show that if $E[Y_0 \mid x]$ and $E[Y_1|x]$ are continuous at π, then in the left and right neighborhoods of the threshold π, the treatment is assigned as if in a randomized experimental design. So the individuals marginally below the threshold represent a valid counterfactual for the treated group just above the threshold. As a result, the expected causal effect of the treatment can be identified as

$$\alpha \equiv E[Y_1 - Y_0 \mid x = \pi] = E[y \mid x = \pi+] - E[y \mid x = \pi-],$$

where $E[y \mid x = \pi+] = \lim_{x \downarrow \pi} E[y \mid x]$, and $E[y \mid x = \pi-] = \lim_{x \uparrow \pi} E[y \mid x]$. In the fuzzy design, T_1 and T_0 are random, but the propensity scores $E[T_1 \mid x = \pi+] \neq E[T_0|x = \pi-]$. In this case, Hahn et al. (2001) show that α can be identified under a further local unconfoundedness condition. Specifically,

$$\alpha \equiv E[Y_1 - Y_0 \mid x = \pi] = \frac{E[y \mid x = \pi+] - E[y \mid x = \pi-]}{E[t \mid x = \pi+] - E[t \mid x = \pi-]}.$$

In both cases, α only involves the difference of two estimable conditional means.

Until today, many estimators of treatment effects in RDDs have been developed. Hahn et al. (2001) and Porter (2003) notice the bias problem in the Nadaraya–Watson estimator (NWE) of $E[y \mid x = \pi+]$, $E[y \mid x = \pi-]$, $E[t \mid x = \pi+]$ and $E[t \mid x = \pi-]$, and the former suggests to use the local linear estimator (LLE), while the latter suggests to use the local polynomial estimator (LPE) which generalizes the LLE. Porter (2003) also puts forward another estimator called the partially linear estimator (PLE). He shows that the PLE can achieve the optimal rate of convergence only under the more stringent assumption (Assumption 2(b) of Porter, 2003) on the data generating process (DGP), while the LPE can achieve the optimal rate under a broader assumption (Assumption 2(a) of Porter,

2003) on the DGP. There is an immediate logic gap: what is the relationship between the PLE and the LPE? Why cannot the PLE achieve the optimal rate under the broader assumption? To shed light on these questions, this paper puts forward a new estimator called the partially polynomial estimator (PPE) which builds a connection between the LPE and PLE. This estimator generalizes the PLE by considering derivative differences of $E[y \mid x]$ (besides the level difference as in the PLE) on the two sides of the threshold π. By locally putting the PPE and the LPE in the threshold regression framework, we show that the PPE can be generated by imposing a relocalization effect on the LPE, and it can achieve the optimal rate of convergence just as the LPE. The second contribution of this paper is to provide a new instrumental variable estimator (IVE) in the fuzzy design. This estimator can aid understanding of the LPE in the fuzzy design. It is well known that the LPE in the fuzzy design can solve the endogeneity problem without introducing extra instrumental variables, but from the construction of the LPE, it is hard to see why endogeneity is even involved. Hahn et al. (2001) interpret the LPE as the Wald estimator in a simple case. Imbens and Lemieux (2008) and Lee and Lemieux (2010) reexpress the LPE in the general case as a 2SLS estimator so that it can be treated as a generalization of the Wald estimator. Despite the numerical equivalence between the LPE and the 2SLS estimator, it is still unclear where the endogeneity is from and how it is eliminated by the LPE. To understand this problem, we localize the model in the neighborhood of the threshold and put forward the new IVE. It is shown that the LPE is constructed essentially by neglecting higher order endogeneities in the model, while the IVE eliminates such endogeneities directly. We derive the asymptotic distributions of the PPE and IVE, and also conduct some simulation studies to confirm the theoretical analysis. A technical contribution of this paper is to linearize the LPE. In the literature where a nonparametric estimator of conditional mean is used as an intermediate input, the local constant estimator combined with a higher order kernel is often employed for technical convenience. In this paper, we show that the LPE can also be used and is asymptotically equivalent to a higher order kernel estimator, while regularity conditions can be somewhat weakened.

The rest of this paper is organized as follows. In Section 2, we construct the PPE in the sharp design, discuss its relationship with the LPE, derive its asymptotic distributions, and provide a variance estimator. Section 3 discusses two new estimators in the fuzzy design: the PPE and the new IVE. Section 4 includes some simulation studies and Section 5 concludes. The proof of theorems and related lemmas are given in three appendices. A word on notations: $\approx$ means the higher-order terms are omitted or a constant term is omitted (depending on the context). Since the LPE of $m(x) \equiv E\left[y_i \mid x_i = x\right]$ (for a response y and an interior point x on the support of x_i) is the building block of the PPE, we here review its main properties and define necessary notations for the following development. From Fan and Gijbels (1996), the pth order LPE of $m(x)$ is a linear functional of

$\mathbf{y} \equiv (y_1, \ldots, y_n)'$:

$$\mathbb{P}_x^n(\mathbf{y}) = \sum_{j=1}^{n} W_j^n(x) y_j, \tag{1}$$

where $W_j^n(x)$ is a weight function depending on the rescaled kernel $k_h(\cdot)$, $\{x_i\}_{i=1}^n$ and x, and $\sum_{j=1}^n W_j^n(x) = 1$. $k_h(\cdot) = \frac{1}{h} k\left(\frac{\cdot}{h}\right)$ with $k(\cdot)$ being a kernel density and h being the bandwidth. The notation $\mathbb{P}$ is due to the fact that the LPE can be treated as a projection estimator; see Mammen et al. (2001). As shown in Lemma 2.1 of Fan et al. (1997), $\mathbb{P}_x^n$ is equivalent to the linear functional $\mathbb{P}_x$ asymptotically:

$$\mathbb{P}_x(\mathbf{y}) = \frac{1}{nhf(x)} \sum_{j=1}^{n} K_p^* \left(\frac{x_j - x}{h}\right) y_j,$$

where $f(\cdot)$ is the density of x_i,

$$K_p^*(u) = e_1' \Gamma^{-1} (1, u, \ldots, u^p)' k(u) \equiv e_1' \Gamma^{-1} \delta(u), \tag{2}$$

is a kernel of order $p+1$ when p is odd and of order $p+2$ when p is even as defined by Gasser et al. (1985),[1] $e_1 = (1, 0, \ldots, 0)'_{(p+1)\times 1}$ whose dimension is determined by the context without further explanation, $\Gamma = (\gamma_{i+j-2})_{1 \le i,j \le p+1}$ is invertible with $\gamma_j = \int u^j k(u) du$, and $\delta(u) = (k(u), uk(u), \ldots, u^p k(u))'$.

2. PARTIALLY POLYNOMIAL ESTIMATION IN THE SHARP DESIGN

This section discusses the partially polynomial estimation in the sharp design. We first review the existing estimators in the literature, then discuss the construction of the PPE and its connection with the LPE, and conclude with the asymptotic theory of the PPE and its variance estimation.

2.1. The Existing Estimators in the Sharp Design

In RDDs, the outcome equation is

$$y = \underline{m}(x) + \alpha(x)t + \varepsilon,$$

where the forcing variable x is some basic determinant of the outcome, $\underline{m}(x)$ is the baseline effect and assumed to be continuous, t is the treatment status, ε may depend on t and is denoted as ε_t if necessary, and $E[\varepsilon|x,t] = 0$, $t = 0, 1$. The treatment effect at

[1]Or equivalently, as shown in Ruppert and Wand (1994), $K_p^*(u) = |\Gamma(u)| / |\Gamma| \, k(u)$, where $\Gamma(u)$ is the same as Γ, but with the first column replaced by $(1, u, \ldots, u^p)'$. When $p = 0$ and 1, $K_p^*(u) = k(u)$ if k is symmetric.

x is $\alpha(x) + \varepsilon_1 - \varepsilon_0$, where $\alpha(x)$ is the average treatment effect at x and is assumed to be continuous in x. In the sharp design, $t = d \equiv 1(x \geq \pi)$, where $1(A)$ is the indicator function with value 1 when the event A is true and 0 otherwise. We are interested in the average treatment effect at $x = \pi$, that is, $E[\alpha(x) + \varepsilon_1 - \varepsilon_0 | x = \pi] = \alpha(\pi) \equiv \alpha$. In the sharp design, the outcome equation can be written as

$$y = m_0(x) + \alpha d + \varepsilon \equiv m(x) + \varepsilon, \tag{3}$$

where $m(x) = E[y \,|\, x] = \underline{m}(x) + d\alpha(x)$, and $m_0(x) \equiv E[y \,|\, x] - \alpha d = \underline{m}(x) + d(\alpha(x) - \alpha)$ shifts $m(x)$ down by the size α on $x \geq \pi$ so is continuous.

Since Porter (2003), the benchmark estimator of α is the local polynomial estimator (LPE). In the sharp design, it is defined as

$$\tilde{\alpha} = \hat{m}_+(\pi) - \hat{m}_-(\pi), \tag{4}$$

where $\hat{m}_+(\pi)$ is the LPE of $m_+(\pi) \equiv E[y \,|\, x = \pi+] = \underline{m}(\pi) + \alpha$ and is determined by the minimizer $\hat{a}$ in the following problem:

$$\min_{a,b_1,\ldots,b_p} \frac{1}{n} \sum_{i=1}^{n} k_h(x_i - \pi)\, d_i \left[y_i - a - b_1(x_i - \pi) - \cdots - b_p(x_i - \pi)^p\right]^2,$$

where p is a nonnegative integer, and $d_i = 1(x_i \geq \pi)$. $\hat{m}_-(\pi)$ is the LPE of $m_-(\pi) \equiv E[y \,|\, x = \pi-] = \underline{m}(\pi)$ and is similarly defined as $\hat{m}_+(\pi)$ with d_i substituted by $d_i^c \equiv 1 - d_i$. When $p = 0$, $\tilde{\alpha}$ is the NWE. As argued in Section 3 of Hahn et al. (2001) and Section 3.2 of Porter (2003), this estimator suffers from the usual boundary problem in conditional mean estimation. Hahn et al. (2001) suggest $p = 1$, which results in the LLE. This estimator avoids the boundary problem of the NWE, and also shares some efficiency property as discussed in Fan (1992, 1993). Imbens and Lemieux (2008) and Lee and Lemieux (2010) also mention a related estimator based on the pooled regression:

$$\min_{a,\alpha,b_1,\beta_1} \frac{1}{n} \sum_{x_i \in N_0} \left[y_i - a - \alpha d_i - b_1(x_i - \pi) - \beta_1 d_i(x_i - \pi)\right]^2 \tag{5}$$

where $N_0 = [\pi - h, \pi + h]$. The resulting estimator of α is numerically equivalent to (4) when k is the uniform kernel and $p = 1$. This estimator can be easily extended to the case with $p > 1$ and a general kernel by considering the following minimization problem:

$$\min_{a,\alpha,b_1,\beta_1,\ldots,b_p,\beta_p} \frac{1}{n} \sum_{i=1}^{n} k_h(x_i - \pi)\left[y_i - a - \alpha d_i - b_1(x_i - \pi) - \beta_1 d_i(x_i - \pi) - \cdots - b_p(x_i - \pi)^p - \beta_p d_i(x_i - \pi)^p\right]^2. \tag{6}$$

A good property of this estimator is that the standard error of the estimated treatment effect can be directly obtained from the regression since the usual standard error of the least square estimation is valid (which will be shown as a corollary of Theorem 5 in Section 3.3). We label this estimator as the least squares estimator (LSE).

Another estimator put forward in Porter (2003) is the PLE. This estimator is motivated by the observation that (3) takes the partially linear form of Robinson (1988), so α can be treated as the parametric coefficient in the partially linear model. The PLE is defined as

$$\begin{aligned}&\arg\min_{\alpha}\sum_{i=1}^{n}\left[y_i-\alpha d_i-\sum_{j=1}^{n}w_j^i\left(y_j-\alpha d_j\right)\right]^2\\&=\arg\min_{\alpha}\sum_{i=1}^{n}\left[y_i-\sum_{j=1}^{n}w_j^i y_j-\alpha\left(d_i-\sum_{j=1}^{n}w_j^i d_j\right)\right]^2,\end{aligned}$$

where $w_j^i=k_h\left(x_i-x_j\right)/\sum_{l=1}^{n}k_h\left(x_i-x_l\right)$. $\sum_{j=1}^{n}w_j^i\left(y_j-\alpha d_j\right)$ can be treated as an estimator of $m_0(x)$ at x_i. Actually, the PLE in Robinson (1988) can be equivalently redefined in this way. Note that $d_i-\sum_{j=1}^{n}w_j^i d_j=0$ when x_i is out of a $O(h)$ neighborhood of π, so only the information in the h-neighborhood of π is used to estimate α. As a result, the PLE only has a nonparametric convergence rate instead of the $\sqrt{n}$ rate in Robinson (1988); see Section 3.3 of Porter (2003) for more discussions on this point.

Porter (2003) shows that the LPE can achieve the optimal rate of convergence for a general form of $m_0(x)$. However, the PLE can achieve this optimal rate only if $m_0(x)$ is smooth enough in a neighborhood of π such as in the constant treatment effects case.

2.2. Construction of the Partially Polynomial Estimator

Because the PLE only explores the information that $m(x)$ (rather than its derivatives) has a jump at π, it cannot achieve the optimal rate of convergence when $m_0(\cdot)$ is known to be only continuous at π. Now, we generalize the PLE to the PPE by explicitly considering the jumps of the derivatives of $m(x)$ at π. Specifically, let

$$y_i=m_q(x_i)+X_i^{d\prime}\theta+\varepsilon_i, \tag{7}$$

where

$$X_i^d=(d_i,d_i(x_i-\pi),\ldots,d_i(x_i-\pi)^q)',\theta=\left(\alpha,\beta_1,\ldots,\beta_q\right)',$$

and $m_q(x_i)\equiv m(x_i)-X_i^{d\prime}\theta$ is an extension of $m_0(x_i)$ in (3) and has continuous derivatives at π to qth order, $\beta_v=\frac{m_+^{(v)}(\pi)-m_-^{(v)}(\pi)}{v!}$, $v=1,\ldots,q$, is the scaled difference of the vth derivatives of $m(x)$ in the left and right neighborhoods of π, and $m_+^{(v)}(\pi)$ and $m_-^{(v)}(\pi)$ are

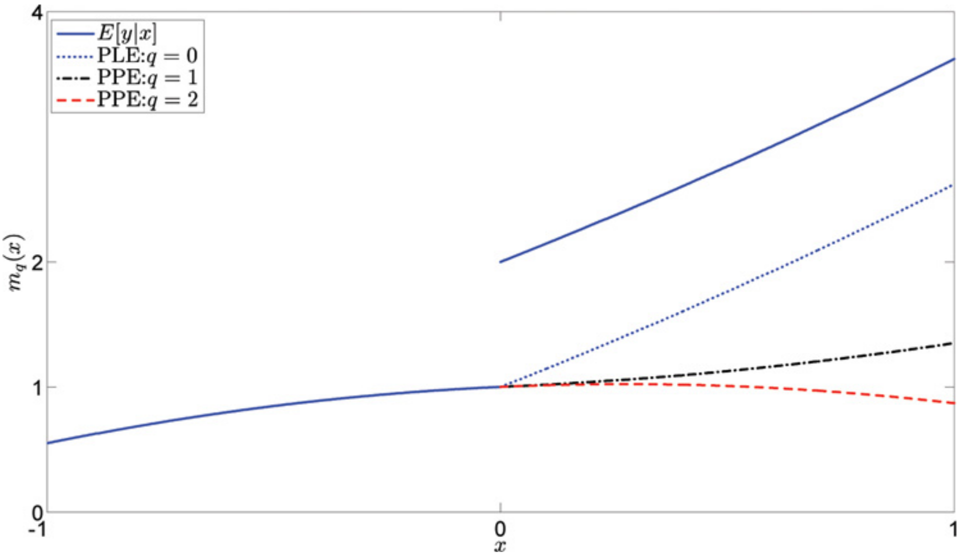

FIGURE 1 $m_q(x)$ in partially polynomial estimation with different orders.

the νth order right and left derivatives of $m(x)$ at π.[2] $m_q(x)$ is shown in Fig. 1, where $m(x) = \begin{cases} 1 + 0.16x - 0.29x^2, & \text{if } x < 0; \\ 2 + 1.43x + 0.19x^2, & \text{if } x \geq 0. \end{cases}$ In this special case, $\alpha = 1$, $\beta_1 = 1.27$, and $\beta_2 = 0.48$. Note that $q = 0$ corresponds to the PLE of Porter (2003). Obviously, its $m_0(x)$ may not be smooth at 0.

The estimator of α, $\hat{\alpha}$, is the first element of the minimizer $\hat{\theta}$ in the following problem:

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \left[\tilde{y}_i(\theta) - \mathbb{P}_{x_i}^{n} \left(\tilde{\mathbf{y}}(\theta) \right) \right]^2, \tag{8}$$

where

$$\tilde{y}_i(\theta) = y_i - X_i^{d\prime}\theta, \tilde{\mathbf{y}}(\theta) = \left(\tilde{y}_1(\theta), \ldots, \tilde{y}_n(\theta) \right)',$$

and $\mathbb{P}_{x_i}^{n}\left(\tilde{\mathbf{y}}(\theta)\right)$ is the pth order LPE of $E\left[\tilde{y}_i(\theta)\,|x_i\right]$ which is equal to $m_q(x_i)$ when θ is evaluated at its true value. To explore the qth order smoothness of $m_q(\cdot)$, we assume $p \geq q$, though p and q are not necessarily the same. From Lemma 2.1 of Fan et al. (1997), $\mathbb{P}_{x_i}^{n}\left(\tilde{\mathbf{y}}(\theta)\right)$ is equivalent to the local constant estimator with a higher-order kernel. Because the kernel function in Porter (2003) is allowed to be higher order, the PPE distinguishes

[2](7) is a partially linear regression in Robinson (1988) because the parametric component of (7) is linear in the parameters. The term PPE is to distinguish (7) from the partially linear regression in Porter (2003).

from the PLE mainly by considering the difference of derivatives at π in (8) rather than using the LPE to estimate $m_q(x_i)$. Since the derivative differences of $m(x)$ at π are taken into account in the PPE, π is more or less like an interior point on x's support, and $\mathbb{P}^n_{x_i}\left(\tilde{\mathbf{y}}(\theta)\right)$ is like estimating the conditional mean at an interior point x_i.

2.3. Connection with the Local Polynomial Estimator

We now build a connection between the PPE $\hat{\alpha}$ and the LPE $\tilde{\alpha}$. For this purpose, we compare the PPE to the LSE in threshold regression; see Chan (1993), Hansen (2000), and Yu (n.d., 2012) for more discussions on threshold regression. A typical setup of the PPE is $p = q$, so we only concentrate on this case. In threshold regression,

$$y = \begin{cases} x'\beta_1 + e_1, z < \pi; \\ x'\beta_2 + e_2, z \geq \pi; \end{cases} \tag{9}$$

where z is the threshold variable used to split the sample, $x \in \mathbb{R}^{p+1}$ is the covariate with the first element being a constant, $\beta \equiv (\beta_1', \beta_2')' \in \mathbb{R}^{2(p+1)}$ and $\sigma \equiv (\sigma_1, \sigma_2)'$ are parameters in mean and variance in the two regimes of (9), the error terms e_1 and e_2 allow for conditional heteroskedasticity and are not necessarily the same, and all the other variables have the same definitions as in the linear regression framework. A useful reparametrization of (9) is

$$y = x'\beta_1 + x'(\beta_2 - \beta_1) 1(z \geq \pi) + e, \tag{10}$$

where $e = e_1$ when $z < \pi$, and $e = e_2$ when $z \geq \pi$. Returning to the regression discontinuity model, (9) is only satisfied *locally*. Note that the approximation in (7) can be written in two equivalent ways:

$$y = \begin{cases} a^- + b_1^-(x-\pi) + \cdots + b_p^-(x-\pi)^p + \varepsilon_0, & x < \pi; \\ a^+ + b_1^+(x-\pi) + \cdots + b_p^+(x-\pi)^p + \varepsilon_1, & x \geq \pi; \end{cases} \tag{11}$$

and

$$\begin{aligned} y = a^- + b_1^-(x-\pi) + \cdots + b_p^-(x-\pi)^p + [\alpha + \beta_1(x-\pi) \\ + \cdots + \beta_p(x-\pi)^p]1(x \geq \pi) + \varepsilon, \end{aligned} \tag{12}$$

where $a^- + b_1^-(x-\pi) + \cdots + b_p^-(x-\pi)^p$ is the Taylor expansion of $m_q(x)$ to order p in the left neighborhood of π with $a^- = m_-(\pi)$ and $b_v^- = m_-^{(v)}(\pi)/v!$, $v = 1, \ldots, p$,

$$\left(a^+, b_1^+, \ldots, b_p^+\right) = \left(a^-, b_1^-, \ldots, b_p^-\right) + \left(\alpha, \beta_1, \ldots, \beta_p\right),$$

and the threshold variable z in (9) is just x. Obviously, $(a^-, b_1^-, \ldots, b_p^-)$, $(\alpha, \beta_1, \ldots, \beta_p)$, $(1\ (x-\pi)\ \cdots\ (x-\pi)^p)'$ and $(\varepsilon_0, \varepsilon_1)$ play the role of $\beta_1, \beta_2 - \beta_1, x$, and (e_1, e_2), respectively, in (10).

The main concern in threshold regression is the threshold point π. In contrast, in RDDs, π is generally known from the design, and the main concern is the mean difference α between the two regimes. In threshold regression, we can set up the objective functions of the least squares estimation for the two equivalent models (9) and (10) as follows:

$$\mathrm{Obj1} = \sum_{i=1}^{n} \left(y_i - x_i'\beta_1 1(z_i < \pi) - x_i'\beta_2 1(z_i \geq \pi)\right)^2,$$

$$\mathrm{Obj2} = \sum_{i=1}^{n} \left(y_i - x_i'\left(\beta_2 - \beta_1\right) 1\left(z_i \geq \pi\right) - x_i'\beta_1\right)^2.$$

Suppose π is known, then in Obj1, $\beta_2 - \beta_1$ is estimated in two steps. First estimate β_2 using the data with $z_i \geq \pi$ and estimate β_1 using the data with $z_i < \pi$, and then take difference of the estimates of β_2 and β_1 in step 1 as the estimator of $\beta_2 - \beta_1$. In contrast, Obj2 uses a profiled procedure: first fix $\beta_2 - \beta_1$ and regress $y_i - x_i'\left(\beta_2 - \beta_1\right) 1\left(z_i \geq \pi\right)$ on x_i to get an estimate of β_1 (as a function of $\beta_2 - \beta_1$), and then minimize Obj2 with respect to $\beta_2 - \beta_1$ to estimate $\beta_2 - \beta_1$.

In RDDs, we only use the data local to π to estimate α, so the weight $k_h(x_i - \pi)$ is imposed on each summand, and also, x_i, β_1, and $\beta_2 - \beta_1$ are substituted by the counterparts of RDDs. The estimates of α based on Obj1 and Obj2 correspond to the LPE and LSE in Section 2.1, respectively. To relate the LPE (or equivalently, the LSE) to the PPE, we express the LSE in (6) as the minimizer $\tilde{\alpha}$ in

$$\min_{\theta} \sum_{i=1}^{n} k_h(x_i - \pi) \left[\tilde{y}_i\left(\theta\right) - \mathbb{P}_{\pi}^{n}\left(\tilde{\mathbf{y}}\left(\theta\right)\right)\right]^2.$$

On the other hand, the PPE is the minimizer $\hat{\alpha}$ in

$$\min_{\theta} \sum_{i=1}^{n} \left[\tilde{y}_i\left(\theta\right) - \mathbb{P}_{x_i}^{n}\left(\tilde{\mathbf{y}}\left(\theta\right)\right)\right]^2.$$

There are two differences between these two objective functions. First, the local weight $k_h(x_i - \pi)$ is removed in the PPE. From Lemma 1 in Appendix C, only the data in the h neighborhood of π will contribute to the objective function of the PPE. Equivalently, the local weight $k_h(x_i - \pi)$ with a uniform kernel is used in the PPE, which may lose some efficiency compared to other kernels such as the Epanechinikov kernel. Second, the LPE of $E[\tilde{y}_i\left(\theta\right)|x_i]$ rather than $E[\tilde{y}_i\left(\theta\right)|x_i = \pi]$ is used in the PPE. Obviously, $\mathbb{P}_{x_i}^{n}\left(\tilde{\mathbf{y}}\left(\theta\right)\right)$ should be a better estimator of $E[\tilde{y}_i\left(\theta\right)|x_i]$ than $\mathbb{P}_{\pi}^{n}\left(\tilde{\mathbf{y}}\left(\theta\right)\right)$. Disregarding the difference in the local weight $k_h(x_i - \pi)$, the PPE will reduce to the LSE by a *relocalization* effect.

Nevertheless, since only $x_i \in N_0$ contributes to the estimation, the performance of these two estimators should be similar; simulations in Section 4 confirm this result. However, the following subsection shows that their asymptotic properties are quite different.

2.4. Asymptotic Theory of $\hat{\alpha}$

First, we specify some regularity conditions required in deriving the asymptotic distribution of $\hat{\alpha}$. These assumptions roughly correspond to those in Section 3.1 of Porter (2003). For example, Assumptions K, F, M(a), M(b), and E correspond to Assumptions 1, the first half of 2(a), the second half of 2(a), 2(b), and 3 in Porter (2003), respectively. See the discussions there for the role of these assumptions in the development of the asymptotic theory.

Assumption K. *$k(\cdot)$ is a symmetric, bounded, Lipschitz function, zero outside a bounded set $[-1, 1]$, and $\int k(u)du = 1$.*

We assume that $k(\cdot)$ has a bounded support $[-1, 1]$ only to simplify the proof. Also, $k(\cdot)$ is a second-order kernel; no higher order kernels are required.

Assumption F. *For some compact interval N of π with $\pi \in int(N)$, f is l_f times continuously differentiable and bounded away from zero.*

This assumption roughly assumes that there is no manipulation of the forcing variable; see McCrary (2008) for more discussions about this assumption and a test on its validity.

Assumption M.

(a) *$m_0(x)$ is l_m times continuously differentiable for $x \in N \setminus \{\pi\}$, and $m_0(x)$ is continuous and has finite right- and left-hand derivatives to order l_m at π.*
(b) *Right- and left-hand derivatives of $m_0(x)$ to order l_m are equal at π.*[3]

The typical case where Assumption M(b) holds is the constant treatment effects model. In such a model, $Y_{1i} - Y_{0i} = \alpha(x) + \varepsilon_1 - \varepsilon_0 = \alpha$ is constant across individuals, so $m_0(x)$ is smooth up to order l_m, and we need not consider the derivative differences.

Assumption E.

(a) *$\sigma^2(x) = E\left[\varepsilon^2 | x\right]$ is continuous for $x \in N \setminus \{\pi\}$, and the right and left-hand limits at π exist.*
(b) *For some $\zeta > 0$, $E\left[\,|\varepsilon|^{2+\zeta}\middle|\, x\right]$ is uniformly bounded on N.*

[3]In this case, estimators of β_ℓ, $\ell = 1, \ldots, q$, converge to zero.

Assumption B below restricts the range of h, which will affect the bias properties of the PPE.

Assumption B. $\frac{n^{\zeta/(2+\zeta)}h}{\ln n} \to \infty$, $\frac{\sqrt{nh}}{\ln n} \to \infty$.

(a) $\sqrt{nh}h^{q+3} \to 0$, $\sqrt{nh}h^{q+1} \to C_a$, *where* $0 \le C_a < \infty$.
(b1) $\sqrt{nh}h^{p+3} \to 0$, $\sqrt{nh}h^{p+1} \to C_{b1}$, *where* $0 \le C_{b1} < \infty$.
(b2) $\sqrt{nh}h^{p+3} \to 0$, $\sqrt{nh}h^{p+2} \to C_{b2}$, *where* $0 \le C_{b2} < \infty$.

The following Theorem 1 provides the asymptotic results for the PPE under different sets of regularity conditions when $q \ge 1$. As in the PLE, since only the data with x_i in an h-neighborhood of π contribute to $\hat{\alpha}$, the convergence rate is $\sqrt{nh}$.

Theorem 1. *Suppose* $p \ge q \ge 1$, *and Assumptions E, F, and K hold with* $l_f \ge 1$.

(a) *If Assumption M(a) holds with* $l_m \ge q+1$, *and Assumption B(a) holds, then*

$$\sqrt{nh}\left(\hat{\alpha} - \alpha\right) \xrightarrow{d} N\left(-C_a B_a, \frac{V}{f(\pi)}\right),$$

Here,

$$B_a = e_1' N_p^{-1}\left[\frac{m_+^{(q+1)}(\pi)}{(q+1)!}Q_{pq}^+ + \frac{m_-^{(q+1)}(\pi)}{(q+1)!}Q_{pq}^-\right],$$
$$V = e_1' N_p^{-1}\left[\sigma_+^2(\pi)\Omega_p^+ + \sigma_-^2(\pi)\Omega_p^-\right]N_p^{-1}e_1$$

with $\sigma_+^2(\pi) = E\left[\varepsilon^2|x = \pi+\right]$, $\sigma_-^2(\pi) = E\left[\varepsilon^2|x = \pi-\right]$,

$$\begin{aligned}
N_p(i,j) &= \int_0^1 K_p^*(\overline{\delta}_{i-1}^+(w))K_p^*(\overline{\delta}_{j-1}^+(w))dw \\
&\quad + \int_{-1}^0 K_p^*(\overline{\delta}_{i-1}^-(w))K_p^*(\overline{\delta}_{j-1}^-(w))dw, \\
Q_{pq}^+(i) &= \int_0^1 K_p^*(\overline{\delta}_{i-1}^+(w))\left(\int_{-w}^1 K_p^*(u)(w+u)^{q+1}du - w^{q+1}\right)dw \\
&\quad + \int_{-1}^0 K_p^*(\overline{\delta}_{i-1}^-(w))\left(\int_{-w}^1 K_p^*(u)(w+u)^{q+1}du\right)dw, \\
Q_{pq}^-(i) &= \int_0^1 K_p^*(\overline{\delta}_{i-1}^+(w))\left(\int_{-1}^{-w} K_p^*(u)(w+u)^{q+1}du\right)dw \\
&\quad + \int_{-1}^0 K_p^*(\overline{\delta}_{i-1}^-(w))\left(\int_{-1}^{-w} K_p^*(u)(w+u)^{q+1}du - w^{q+1}\right)dw,
\end{aligned}$$

$$\Omega_p^+(i,j) = \int_0^1 \Big[K_p^*(\overline{\delta}_{i-1}^+(w)) - \Big(\int_0^1 K_p^*(\overline{\delta}_{i-1}^+(v))K_p^*(w-v)\,dv + \int_{-1}^0 K_p^*(\overline{\delta}_{i-1}^-(v))K_p^*(w-v)\,dv\Big)\Big] \Big[K_p^*(\overline{\delta}_{j-1}^+(w)) - \Big(\int_0^1 K_p^*(\overline{\delta}_{j-1}^+(v))K_p^*(w-v)\,dv + \int_{-1}^0 K_p^*(\overline{\delta}_{j-1}^-(v))K_p^*(w-v)\,dv\Big)\Big]dw,$$

$$\Omega_p^-(i,j) = \int_{-1}^0 \Big[K_p^*(\overline{\delta}_{i-1}^-(w)) - \Big(\int_0^1 K_p^*(\overline{\delta}_{i-1}^+(v))K_p^*(w-v)\,dv + \int_{-1}^0 K_p^*(\overline{\delta}_{i-1}^-(v))K_p^*(w-v)\,dv\Big)\Big] \Big[K_p^*(\overline{\delta}_{j-1}^-(w)) - \Big(\int_0^1 K_p^*(\overline{\delta}_{j-1}^+(v))K_p^*(w-v)\,dv + \int_{-1}^0 K_p^*(\overline{\delta}_{j-1}^-(v))K_p^*(w-v)\,dv\Big)\Big]dw,$$

and

$$K_p^*(\overline{\delta}_{i-1}^+(w)) = w^{i-1} - \int_{-w}^1 K_p^*(u)(w+u)^{i-1}du,$$
$$K_p^*(\overline{\delta}_{i-1}^-(w)) = -\int_{-w}^1 K_p^*(u)(w+u)^{i-1}du,$$

$i, j = 1, \ldots, q+1$, *$K_p^*(u)$ being defined in (2).*

(b1) *If Assumption M(b) holds with $l_m \geq p+1$, and Assumption B(b1) holds, then when p is odd,*

$$\sqrt{nh}\left(\hat{\alpha} - \alpha\right) \xrightarrow{d} N\left(-C_{b1}B_{b1}, \frac{V}{f(\pi)}\right),$$

where

$$B_{b1} = \left(\int_{-1}^1 K_p^*(u)\,u^{p+1}du\right)\frac{m_0^{(p+1)}(\pi)}{(p+1)!}e_1'N_p^{-1}Q_p$$

with

$$Q_p(i) = \int_0^1 K_p^*(\overline{\delta}_{i-1}^+(w))dw + \int_{-1}^0 K_p^*(\overline{\delta}_{i-1}^-(w))dw,$$

$i = 1, \ldots, q+1.$

(b2) *If Assumption M(b) holds with $l_m \geq p+2$, and Assumption B(b2) holds, then when p is even,*

$$\sqrt{nh}\left(\hat{\alpha}-\alpha\right) \xrightarrow{d} N\left(-C_{b2}B_{b2}, \frac{V}{f(\pi)}\right),$$

where

$$B_{b2} = \left(\int_{-1}^{1} K_p^*(u)\, u^{p+2} du\right)\left(\frac{m_0^{(p+1)}(\pi) f'(\pi)}{(p+1)! f(\pi)} + \frac{m_0^{(p+2)}(\pi)}{(p+2)!}\right) e_1' N_p^{-1} Q_p.$$

Theorem 1 is surprising in two aspects. First, under Assumption M, the PPE can achieve the optimal rate by incorporating the derivative differences in the left and right neighborhoods of π. For example, if $m(x)$ is in $\overline{C}_r$, $r > 1$, of Porter (2003), where $\overline{C}_r$ is the set of functions satisfying Assumption M(a) with $l_m = r$, then the PPE with $p \geq q = r-1$ can achieve the optimal convergence rate. If $m(x)$ is in C_r, $r > 2$, of Porter (2003), where C_r is the set of functions satisfying Assumption M(b) with $l_m = r$, then the PPE with $0 < q \leq p = r-1$ ($r-2$ when r is even) can achieve the optimal convergence rate. Second, the PLE is indeed very special. In our notation, when $q = 0$, $Q_{pq}^+ = Q_{pq}^- = 0$, so the bias in (a) is $O(\sqrt{nh}h^2)$ instead of $O(\sqrt{nh}h)$ as illustrated in Theorem 2(a) of Porter (2003). In (b1) and (b2), $Q_p = 0$, so a higher-order bias $O(\sqrt{nh}h^{p+2+1(p \text{ is even})})$ appears as shown in Theorem 2(b) of Porter (2003). This is basically because $1(x_i \geq \pi)$ and $1(x_i < \pi)$ are symmetric, and the lower-order biases in the left and right neighborhoods of π offset each other. In the PPE, $(x_i - \pi)^l 1(x_i \geq \pi)$, $l \geq 1$, and $1(x_i < \pi)$ are not symmetric, so the lower-order bias remains. The order of the biases in the PLE, LPE and PPE is summarized in Table 1. Note that when the kernel is symmetric, the order s of the kernel $k(\cdot)$ in the PLE of Porter (2003) must be even. Roughly speaking, s plays a similar role as $p+1$ when p is odd and $p+2$ when p is even in the PPE. In the LPE, when p is odd and Assumption M(b) holds, the lower-order biases in the two neighborhoods of π offset each other, and a higher-order bias appears.

As discussed above, the PLE with a higher-order kernel is essentially equivalent to the PPE with $q = 0$ and some $p > q$. But there is indeed some subtle difference between them: Theorem 1 needs less stringent conditions on the smoothness of $f(x)$ than Theorem 2 of

TABLE 1
Biases of Four Estimators (the b in $\sqrt{nh}h^b$)

	Assumption M(a)/A(a)	*Assumption M(b)/A(b)*
PLE ($q = 0, p \geq q$)	2	$p+2+1\,(p$ is even$)$
PPE ($p \geq q > 0$)	$q+1$	$p+1+1\,(p$ is even$)$
LPE ($p \geq 0$)	$p+1$	$p+1+1\,(p$ is odd$)$
IVE ($p \geq q \geq 0$)	$q+1$	$p+1$

Porter (2003). For example, in (a), Porter (2003) requires $l_f \geq 2$ while Theorem 1 only requires $l_f \geq 1$; in (b1) and (b2), Porter (2003) requires $l_f \geq s$, while Theorem 1 only requires $l_f \geq 1$. This is the role played by the PPE more than the higher-order kernel estimator; that is, the PPE adapts automatically to the smoothness of the density of x.

Note that the first parts of B_{b1} and B_{b2} are the same as those appearing in Theorem 4.1 of Ruppert and Wand (1994) where the conditional mean at an interior point is estimated, which confirms our intuition that π can be treated as an interior point in the PPE. In case (a), the optimal bandwidth to minimize the mean squared error (MSE) is $O(n^{-\frac{1}{2q+3}})$; in case (b1), the optimal bandwidth is $O(n^{-\frac{1}{2p+3}})$; in case (b2), the optimal bandwidth is $O(n^{-\frac{1}{2p+5}})$. So when we have more smoothness in $m(x)$, the optimal bandwidth gets larger. Note also that N_p, Ω_p^+, Ω_p^-, Q_{pq}^+, Q_{pq}^+ and Q_p only depend on the kernel function, which validates the conventional insight that the bandwidth affects the convergence rate while the kernel only affects the efficiency constant. Also, $K_p^*(\cdot)$ instead of $k(\cdot)$ appears in these notations. This is consistent with the observation in the introduction that the LPE at a interior point is equivalent to the local constant estimator with a higher-order kernel. When $q = p = 0$, $K_p^*(u) = k(u)$, and N_p reduces to $2\int_0^1 K_0^2(w)dw$ in Porter (2003), where $K_0(w) = \int_w^1 k(u)du$. We can check some special cases of (a) to show the results in Theorem 1 are correct. Suppose $m_+^{(q+1)}(\pi) = m_-^{(q+1)}(\pi)$, then

$$\begin{aligned}
&Q_{pq}^+(i) + Q_{pq}^-(i) \\
&= \int_0^1 K_p^*(\overline{\delta}_{i-1}^+(w)) \left(\int_{-1}^1 K_p^*(u)\,(w+u)^{q+1}\,du - w^{q+1} \right) dw \\
&\quad + \int_{-1}^0 K_p^*(\overline{\delta}_{i-1}^-(w)) \left(\int_{-1}^1 K_p^*(u)\,(w+u)^{q+1}\,du - w^{q+1} \right) dw \\
&= \begin{cases} 0, & \text{if } q < p; \\ \left(\int_{-1}^1 K_p^*(u)\,u^{p+1}du \right) Q_p(i), & \text{if } q = p \text{ and } p \text{ odd}; \\ 0, & \text{if } q = p \text{ and } p \text{ even}; \end{cases}
\end{aligned}$$

which matches the asymptotic biases in (b1) and (b2).

Another good property of the PPE is that it automatically generates the (scaled) derivative difference of $m_0(\cdot)$ at the left and right side of π. From the proof of Theorem 1, we can show that

$$\sqrt{nh}H\left(\hat{\theta} - \theta\right) \xrightarrow{d} N\left(-C\mathbf{B}, \frac{\mathbf{V}}{f(\pi)}\right), \tag{13}$$

where C is the constant in each case of Theorem 1, $H = \text{diag}\{1, h, \ldots, h^q\}_{(q+1)\times(q+1)}$, and $\mathbf{B}$ and $\mathbf{V}$ are defined as B and V in each case with e_1 and e_1' deleted. To estimate the derivative difference, we can multiply the left hand side of (13) by $D = \text{diag}\{0!, 1!, \ldots, q!\}$

to get

$$\sqrt{nh}HD\left(\hat{\theta}-\theta\right) \xrightarrow{d} N\left(-CD\mathbf{B}, \frac{D\mathbf{V}D}{f(\pi)}\right).$$

This asymptotic result can be used to test hypotheses like $D\theta = 0$; that is, there is no treatment effect up to qth order derivative.

2.5. Variance Estimation

For inference, we need to estimate the asymptotic bias and variance of $\hat{\alpha}$. For the bias, higher order derivatives of $m(x)$ at $x = \pi$ are involved. It is a standard exercise to estimate these derivatives, see, e.g., Härdle (1990), Pagan and Ullah (1999), and Li and Racine (2007). In practice, it is more popular to use undersmoothing to avoid calculating the bias. As to the variance, we need only to estimate $\sigma_+^2(\pi)$, $\sigma_-^2(\pi)$, and $f(\pi)$ since other components are just complicated functionals of the kernel. The estimation of $f(\pi)$ is straightforward, so we concentrate on the estimation of $\sigma_+^2(\pi)$ and $\sigma_-^2(\pi)$ in the following.

First get the sample analog of ε_i:

$$\hat{\varepsilon}_i = y_i - X_i^{d\prime}\hat{\theta} - \widehat{m}(x_i),$$

where $X_i^d = (d_i, d_i(x_i - \pi), \ldots, d_i(x_i - \pi)^q)$, and $\hat{m}(x_i)$ is determined by the minimizer $\hat{a}$ in

$$\min_{a,b_1,\ldots,b_p} \frac{1}{n}\sum_{j=1}^{n} k_h\left(x_j - x\right)\left[y_j - X_j^{d\prime}\hat{\theta} - a - b_1\left(x_j - x_i\right) - \cdots - b_p\left(x_j - x_i\right)^p\right]^2.$$

Then $\sigma_+^2(\pi)$ is estimated as the minimizer $\hat{a}$ in

$$\min_{a,b_1,\ldots,b_p} \frac{1}{n}\sum_{i=1}^{n} k_h\left(x_i - \pi\right) d_i\left[\hat{\varepsilon}_i^2 - a - b_1\left(x_i - \pi\right) - \cdots - b_p\left(x_i - \pi\right)^p\right]^2,$$

and $\sigma_-^2(\pi)$ is similarly estimated with d_i replacing d_i^c. The estimators are denoted as $\hat{\sigma}_+^2(\pi)$ and $\hat{\sigma}_-^2(\pi)$. The following theorem shows the consistency of $\hat{\sigma}_+^2(\pi)$ and $\hat{\sigma}_-^2(\pi)$.

Theorem 2. *If the assumptions in part (a) of Theorem 1 holds with ζ in Assumption E satisfying $\zeta \geq 2$, and $\frac{\sqrt{nh^2}}{\ln n} \to \infty$, then*

$$\hat{\sigma}_+^2(\pi) \xrightarrow{p} \sigma_+^2(\pi) \quad \text{and} \quad \hat{\sigma}_-^2(\pi) \xrightarrow{p} \sigma_-^2(\pi).$$

3. THE FUZZY DESIGN

In the fuzzy design, we study two estimators, the PPE and the newly proposed IVE. As in the sharp design, we first review the existing estimators in the fuzzy design.

3.1. The Existing Estimators in the Fuzzy Design

As (3), we can write t in the form of $s(x) + \beta d + \eta$, $E[\eta \mid x] = 0$. Note here that it is not necessary to introduce the notation $\beta(x)$ for the following development, and $s(x)$ is similar to $m_0(x)$ in (3). Note also that in the fuzzy design, y generally cannot be written as $y = m_0(x) + \alpha t + \varepsilon$ for some $m_0(x)$. This is because $y = [\underline{m}(x) + t(\alpha(x) - \alpha)] + t\alpha + \varepsilon \equiv m_0(x, t) + t\alpha + \varepsilon$, where $m_0(x, t)$ depends on t unless $\alpha(x) = \alpha$. To express y in this form, we need to redefine the error term:

$$
\begin{aligned}
y &= [\underline{m}(x) + (s(x) + d\beta + \eta)(\alpha(x) - \alpha)] + t\alpha + \varepsilon \\
&= [\underline{m}(x) + (s(x) + d\beta)(\alpha(x) - \alpha)] + t\alpha + [\eta(\alpha(x) - \alpha) + \varepsilon].
\end{aligned}
$$

We can also express y in the form of (3):

$$
\begin{aligned}
y &= t(\underline{m}(x) + \alpha(x) + \varepsilon_1) + (1 - t)(\underline{m}(x) + \varepsilon_0) \\
&= (s(x) + d\beta + \eta)(\underline{m}(x) + \alpha(x) + \varepsilon_1) + (1 - s(x) - d\beta - \eta)(\underline{m}(x) + \varepsilon_0) \\
&= (s(x) + d\beta)\big(\underline{m}(x) + \alpha(x)\big) + (1 - s(x) - d\beta)\,\underline{m}(x) \\
&\quad + (s(x) + d\beta)\varepsilon_1 + (1 - s(x) - d\beta)\varepsilon_0 + \alpha(x)\eta + \eta(\varepsilon_1 - \varepsilon_0) \\
&\equiv [\underline{m}(x) + s(x)\alpha(x) + d\beta(\alpha(x) - \alpha)] + d\beta\alpha + R \\
&\equiv m_0(x) + d\beta\alpha + R \equiv m(x) + R, \qquad (14)
\end{aligned}
$$

where

$$
R = (s(x) + d\beta)\varepsilon_1 + (1 - s(x) - d\beta)\varepsilon_0 + \alpha(x)\eta + \eta(\varepsilon_1 - \varepsilon_0). \qquad (15)
$$

So the jump size of $E[y \mid x]$ at π is $\Delta \equiv \beta\alpha$, and the error term changes to R.

The LPE (or LSE) can be easily extended to the fuzzy design. The resulting estimator

$$
\tilde{\alpha}_f = \frac{\widetilde{\Delta}}{\tilde{\beta}},
$$

where $\widetilde{\Delta}$ and $\tilde{\beta}$ are the LPEs (or LSEs) based on $\{y_i, x_i\}_{i=1}^n$ and $\{t_i, x_i\}_{i=1}^n$, respectively. Hahn et al. (2001) show that this estimator is numerically equivalent to the Wald estimator when the uniform kernel is used and $p = 0$. Imbens and Lemieux (2008) and Lee and Lemieux (2010) mention that this estimator is numerically equivalent to the 2SLS estimator based on the following model of regression with endogeneity when the uniform kernel is used and $p = 1$:

$$y_i = \phi_0 + \phi_1 d_i (x_i - \pi) + \phi_2 d_i^c (x_i - \pi) + t_i \alpha + r_i,$$
$$t_i = \psi_0 + \psi_1 d_i (x_i - \pi) + \psi_2 d_i^c (x_i - \pi) + d_i \beta + \eta_i,$$

where only the data such that $x_i \in N_0$ are used in the estimation, the endogenous variable is t_i, and the excluded exogenous variable is d_i. As in the LSE of the sharp design, the standard error of the 2SLS estimator is valid. Also, this estimator can be easily extended to the case with $p > 1$ and a general kernel as in (6). When Δ and β are estimated by the PLE (instead of the LPE), we get the PLE of α in the fuzzy design.

3.2. PPE

For the PPE in the fuzzy design, we estimate α by

$$\hat{\alpha}_f = \frac{\widehat{\Delta}}{\hat{\beta}},$$

where $\widehat{\Delta}$ and $\hat{\beta}$ are the PPEs based on $\{y_i, x_i\}_{i=1}^n$ and $\{t_i, x_i\}_{i=1}^n$, respectively. The following theorem states the asymptotic distribution of $\hat{\alpha}_f$. First, we give some extra assumptions. Assumption S is the counterpart of Assumption M for $s(x)$. Also, the original Assumption M is replaced by Assumption A.

Assumption S.

(a) *$s(x)$ is l_s times continuously differentiable for $x \in N \backslash \{\pi\}$, and $s(x)$ is continuous and has finite right and left-hand derivatives to order l_s at π.*
(b) *Right- and left-hand derivatives of $s(x)$ to order l_s are equal at π.*

Assumption A. *$\underline{m}(x)$ and $\alpha(x)$ are l_m and l_α times continuously differentiable for $x \in N$, respectively.*

(a) $\alpha^{(q+1)}(\pi) \neq 0.$
(b) $\alpha^{(\ell)}(\pi) = 0,\ \ell = 1, \ldots, l_\alpha.$

Assumption E in Section 3.3 is replaced by Assumption E′ below.

Assumption E′.

(a) *$\sigma_1^2(x) \equiv E\left[\varepsilon_1^2 | x\right]$ and $\sigma_0^2(x) \equiv E\left[\varepsilon_0^2 | x\right]$ are continuous for $x \in N$.*
(b) *For some $\zeta > 0$, $E\left[\left|\varepsilon_1\right|^{2+\zeta} \middle| x\right]$ and $E\left[\left|\varepsilon_0\right|^{2+\zeta} \middle| x\right]$ are uniformly bounded on N.*

We also need the local unconfoundedness (LU) condition of Hahn et al. (2001) (see their Theorem 2).

Assumption LU. $E[\eta(\varepsilon_1 - \varepsilon_0)|x] = 0$ *for* $x \in N$.

Without this assumption, $E[R|x] \neq 0$ for $x \in N$, where R is defined in (15).

Theorem 3. *Suppose* $p \geq q$, $q \geq 1$, *and Assumptions* E', *F, K, and LU hold with* $l_f \geq 1$.

(a) *If Assumption A(a) and S(a) hold with* $l_m \geq q+1$, $l_\alpha \geq q+1$, $l_s \geq q+1$, *and Assumption B(a) holds, then*

$$\sqrt{nh}\left(\hat{\alpha}_f - \alpha\right) \xrightarrow{d} \frac{1}{\beta} N\left(-C_a\left(B_a^\Delta - \alpha B_a^\beta\right), \frac{V_\Delta - 2\alpha C_{\Delta\beta} + \alpha^2 V_\beta}{f(\pi)}\right),$$

where

$$B_a^\Delta = e_1' N_p^{-1}\left[\frac{m_+^{(q+1)}(\pi)}{(q+1)!} Q_{pq}^+ + \frac{m_-^{(q+1)}(\pi)}{(q+1)!} Q_{pq}^-\right],$$
$$B_a^\beta = e_1' N_p^{-1}\left[\frac{s_+^{(q+1)}(\pi)}{(q+1)!} Q_{pq}^+ + \frac{s_-^{(q+1)}(\pi)}{(q+1)!} Q_{pq}^-\right],$$
$$V_\Delta = e_1' N_p^{-1}\left[E\left[R^2|x=\pi+\right]\Omega_p^+ + E\left[R^2|x=\pi-\right]\Omega_p^-\right] N_p^{-1} e_1,$$
$$C_{\Delta\beta} = e_1' N_p^{-1}\left[E[R\eta\,|\,x=\pi+]\Omega_p^+ + E[R\eta\,|\,x=\pi-]\Omega_p^-\right] N_p^{-1} e_1,$$
$$V_\beta = e_1' N_p^{-1}\left[(s(\pi)+\beta)(1-s(\pi)-\beta)\Omega_p^+ + s(\pi)(1-s(\pi))\Omega_p^-\right] N_p^{-1} e_1,$$

with $s_+^{(q+1)}(\pi)$ *and* $s_-^{(q+1)}(\pi)$ *being the* $(q+1)$*th order right and left derivatives of* $s(x)$ *at* π *and* $m(x)$ *being defined in (14).*

(b1) *If Assumption A(b) and S(b) hold with* $l_m \geq p+1$, $l_\alpha \geq p+1$, $l_s \geq p+1$, *and Assumption B(b1) holds, then when p is odd,*

$$\sqrt{nh}\left(\hat{\alpha}_f - \alpha\right) \xrightarrow{d} \frac{1}{\beta} N\left(-C_{b1}\left(B_{b1}^\Delta - \alpha B_{b1}^\beta\right), \frac{V_\Delta - 2\alpha C_{\Delta\beta} + \alpha^2 V_\beta}{f(\pi)}\right),$$

where

$$B_{b1}^\Delta = \left(\int_{-1}^1 K_p^*(u)\, u^{p+1} du\right) \frac{m_0^{(p+1)}(\pi)}{(p+1)!} e_1' N_p^{-1} Q_p,$$
$$B_{b1}^\beta = \left(\int_{-1}^1 K_p^*(u)\, u^{p+1} du\right) \frac{s^{(p+1)}(\pi)}{(p+1)!} e_1' N_p^{-1} Q_p,$$

with $m_0(x)$ *being defined in (14).*

(b2) *If Assumption A(b) and S(b) hold with $l_m \geq p+2$, $l_\alpha \geq p+2$, $l_s \geq p+2$, and Assumption B(b2) holds, then when p is even,*

$$\sqrt{nh}\left(\hat{\alpha}_f - \alpha\right) \xrightarrow{d} \frac{1}{\beta} N\left(-C_{b2}\left(B_{b2}^{\Delta} - \alpha B_{b2}^{\beta}\right), \frac{V_\Delta - 2\alpha C_{\Delta\beta} + \alpha^2 V_\beta}{f(\pi)}\right),$$

where

$$B_{b2}^{\Delta} = \left(\int_{-1}^{1} K_p^*(u)\, u^{p+2} du\right)\left(\frac{m_0^{(p+1)}(\pi) f'(\pi)}{(p+1)! f(\pi)} + \frac{m_0^{(p+2)}(\pi)}{(p+2)!}\right) e_1' N_p^{-1} Q_p,$$

$$B_{b2}^{\beta} = \left(\int_{-1}^{1} K_p^*(u)\, u^{p+2} du\right)\left(\frac{s^{(p+1)}(\pi) f'(\pi)}{(p+1)! f(\pi)} + \frac{s^{(p+2)}(\pi)}{(p+2)!}\right) e_1' N_p^{-1} Q_p,$$

with $m_0(x)$ being defined in (14).

In (a), by the form of $m(x)$, we can see $m_+^{(q+1)}(\pi) = \underline{m}^{(q+1)}(\pi) + (s(\pi) + \beta)\alpha^{(q+1)}(\pi) + \alpha s_+^{(q+1)}(\pi)$, and $m_-^{(q+1)}(\pi) = \underline{m}^{(q+1)}(\pi) + s(\pi)\alpha^{(q+1)}(\pi) + \alpha s_-^{(q+1)}(\pi)$. In (b1), $m_0^{(p+1)}(\pi) = \underline{m}^{(p+1)}(\pi) + \alpha s^{(p+1)}(\pi)$ and in (b2), $m_0^{(p+2)}(\pi) = \underline{m}^{(p+2)}(\pi) + \alpha s^{(p+2)}(\pi)$. The comments on Theorem 1 can still be applied here. Theorem 3 assumes that $\underline{m}(\cdot)$, $\alpha(\cdot)$ and $s(\cdot)$ are similarly smooth. When this assumption does not hold, we should adjust the biases and variances in this theorem, but we will not pursue this point in this paper.

3.3. IVE

To define the IVE, we first put regression discontinuity designs in the usual regression framework with endogeneity:

$$y = \underline{m}(x) + t\alpha(x) + \varepsilon_t,$$
$$t = s(x) + d\beta + \eta,$$

where $\varepsilon_t = \varepsilon_0 + t(\varepsilon_1 - \varepsilon_0)$, t is endogenous, and d is an instrumental variable (IV). Local to π, we approximate $\underline{m}(x)$, $\alpha(x)$, and $s(x)$ by constants, and then the IV estimator is just the Wald estimator. If we approximate $\underline{m}(x)$, $\alpha(x)$, and $s(x)$ by polynomials around π, we will get

$$\begin{aligned} y &\approx \phi_0 + \phi_1(x-\pi) + \cdots + \phi_p(x-\pi)^p \\ &\quad + t(\alpha + \gamma_1(x-\pi) + \cdots + \gamma_q(x-\pi)^q) + \varepsilon_t, \\ t &\approx \psi_0 + \psi_1(x-\pi) + \cdots + \psi_p(x-\pi)^p \\ &\quad + d(\beta + \xi_1(x-\pi) + \cdots + \xi_q(x-\pi)^q) + \eta, \end{aligned} \tag{16}$$

where we localize $\underline{m}(\cdot)$ and $s(\cdot)$ around π (instead of x_i as in the PPE).[4] The endogenous variables are $(t, t(x-\pi), \ldots, t(x-\pi)^q)$, and the excluded IVs are $(d, d(x-\pi), \ldots, d(x-\pi)^q)$. Indeed, the instruments may be correlated with ε_t as argued in Hahn et al. (2001), but since the arguments are local to π, the corresponding orthogonality condition should be

$$E\left[(d, d\,(x-\pi), \ldots, d\,(x-\pi)^q)'\varepsilon_t | x \in N\right] = 0,$$

which reduces to $E\left[(d, d\,(x-\pi), \ldots, d\,(x-\pi)^q)'\,(\varepsilon_0 + t(\varepsilon_1 - \varepsilon_0))\,|x\right] = 0$ as long as $f(x) > 0$, $x \in N$. Notice that

$$\begin{aligned} &E\left[(d, d\,(x-\pi), \ldots, d\,(x-\pi)^q)'\,(\varepsilon_0 + t(\varepsilon_1 - \varepsilon_0))\,|x\right] \\ &\quad = E[(d, d\,(x-\pi), \ldots, d\,(x-\pi)^q)' \\ &\qquad (\varepsilon_0 + (s(x) + d\beta + \eta)\,(\varepsilon_1 - \varepsilon_0))\,|x] = 0, \end{aligned}$$

where the last equality is from the LU assumption. But it is easy to see that

$$E\left[(t, t\,(x-\pi), \ldots, t\,(x-\pi)^q)'\,(\varepsilon_0 + t(\varepsilon_1 - \varepsilon_0))\,|x\right]$$

is generally not zero, so there is endogeneity even in the neighborhood N. In summary, the validity of the orthogonality condition relies on the smoothness of $\underline{m}(x)$ and $\alpha(x)$ such that they can be approximated as polynomials for $x \in N$ and also the LU condition, which are exactly the conditions required for identification in Theorem 2 of Hahn et al. (2001). Given the IVs, the IVE of $\vartheta \equiv (\alpha, \gamma_1, \ldots, \gamma_q, \phi_0, \ldots, \phi_p)'$ is

$$\hat{\vartheta} = \left[\sum_{x_i \in N_0} \begin{pmatrix} X_i^d \\ X_i \end{pmatrix} \left(X_i^{t\prime}\ X_i'\right)\right]^{-1} \sum_{x_i \in N_0} \begin{pmatrix} X_i^d \\ X_i \end{pmatrix} y_i, \tag{17}$$

where $N_0 = [\pi - h, \pi + h]$, $X_i = (1, (x_i - \pi), \ldots, (x_i - \pi)^p)'$, $X_i^t = (t_i, t_i(x_i - \pi), \ldots, t_i(x_i - \pi)^q)'$, and $X_i^d = (d_i, d_i(x_i - \pi), \ldots, d_i(x_i - \pi)^q)'$. The following theorem states the asymptotic distribution of $\hat{\alpha}_I$, which is the first element of $\hat{\vartheta}$.

Theorem 4. *Suppose $p \geq q \geq 0$ and Assumptions E′, F, LU, S hold with $l_f \geq 0$ and $l_s \geq 0$.*

(i) *Under Assumption A(a) with $l_m \geq p+1$, and $l_\alpha \geq q+1$, $\sqrt{nh}h^{q+1} \to C_a$ with $0 \leq C_a < \infty$,*

$$\sqrt{nh}\left(\hat{\alpha}_I - \alpha\right) \xrightarrow{d} N\left(C_a B_a^I, \frac{V_I}{f(\pi)}\right),$$

[4]Note here that we localize $s(x)$ around π as $\psi_0 + \psi_1(x-\pi) + \cdots + \psi_p(x-\pi)^p + d(\xi_1(x-\pi) + \cdots + \xi_q(x-\pi)^q)$ just as we localize $m_0(x)$ in (7).

where

$$B_a^I = e_1' \begin{pmatrix} (s(\pi)+\beta)\,\Gamma_+^{qq} & \Gamma_+^{qp} \\ s(\pi)\Gamma^{pq}+\beta\Gamma_+^{pq} & \Gamma^{pp} \end{pmatrix}^{-1} \left[\begin{array}{c} 1(p=q)\frac{m^{(p+1)}(\pi)}{(p+1)!}\begin{pmatrix} \mu_{p+1,p+q+1}^{+} \\ \mu_{p+1,2p+1} \end{pmatrix} \\ +\frac{\alpha^{(q+1)}(\pi)}{(q+1)!}\begin{pmatrix} (s(\pi)+\beta)\,\mu_{q+1,2q+1}^{+} \\ s(\pi)\mu_{q+1,p+q+1}+\beta\mu_{q+1,p+q+1}^{+} \end{pmatrix} \end{array} \right],$$

$$V_I = e_1' \begin{pmatrix} (s(\pi)+\beta)\,\Gamma_+^{qq} & \Gamma_+^{qp} \\ s(\pi)\Gamma^{pq}+\beta\Gamma_+^{pq} & \Gamma^{pp} \end{pmatrix}^{-1} \begin{pmatrix} E\left[\varepsilon_t^2 \middle| x=\pi+\right]\Gamma_+^{qq} & E\left[\varepsilon_t^2 \middle| x=\pi+\right]\Gamma_+^{qp} \\ E\left[\varepsilon_t^2 \middle| x=\pi+\right]\Gamma_+^{pq} & E\left[\varepsilon_t^2 \middle| x=\pi+\right]\Gamma_+^{pp} + E\left[\varepsilon_t^2 \middle| x=\pi-\right]\Gamma_-^{pp} \end{pmatrix} \cdot \begin{pmatrix} (s(\pi)+\beta)\,\Gamma_+^{qq} & s(\pi)\Gamma^{qp}+\beta\Gamma_+^{qp} \\ \Gamma_+^{pq} & \Gamma^{pp} \end{pmatrix}^{-1} e_1,$$

with

$$\Gamma_+^{qq} = (\gamma_{i+j-2}^{+})_{1\le i,j\le q+1}, \qquad \Gamma^{pp} = (\gamma_{i+j-2})_{1\le i,j\le p+1},$$
$$\mu_{q+1,2q+1}^{+} = \left(\gamma_{q+1}^{+},\ldots,\gamma_{2q+1}^{+}\right)', \qquad \mu_{p+1,2p+1} = \left(\gamma_{p+1},\ldots,\gamma_{2p+1}\right)',$$
$$\gamma_j^{+} = \int_0^1 u^j du, \gamma_j = \int_{-1}^1 u^j du,$$

and other Γ and μ terms being similarly defined.

(ii) *Under Assumption A(b) with $l_m \ge p+1$, and $l_\alpha \ge p+1$, $\sqrt{nh}h^{p+1} \to C_b$ with $0 \le C_b < \infty$,*

$$\sqrt{nh}\left(\hat{\alpha}_I - \alpha\right) \xrightarrow{d} N\left(C_b B_b^I, \frac{V_I}{f(\pi)}\right),$$

where

$$B_b^I = \frac{m^{(p+1)}(\pi)}{(p+1)!} e_1' \begin{pmatrix} (s(\pi)+\beta)\,\Gamma_+^{qq} & \Gamma_+^{qp} \\ s(\pi)\Gamma^{pq}+\beta\Gamma_+^{pq} & \Gamma^{pp} \end{pmatrix}^{-1} \begin{pmatrix} \mu_{p+1,p+q+1}^{+} \\ \mu_{p+1,2p+1} \end{pmatrix}$$

From Theorem 4, the bias is $O(\sqrt{nh}h^{q+1})$ under Assumption A(a) and $O(\sqrt{nh}h^{p+1})$ under Assumption A(b). This implies that the Wald estimator has a bias of order $O_p(\sqrt{nh}h)$, same as the NW estimator in Section 3.2 of Porter (2003). This bias information is added to Table 1; the bias properties of the IVE are comparable with the LPE and PPE. Note also that $s(x)$ is only required to be continuous for the IVE while the LPE requires $l_s \geq p+1$.

In what follows, we point out some connection between the IV estimator and the 2SLS estimator in Section 4.3 of Imbens and Lemieux (2008) and Section 4.3.2 of Lee and Lemieux (2010). As mentioned in Section 3.1, those authors claim that for $x \in N_0$, the model can be approximated as

$$\begin{aligned} y &\approx \phi_0 + \phi_1 d\,(x-\pi) + \phi_2 d^c\,(x-\pi) + t\alpha + r, \\ t &\approx \psi_0 + \psi_1 d\,(x-\pi) + \psi_2 d^c\,(x-\pi) + d\beta + \eta, \end{aligned} \tag{18}$$

while the approximation in (16) is

$$\begin{aligned} y &\approx \phi_0 + \phi_1\,(x-\pi) + t\alpha + \gamma_1 t\,(x-\pi) + \varepsilon_t, \\ t &\approx \psi_0 + \psi_1\,(x-\pi) + d\beta + \xi_1 d\,(x-\pi) + \eta. \end{aligned} \tag{19}$$

Here, we take the local linear form to emphasize the essence of the problem. In (18), the endogenous variable is only t, while in (19), the endogenous variables include both t and $t(x-\pi)$. If we substitute t in the second equation of (19) to $t\,(x-\pi)$ in the first equation and neglect higher order terms of $(x-\pi)$, then we have

$$y \approx \phi_0 + (\phi_1 + \gamma_1\psi_0)\,(x-\pi) + \gamma_1\beta d\,(x-\pi) + t\alpha + \varepsilon_t,$$

which is exactly the approximation in (18). In this sense, the 2SLS estimator first substitutes the t in higher order endogeneity ($t(x-\pi)$ is higher order endogeneity relative to t for $x \in N_0$) by its reduced form, and then apply the IV estimation. In contrast, our IVE apply the IV estimation directly to all endogeneities. Given that the system is just identified, it is easy to show that the 2SLS estimator based on (18) is numerically equivalent to the LPE, so the discussion above provides a connection between the IVE and the LPE. Since the two estimators are constructed differently, their asymptotic distributions are quite different. Putting all discussions together, we get the relationships in Fig. 2. All estimators in the figure can be applied in the fuzzy design, while the IVE and the 2SLS estimator cannot be applied in the sharp design due to obvious reasons.

Given the formulas of the biases and variances, we can estimate them by their sample analogs as usual, so we will not pursue them here. As an alternative of the usual variance estimation, we propose to use the standard error of the IVE as the standard error of $\hat{\alpha}_I$ since it can be easily read from popular econometric software packages such as Stata.

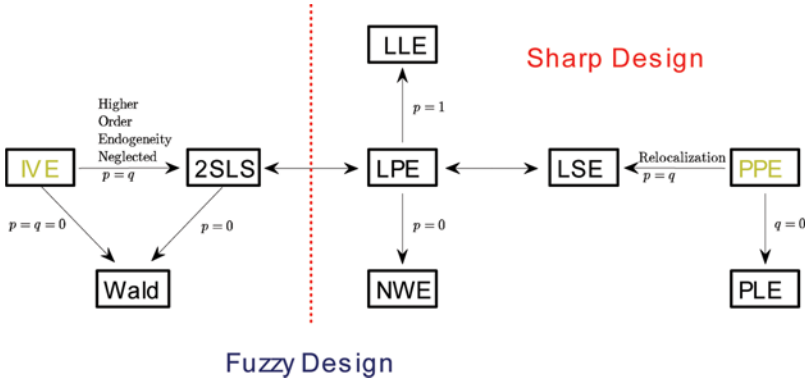

FIGURE 2 Relationship between known estimators.

Recall that for the IVE, the standard error can be obtained from the matrix

$$\widehat{\Sigma} = \left[\sum_{x_i \in N_0} \begin{pmatrix} X_i^d \\ X_i \end{pmatrix} \left(X_i^{t\prime}\ X_i^{\prime}\right)\right]^{-1} \left[\sum_{x_i \in N_0} \begin{pmatrix} X_i^d \\ X_i \end{pmatrix} \left(X_i^{d\prime}\ X_i^{\prime}\right) \widehat{\varepsilon}_{ti}^2\right] \left[\sum_{x_i \in N_0} \begin{pmatrix} X_i^t \\ X_i \end{pmatrix} \left(X_i^{d\prime}\ X_i^{\prime}\right)\right]^{-1},$$

where $\hat{\varepsilon}_{ti} = y_i - \left(X_i^{t\prime}\ X_i^{\prime}\right) \hat{\vartheta}$. We show below the consistency of this estimator.

Theorem 5. *If the assumptions in part (a) of Theorem 4 hold with ζ in Assumption E′ satisfying $\zeta \geq 2$, then*

$$nhe_1^{\prime} \widehat{\Sigma} e_1 \xrightarrow{p} \frac{V_I}{f(\pi)}.$$

Theorem 5 implies that the usual standard error in the IV estimation is valid in RDDs. An immediate corollary is that the usual standard error for the 2SLS estimator is also consistent for the LPE. Another corollary is that the usual standard error in regressing y_i on $(X_i^{d\prime}\ X_i^{\prime})$ for $x_i \in N_0$ in the sharp design is consistent; that is, the standard error of the usual least squares regression is valid for the LSE in the sharp design, where we need only change X_i^t in $\widehat{\Sigma}$ to X_i^d and $\hat{\varepsilon}_{ti}$ to $\tilde{\varepsilon}_i = y_i - (X_i^{d\prime}\ X_i^{\prime})(\tilde{\theta}^{\prime}, \tilde{\beta}_-^{\prime})^{\prime}$ with $(\tilde{\theta}^{\prime}, \tilde{\beta}_-^{\prime})^{\prime}$ being the LSE of $(\alpha, \beta_1, \ldots, \beta_p, a^-, b_1^-, \ldots, b_p^-)^{\prime}$ in (12). Also, we see from Theorems 4 and 5 that the IVE uses the uniform kernel. Of course, we can use other kernels in practice, and then $\hat{\vartheta}$ will change to

$$\hat{\vartheta} = \left[\sum_{i=1}^{n} \begin{pmatrix} X_i^d \\ X_i \end{pmatrix} \left(X_i^{t\prime}\ X_i^{\prime}\right) k\left(\frac{x_i - \pi}{h}\right)\right]^{-1} \sum_{i=1}^{n} \begin{pmatrix} X_i^d \\ X_i \end{pmatrix} y_i k\left(\frac{x_i - \pi}{h}\right).$$

For $\widehat{\Sigma}$, we just multiply $\left(X_i^{d\prime}\ X_i^{t\prime}\ X_i^{\prime}\ y_i\right)$ by $\sqrt{k\left(\frac{x_i-\pi}{h}\right)}$ and plug in the formula of $\widehat{\Sigma}$. Obviously, $\hat{\vartheta}$ is also equivalent to (17) under such a substitution.

4. SIMULATIONS

In this section, we conduct some simulations to check the finite-sample performance of the estimators discussed in this paper. We will use similar specifications as in Fig. 1. Specifically, the following two DGPs for y are used:

$$\text{DGPy1} : y = 1 + 0.16x - 0.29x^2 + t + \varepsilon,$$
$$\text{DGPy2} : y = 1 + 0.16x - 0.29x^2 + t\left(1 + 1.27x + 0.48x^2\right) + \varepsilon,$$

where ε follows $N(0, 0.2^2)$, and x follows the uniform distribution on $[-1, 1]$. DGPy1 corresponds to constant treatment effects: (Assumption M(b) or A(b)) and DGPy2 corresponds to variable treatment effects (Assumption M(a) or A(a)). t in the fuzzy design also follows two DGPs:

$$\text{DGPt1} : t = 0.25 + 0.2x + 0.05x^2 + 0.5 \cdot 1\,(x \geq 0) + \eta,$$
$$\text{DGPt2} : t = 0.25 + 0.2x + 0.05x^2 + \left(0.5 + 0.15x - 0.2x^2\right) \cdot 1\,(x \geq 0) + \eta,$$

where $s(x)$ in DGPt1 (DGPt2) satisfies Assumption S(b) (S(a)), and η and ε are independent. For each DGP, we consider the PLE, PPE, LPE, and IV estimator for $p = q = 0, 1$ or 2. The kernel function is set as the Epanechinikov kernel $k(u) = \frac{3}{4}\left(1 - u^2\right) 1\,(|u| \leq 1)$. In the PLE, the corresponding equivalent kernels are used. Both the sample size n and the number of replications are set as 500. The bandwidth is set as fixed from $[0.1, 1]$. We do not discuss the bandwidth selection in this paper; see Porter and Yu (2010) for a summary of the existing bandwidth selection methods in RDDs. Figures 3 and 4 summarize the bias and root mean squared error (RMSE) of the estimators in the sharp design, and Figs. 5 and 6 in the fuzzy design.

From Figs. 3 and 4, a few results of interest are summarized as follows. First, from $p = 0$ in both figures, the NWE (the LPE in the figures) indeed has larger biases relative to the PLE (which is equivalent to the PPE with $p = 0$). Second, the biases and RMSEs of the PLE is not quite stable compared to the PPE and LPE when $p = 1$ and 2 especially in the variable treatment effects case.[5] Third, the performances of the PPE and the LPE are quite similar. When $p = 1$ or 2, their biases almost disappear. Compared with $p = 1$, $p = 2$ seems to have smaller bias but larger variance just as expected. We then switch

[5]Note that the lines labeled as PLE when $p = 1$ is the same as when $p = 0$. They are drawn on for comparison. Note also that in the constant treatment effects case with $p = 2$, the PLE seems to have better bias and RMSE properties than the PPE and LPE for some range of bandwidth; this is understandable from the theoretical analysis: the bias is $O(h^5)$ for the PLE and is $O(h^4)$ for the PPE and $O(h^3)$ for the LPE.

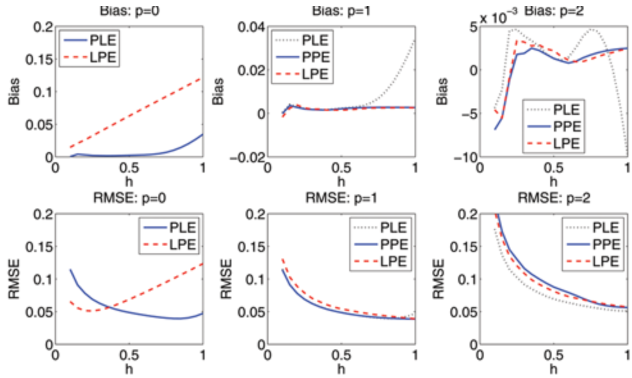

FIGURE 3 Bias and RMSE of PLE, PPE, and LPE in DGPy1 of sharp design.

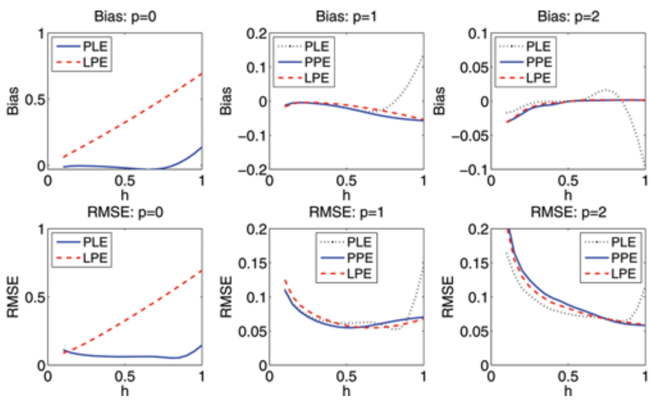

FIGURE 4 Bias and RMSE of PLE, PPE, and LPE in DGPy2 of sharp design.

to the fuzzy design. In Figs. 5 and 6, we delete the results for the PLE and PPE since they are much worse than the LPE and IVE.[6] From these two figures, the LPE and IVE perform similarly especially when $p = 1$. $p = 0$ will have large biases while $p = 2$ will have large variances. These four figures also indicate two well-known results: (i) models with variable treatment effects are harder to estimate than models with constant treatment effects; (ii) fuzzy designs are harder to estimate than sharp designs.

From this simulation study, our suggestion is as follows: in the sharp design, use the LPE with $p = 1$, and in the fuzzy design, use the LPE or IVE with $p = 1$. Our suggestion

[6]Given that the performance of the PPE is similar to the LPE in the sharp design, we can conclude that the PPE does not work well in estimating β; recall β is the jump size of the propensity score $E[t \mid x] = s(x) + \beta d$.

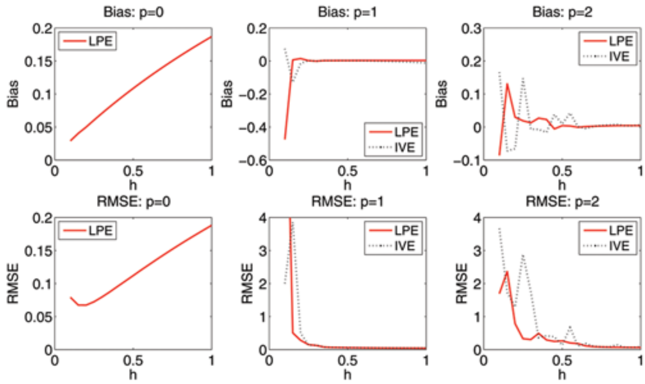

FIGURE 5 Bias and RMSE of LPE and IVE in DGPy1 and DGPt1 of fuzzy design.

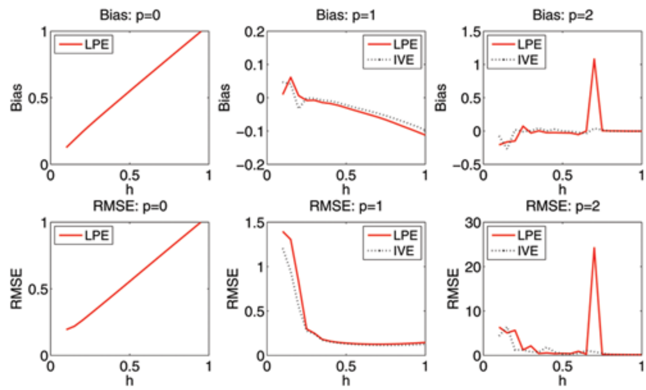

FIGURE 6 Bias and RMSE of LPE and IVE in DGPy2 and DGPt2 of fuzzy design.

is based on two facts: (i) These estimators have good balance between bias and variance; (ii) It is easy to report standard errors for these estimators.

5. CONCLUSION

This paper tries to deepen understanding of the existing estimators in regression discontinuity designs such as the local polynomial estimator and the partially linear estimator. For this purpose, we propose two new estimators of treatment effects. The first estimator is the partially polynomial estimator which extends the partially linear estimator in Porter (2003). Unlike the partially linear estimator, this estimator can achieve the optimal rate of convergence even under broader conditions of the data generating process. This estimator is also related to the popular local polynomial estimator by a relocalization effect. The second estimator is a new instrumental variable estimator in the fuzzy design. This estimator will reduce to the local polynomial estimator if higher order endogeneities are neglected. We study the asymptotic properties of these two estimators and use simulation studies to confirm the theoretical analysis.

APPENDIX A: PROOF OF THEOREM 1

First, we review the LPE of $m(x) \equiv E\left[y_i|x_i = x\right]$ at the end of the introduction and introduce some notations:

$$\begin{aligned}
\hat{m}(x) &= \mathbb{P}_x^n(\mathbf{y}) \equiv e_1'\left(X(x)'K(x)X(x)\right)^{-1}X(x)'K(x)\mathbf{y}, \\
&= e_1'\left(H^{-1}X(x)'K_h(x)X(x)H^{-1}\right)^{-1}H^{-1}X(x)'K_h(x)\mathbf{y}, \\
&\equiv e_1'\left(Z(x)'K_h(x)Z(x)\right)^{-1}Z(x)'K_h(x)\mathbf{y}, \\
&= e_1'\left(\frac{1}{n}\sum_{j=1}^{n}Z_j(x)Z_j'(x)k_h\left(x_j - x\right)\right)^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}Z_j(x)k_h\left(x_j - x\right)y_j\right), \\
&\equiv e_1'S_n^{-1}(x)\tilde{r}\left(y(x)\right), \qquad (20)
\end{aligned}$$

where

$$\begin{aligned}
X(x) &= \begin{pmatrix} 1 & x_1 - x & \cdots & (x_1 - x)^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n - x & \cdots & (x_n - x)^p \end{pmatrix}_{n\times(p+1)} \\
&\equiv \begin{pmatrix} X_1(x)' \\ \vdots \\ X_n(x)' \end{pmatrix} \equiv \left(X^0(x), \ldots, X^p(x)\right), \\
K(x) &= \operatorname{diag}\left\{k\left(\frac{x_1 - x}{h}\right), \ldots, k\left(\frac{x_n - x}{h}\right)\right\}_{n\times n}, \\
K_h(x) &= \operatorname{diag}\left\{k_h(x_1 - x), \ldots, k_h(x_n - x)\right\}_{n\times n}, \\
e_1 &= (1, 0, \ldots, 0)'_{(p+1)\times 1}, \qquad H = \operatorname{diag}\left\{1, h, \ldots, h^p\right\}_{(p+1)\times(p+1)}, \\
Z(x) &= X(x)H^{-1}, \\
Z_j(x) &= \left(1, \frac{x_j - x}{h}, \ldots, \left(\frac{x_j - x}{h}\right)^p\right)'_{(p+1)\times 1} \equiv \left(Z_j^0(x), Z_j^1(x), \ldots, Z_j^p(x)\right)'.
\end{aligned}$$

The dimensions of e_1 and H are determined by the context without further explanation. Denote $e_1'\left(X(x)'K(x)X(x)\right)^{-1} \cdot X(x)'K(x)$ as $W^n(x)' = \left(W_1^n(x), \ldots, W_n^n(x)\right)$, which is the weight in (1). $S_n(x)$ converges in probability to $S(x) \equiv \Gamma f(x)$, which generates the equivalent kernel $K_p^*(\cdot)$.

Some calculus shows that in (8),

$$\hat{\theta} = \left(\underline{X}^{d\prime}\underline{X}^{d}\right)^{-1}\underline{X}^{d\prime}\underline{\mathbf{y}} \quad \text{and} \quad \hat{\alpha} = e_1'\left(\underline{X}^{d\prime}\underline{X}^{d}\right)^{-1}\underline{X}^{d\prime}\underline{\mathbf{y}}, \tag{21}$$

where

$$\begin{aligned}
\underline{X}^d &= \begin{pmatrix} X_1^{d\prime} - \mathbb{P}_{x_1}^n\left(X^d\right)' \\ \vdots \\ X_n^{d\prime} - \mathbb{P}_{x_n}^n\left(X^d\right)' \end{pmatrix} \equiv \begin{pmatrix} \underline{X}_1^{d\prime} \\ \vdots \\ \underline{X}_n^{d\prime} \end{pmatrix}_{n\times(q+1)} \equiv \left(\underline{X}^{0d}, \ldots, \underline{X}^{qd}\right)_{n\times(q+1)}, \\
&= X^d - \mathbf{e}_1'\left(\mathbf{X}'\mathbf{K}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{K}\mathbf{I}X^d = \left(I_n - \mathbf{e}_1'\left(\mathbf{X}'\mathbf{K}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{K}\mathbf{I}\right)X^d,
\end{aligned}$$

with $\mathbb{P}_{x_1}^n\left(X^d\right)$ operating on each column of X^d to get a row vector,

$$\begin{aligned}
X^d &= \begin{pmatrix} 1(x_1 \geq \pi) & (x_1 - \pi)1(x_1 \geq \pi) & \cdots & (x_1 - \pi)^q 1(x_1 \geq \pi) \\ \vdots & \vdots & \vdots & \vdots \\ 1(x_n \geq \pi) & (x_n - \pi)1(x_n \geq \pi) & \cdots & (x_n - \pi)^q 1(x_n \geq \pi) \end{pmatrix} \\
&\equiv \begin{pmatrix} X_1^{d\prime} \\ \vdots \\ X_n^{d\prime} \end{pmatrix}_{n\times(q+1)} \equiv \left(X^{0d}, \ldots, X^{qd}\right)_{n\times(q+1)}, \\
I_n &= \operatorname{diag}\{1, \ldots, 1\}_{n\times n}, \mathbf{e}_1 = \operatorname{diag}\{e_1, \ldots, e_1\}_{n(p+1)\times n} = I_n \otimes e_1, \\
\mathbf{X} &= \operatorname{diag}\{X(x_1), \ldots, X(x_n)\}_{n^2\times n(p+1)}, \\
e &= (1, 1, \ldots, 1)_{n\times 1}', \mathbf{I} = (e \otimes I_n)_{n^2\times n}, \otimes \text{ is the Kronecker product,} \\
\mathbf{K} &= \operatorname{diag}\{K_h(x_1), \ldots, K_h(x_n)\}_{n^2\times n^2},
\end{aligned}$$

and

$$\underline{\mathbf{y}} = \begin{pmatrix} y_1 - \mathbb{P}_{x_1}^n(\mathbf{y}) \\ \vdots \\ y_n - \mathbb{P}_{x_n}^n(\mathbf{y}) \end{pmatrix} = \left(I_n - \mathbf{e}_1'\left(\mathbf{X}'\mathbf{K}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{K}\mathbf{I}\right)\mathbf{y},$$

with

$$\mathbf{y} = m_q(\mathbf{x}) + X^d\theta + \boldsymbol{\varepsilon} \equiv \tilde{\mathbf{y}} + X^d\theta, m_q(\mathbf{x}) = \left(m_q(x_1), \ldots, m_q(x_n)\right)',$$

$$\mathbf{x} = (x_1, \ldots, x_n)', \boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)',$$

$$\tilde{\mathbf{y}} = m_q(\mathbf{x}) + \boldsymbol{\varepsilon} = (\tilde{y}_1, \ldots, \tilde{y}_n)', \ \tilde{y}_i \text{ is } \tilde{y}_i(\theta) \text{ evaluated at the true value of } \theta.$$

To simplify notations, we use $m(x)$ to denote $m_q(x)$ during the proof of Theorem 1.

Some explanations on $\hat{\theta}$ are in order. $\underline{X}^d$ and $\underline{\mathbf{y}}$ are the demeaned X^d and $\mathbf{y}$ by the "local polynomial operator" $\mathbb{P}_x^n$. $I_n - \mathbf{e}_1' (\mathbf{X}'\mathbf{K}\mathbf{X})^{-1} \mathbf{X}'\mathbf{K}\mathbf{I} \equiv I_n - \mathbb{P}_{\mathbf{x}}^n$ is like a demeaned operator on a vector in $\mathbb{R}^n$ at $\mathbf{x}$. Note that

$$\begin{aligned}
\left(\underline{X}^{d\prime}\underline{X}^d\right)^{-1} \underline{X}^{d\prime}\underline{\mathbf{y}} &= \left(\underline{X}^{d\prime}\underline{X}^d\right)^{-1} \underline{X}^{d\prime} \left(\underline{X}^d\theta + \tilde{\mathbf{y}} - \mathbb{P}_{\mathbf{x}}^n(\tilde{\mathbf{y}})\right) \\
&= \theta + H^{-1}\left(\frac{1}{nh}H^{-1}\underline{X}^{d\prime}\underline{X}^d H^{-1}\right)^{-1} \frac{1}{nh}H^{-1}\underline{X}^{d\prime}\underline{\tilde{\mathbf{y}}} \\
&\equiv \theta + H^{-1}\left(\frac{1}{nh}\underline{Z}^{d\prime}\underline{Z}^d\right)^{-1} \frac{1}{nh}\underline{Z}^{d\prime}\left(m(\mathbf{x}) - \tilde{m}(\mathbf{x}) + \boldsymbol{\varepsilon} - \tilde{\boldsymbol{\varepsilon}}\right) \\
&= \theta + H^{-1}\left(\frac{1}{nh}\sum_{l=1}^n \underline{Z}_l^d \underline{Z}_l^{d\prime}\right)^{-1} \\
&\quad \times \left(\frac{1}{nh}\sum_{l=1}^n \underline{Z}_l^d \left((m(x_l) - \tilde{m}(x_l) + \varepsilon_l - \tilde{\varepsilon}_l)\right)\right),
\end{aligned} \tag{22}$$

where $\underline{Z}^d = \underline{X}^d H^{-1}$ is the normalized $\underline{X}^d$ like $Z(x)$ in $\mathbb{P}_x^n$, $\underline{Z}_l^d = H^{-1}\underline{X}_l^d$, $l = 1, \ldots, n$, and $\underline{\tilde{\mathbf{y}}} = \tilde{\mathbf{y}} - \tilde{y}(\mathbf{x})$ with

$$\begin{aligned}
\tilde{y}(\mathbf{x}) &= (\tilde{y}(x_1), \ldots, \tilde{y}(x_n))' = \mathbb{P}_{\mathbf{x}}^n(\tilde{\mathbf{y}}) \\
&= \left(\mathbb{P}_{x_1}^n(m(\mathbf{x})), \ldots, \mathbb{P}_{x_n}^n(m(\mathbf{x}))\right)' + \left(\mathbb{P}_{x_1}^n(\boldsymbol{\varepsilon}), \ldots, \mathbb{P}_{x_n}^n(\boldsymbol{\varepsilon})\right)' \\
&\equiv (\tilde{m}(x_1), \ldots, \tilde{m}(x_n))' + (\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n)' \\
&\equiv \tilde{m}(\mathbf{x}) + \tilde{\boldsymbol{\varepsilon}}.
\end{aligned}$$

From Lemma 1 in Appendix C, $\underline{X}_l^d = 0$ for $|x_l - \pi| > h$, $l = 1, \ldots, n$, so only the x_l's in the h neighborhood of π will contribute to $\hat{\alpha}$. In consequence, the convergence rate of $\hat{\alpha}$ is $\sqrt{nh}$ instead of $\sqrt{n}$. In the proof follows, we will show that $\underline{Z}^{d\prime}(m(\mathbf{x}) - \tilde{m}(\mathbf{x}))$ contributes to the bias, and $\underline{Z}^{d\prime}(\boldsymbol{\varepsilon} - \tilde{\boldsymbol{\varepsilon}})$ contributes to the variance. Presence of $\tilde{\boldsymbol{\varepsilon}}$ in $\underline{Z}^{d\prime}(\boldsymbol{\varepsilon} - \tilde{\boldsymbol{\varepsilon}})$ makes the asymptotic variance derivation much more complicated than the usual LPE.

From (21) and (22),

$$\sqrt{nh}(\hat{\alpha} - \alpha) = e_1'\left(\frac{1}{nh}\sum_{l=1}^n \underline{Z}_l^d \underline{Z}_l^{d\prime}\right)^{-1}\left(\frac{1}{\sqrt{nh}}\sum_{l=1}^n \underline{Z}_l^d (m(x_l) - \tilde{m}(x_l) + \varepsilon_l - \tilde{\varepsilon}_l)\right).$$

We first analyze the numerator, then the denominator. For $1 \le i \le q+1$, the ith term of $\underline{Z}_l^d$ is

$$\begin{aligned}
&\left(\frac{x_l-\pi}{h}\right)^{i-1} 1\,(x_l \ge \pi) - \frac{1}{h^{i-1}} \mathbb{P}^n_{x_l}(X^{i-1,d}) \\
&= e_1' S_n^{-1}(x_l)(S_n^+(x_l) + S_n^-(x_l))e_1 \left(\frac{x_l-\pi}{h}\right)^{i-1} d_l \\
&\quad - e_1' S_n^{-1}(x_l) \frac{1}{n}\sum_{j=1}^n Z_j\,(x_l)\, k_h\left(x_j - x_l\right)\left(\frac{x_j-\pi}{h}\right)^{i-1} d_j \\
&= e_1' S_n^{-1}(x_l) \frac{1}{n}\sum_{j=1}^n Z_j\,(x_l)\, k_h\left(x_j - x_l\right)\left(Z_l^{i-1}(\pi)d_l - Z_j^{i-1}(\pi)\right) d_j \\
&\quad + e_1' S_n^{-1}(x_l) \frac{1}{n}\sum_{j=1}^n Z_j\,(x_l)\, k_h\left(x_j - x_l\right) Z_l^{i-1}(\pi) d_l d_j^c \\
&\equiv e_1' S_n^{-1}(x_l)\delta^+_{n,i-1}(x_l) + e_1' S_n^{-1}(x_l)\delta^-_{n,i-1}(x_l) \equiv e_1' S_n^{-1}(x_l)\delta_{n,i-1}(x_l). \qquad (23)
\end{aligned}$$

Here, $S_n^+(x)$ ($S_n^-(x)$) is replacing $Z_j(x)$ in $S_n(x)$ by $Z_j(x)d_j$ ($Z_j(x)d_j^c$), $\delta^+_{n,i-1}(x_l)$ plays the role of $-\hat{f}_+(x_l)d_l^c$, $\delta^-_{n,i-1}(x_l)$ plays the role of $\hat{f}_-(x_l)d_l$, and $S_n(x_l)$ plays the role of $\hat{f}(x_l)$ in Porter (2003).

Numerator

Concentrate on the ith term and take an expansion to linearize. We need different linearizations under Assumptions M(a) and M(b). We first discuss the case under Assumption M(a), and then under Assumption M(b).

Under Assumption M(a)

The ith term of the numerator is

$$\begin{aligned}
&\frac{1}{\sqrt{nh}}\sum_{l=1}^n e_1' S_n^{-1}(x_l)\delta_{n,i-1}(x_l)\,(m(x_l) - \tilde{m}(x_l) + \varepsilon_l - \tilde{\varepsilon}_l) \\
&= \frac{1}{\sqrt{nh}}\sum_{l=1}^n e_1' \overline{S}(x_l)^{-1}\overline{\delta}_{i-1}(x_l)\left(-\overline{L}(\Delta m(x_l)) + \varepsilon_l - \mathbb{P}_{x_l}(\boldsymbol{\varepsilon})\right) \\
&\quad + \frac{1}{\sqrt{nh}}\sum_{l=1}^n e_1' \overline{S}(x_l)^{-1}\overline{\delta}_{i-1}(x_l)\left(\overline{L}(\Delta m(x_l)) - L(\Delta m(x_l))\right)
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{1}{\sqrt{nh}} \sum_{l=1}^{n} L_{i-1}(x_l)\varepsilon_l + R_n \\
&\equiv \text{Term1} + \text{Term2} + \text{Term3} + R_n,
\end{aligned}
$$

where

$$
\begin{aligned}
L(\Delta m(x)) &= e_1' S^{-1}(x)\, \overline{r}(\Delta m(x)) - e_1' S^{-1}(x)\,(S_n(x) - S(x))\, S^{-1}(x)\, \overline{r}(\Delta m(x)) \\
&\quad + e_1' S^{-1}(x)\,(\tilde{r}(\Delta m(x)) - \overline{r}(\Delta m(x))), \\
\overline{L}(\Delta m(x)) &= e_1' S^{-1}(x)\, \overline{r}(\Delta m(x)) - e_1' S^{-1}(x)\left(\overline{S}(x) - S(x)\right) S^{-1}(x)\, \overline{r}(\Delta m(x)), \\
L_{i-1}(x) &= e_1' \overline{S}^{-1}(x)\left(\delta_{n,i-1}(x) - \overline{\delta}_{i-1}(x)\right) \\
&\quad - e_1' \overline{S}^{-1}(x)\left(S_n(x) - \overline{S}(x)\right) \overline{S}^{-1}(x) \overline{\delta}_{i-1}(x), \\
\mathbb{P}_x(\boldsymbol{\varepsilon}) &= e_1' S^{-1}(x)\, \tilde{r}(\varepsilon(x)), \\
\overline{\delta}_{i-1}(x) &= \overline{\delta}_{i-1}^{+}(x) + \overline{\delta}_{i-1}^{-}(x),
\end{aligned}
$$

with

$$
\begin{aligned}
\tilde{r}(\Delta m(x)) &= \frac{1}{n} \sum_{j=1}^{n} Z_j(x) k_h\left(x_j - x\right) \left\{ m(x_j) - m(x) - \sum_{\ell=1}^{q} \frac{m^{(\ell)}(x)}{\ell!} \left(x_j - x\right)^{\ell} \right\} \\
\tilde{r}(\varepsilon(x)) &= \frac{1}{n} \sum_{j=1}^{n} Z_j(x) k_h\left(x_j - x\right) \varepsilon_j, \\
\overline{r}(\Delta m(x)) &= \int \delta(u) f(x+uh) \left\{ m(x+uh) - m(x) - \sum_{\ell=1}^{q} \frac{m^{(\ell)}(x)}{\ell!} (uh)^{\ell} \right\} du, \\
\overline{S}(x) &= E\left[Z_j(x)\, Z_j'(x)\, k_h\left(x_j - x\right)\right] \\
&= \left(\int u^{i+j-2} k(u)\, f(x+uh) du \right)_{(p+1)\times(p+1)}, \\
\overline{\delta}_{i-1}^{+}(x) &= E\left[Z_j(x)\, k_h\left(x_j - x\right) \left(\left(\frac{x-\pi}{h}\right)^{i-1} 1(x \geq \pi) - \left(\frac{x_j - \pi}{h}\right)^{i-1} \right) d_j \right] \\
&= \left(\int_{-1}^{1} \delta(u) \left(\left(\frac{x-\pi}{h}\right)^{i-1} 1(x \geq \pi) - \left(\frac{x-\pi}{h} + u\right)^{i-1} \right) \right. \\
&\qquad \left. f(x+uh) 1\left(x + uh \geq \pi\right) du \right)_{(p+1)\times 1},
\end{aligned}
$$

$$\overline{\delta}_{i-1}^{-}(x) = E\left[Z_j(x)\,k_h\left(x_j - x\right)\left(\frac{x-\pi}{h}\right)^{i-1} 1(x \geq \pi)d_j^c\right]$$
$$= \left(\int_{-1}^{1} \delta(u)\left(\frac{x-\pi}{h}\right)^{i-1} 1(x \geq \pi) f(x+uh) 1(x+uh < \pi) du\right)_{(p+1)\times 1},$$

and R_n is the remainder term including quadratic terms in the expansion:

$$R_n = -\frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}(x_l)R(\tilde{y}(x_l))$$
$$+ \frac{1}{\sqrt{nh}}\sum_{l=1}^{n} R_{i-1}(x_l)\left(m(x_l) - \tilde{y}(x_l) + \varepsilon_l\right)$$
$$+ \frac{1}{\sqrt{nh}}\sum_{l=1}^{n} L_{i-1}(x_l)\left(m(x_l) - \tilde{y}(x_l)\right),$$

with

$$R(\tilde{y}(x)) = e_1'S^{-1}(x)\left(S_n(x) - S(x)\right)S^{-1}(x)\left(S_n(x) - S(x)\right)S_n^{-1}(x)\overline{r}(\Delta m(x))$$
$$- e_1'S^{-1}(x)\left(S_n(x) - S(x)\right)S_n^{-1}(x)\left(\tilde{r}(\tilde{y}(x)) - \overline{r}(\Delta m(x))\right),$$
$$\tilde{r}(\tilde{y}(x)) = \tilde{r}(\Delta m(x)) + \tilde{r}(\varepsilon(x)),$$
$$R_{i-1}(x) = e_1'\overline{S}^{-1}(x)\left(S_n(x) - \overline{S}(x)\right)\overline{S}^{-1}(x)\left(S_n(x) - \overline{S}(x)\right)S_n^{-1}(x)\overline{\delta}_{i-1}(x)$$
$$- e_1'\overline{S}^{-1}(x)\left(S_n(x) - \overline{S}(x)\right)S_n^{-1}(x)\left(\delta_{n,i-1}(x) - \overline{\delta}_{i-1}(x)\right).$$

The validity of including qth order Taylor expansion of $m(\cdot)$ in $\tilde{r}(\Delta m(x))$ and $\overline{r}(\Delta m(x))$ can be justified by the discrete orthogonality relation in the LPE (see, e.g., (2.4) of Fan et al. (1997); note that $p \geq q$), $L(\Delta m(x))$ is the linear expansion of $\mathbb{P}_x^n(m(\mathbf{x})) - m(x)$ as shown in Lemma 2 of Appendix C, and $\overline{L}(\Delta m(x))$ is its mean. $L_{i-1}(x)$ is the linear expansion of $e_1'S_n^{-1}(x)\delta_{n,i-1}(x)$ at $e_1'\overline{S}^{-1}(x)\overline{\delta}_{i-1}(x)$. Note that $e_1'S_n^{-1}(x)\delta_{n,i-1}(x)$ is linearized at $\overline{S}^{-1}(x)$ and $\overline{\delta}_{i-1}(x)$ instead of their limits which are $S^{-1}(x)$ and 0, respectively.[7] This is mainly because $\overline{\delta}_{i-1}(x)$ is not a smooth function of x when x is in a neighborhood of π. As a result, $S_n^{-1}(x)$ cannot be linearized at $S^{-1}(x)$; otherwise, $R_{i-1}(x)$ cannot be a higher-order term.

Our analysis includes three steps. In step 1, we show $R_n = o_p(1)$. In step 2, we show Term 3 $= o_p(1)$ and Term 2 $= o_p(1)$. In step 3, we show $-\overline{L}(m(x_l))$ in Term 1

[7] In Porter (2003), $\overline{f}_+(x_l)d_l^c$ and $\overline{f}_-(x_l)d_l$ converges to 0 for a fixed x_l when h converges to zero. This result can be applied to $\overline{\delta}_0^-(x_l)$ and $\overline{\delta}_0^+(x_l)$. For $i > 1$, it is still true for $h^{i-1}\overline{\delta}_{i-1}^+(x_l)$ and $h^{i-1}\overline{\delta}_{i-1}^-(x_l)$.

contributes to the bias, and $\varepsilon_l - \tilde{\varepsilon}_l$ contributes to the variance. Although there is randomness in Term 2, it does not contribute to the asymptotic distribution. With the three steps in hand, the Liapunov central limit theorem is applied to find the asymptotic distribution.

Step 1. First, some basic results. From Lemma B5 of Porter (2003) and Lemmas 3 and 4 of Appendix C, $\sup_{x\in N_0}|S_n(x)-S(x)| = \left(O_p\left(\sqrt{\frac{\ln n}{nh}}\right)+h\right)$, $\sup_{x\in N_0} S_n^{-1}(x) = \sup_{x\in N_0}\overline{S}^{-1}(x) + o_p(1) = O_p(1)$, $\sup_{x\in N_0}|\delta_{n,i-1}(x)-\overline{\delta}_{i-1}(x)| = O_p\left(\sqrt{\frac{\ln n}{nh}}\right)$, $\sup_{x\in N_0}\overline{\delta}_{i-1}(x) = O(1)$, $\sup_{x\in N_0} e_1'\overline{S}^{-1}(x)\overline{\delta}_{i-1}(x) = O(1)$, $\sup_{x\in N_0}\overline{r}(\Delta m(x)) = O(h^{q+1})$, $\sup_{x\in N_0}|\tilde{r}(\tilde{y}(x))-\overline{r}(\Delta m(x))| = O_p\left(\sqrt{\frac{\ln n}{nh}}\right)$, $\sup_{x\in N_0}|\tilde{y}(x)-m(x)| = O_p\left(\sqrt{\frac{\ln n}{nh}}+h^{q+1}\right)$, $\frac{1}{nh}\sum_{l=1}^n|\varepsilon_l|1(\pi-h\le x_l\le\pi+h) = O_p(1)$, $\sup_{x\in N_0}\frac{1}{f(x)} = O(1)$, where $N_0=[\pi-h,\pi+h]$.

(i)
$$\frac{1}{\sqrt{nh}}\sum_{l=1}^n e_1'\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}(x_l)\cdot e_1'S^{-1}(x_l)\left(S_n(x_l)-S(x_l)\right)S^{-1}(x_l)$$
$$\left(S_n(x_l)-S(x_l)\right)S_n^{-1}(x_l)\overline{r}(\Delta m(x_l))$$
$$\approx \frac{1}{\sqrt{nh}}\sum_{l=1}^n e_1'\Gamma^{-1}\overline{\delta}_{i-1}(x_l)\cdot e_1'\Gamma^{-1}$$
$$\left(S_n(x_l)-S(x_l)\right)\Gamma^{-1}\left(S_n(x_l)-S(x_l)\right)\Gamma^{-1}\overline{r}(\Delta m(x_l))$$
$$\approx \frac{1}{\sqrt{nh}}\sum_{l=1}^n O(1)\left(O_p\left(\sqrt{\frac{\ln n}{nh}}\right)+h\right)\left(O_p\left(\sqrt{\frac{\ln n}{nh}}\right)+h\right)O(h^{q+1})$$
$$= \sqrt{nh}O_p\left(\sqrt{\frac{\ln n}{nh}}+h\right)O_p\left(\sqrt{\frac{\ln n}{nh}}+h\right)O(h^{q+1})$$
$$= O_p\left(\left(\frac{\ln n}{\sqrt{nh}}+h\sqrt{\ln n}+h^2\sqrt{nh}\right)h^{q+1}\right),$$

and

$$\frac{1}{\sqrt{nh}}\sum_{l=1}^n R_{i-1}(x_l)(m(x_l)-\tilde{m}(x_l))$$
$$\approx \sqrt{nh}\left[O_p\left(\sqrt{\frac{\ln n}{nh}}\right)O_p\left(\sqrt{\frac{\ln n}{nh}}\right)+O_p\left(\sqrt{\frac{\ln n}{nh}}\right)O_p\left(\sqrt{\frac{\ln n}{nh}}\right)\right]$$
$$\times\left(O_p\left(\sqrt{\frac{\ln n}{nh}}\right)+h^{q+1}\right) = O_p\left(\frac{\ln n\sqrt{\ln n}}{nh}+\frac{\ln n}{\sqrt{nh}}h^{q+1}\right).$$

(ii)
$$\frac{1}{\sqrt{nh}}\sum_{l=1}^{n} R_{i-1}(x_l)\varepsilon_l \approx \sqrt{nh}O_p\left(\sqrt{\frac{\ln n}{nh}}\right)O_p\left(\sqrt{\frac{\ln n}{nh}}\right) \times \left(\frac{1}{nh}\sum_{l=1}^{n}\left|\varepsilon_l\right| 1\left(\pi - h \le x_l \le \pi + h\right)\right) = O_p\left(\frac{\ln n}{\sqrt{nh}}\right).$$

(iii)
$$\frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}(x_l)\cdot e_1' S^{-1}(x_l)\left(S_n(x_l) - S(x_l)\right)S_n^{-1}(x_l) \times \left(\tilde{r}(\tilde{y}(x_l)) - \overline{r}(\Delta m(x_l))\right) \approx \sqrt{nh}O_p\left(\sqrt{\frac{\ln n}{nh}} + h\right)O_p\left(\sqrt{\frac{\ln n}{nh}}\right) = O_p\left(\frac{\ln n}{\sqrt{nh}} + h\sqrt{\ln n}\right).$$

(iv)
$$\frac{1}{\sqrt{nh}}\sum_{l=1}^{n} L_{i-1}(x_l)\left(m(x_l) - \tilde{y}(x_l)\right) \approx \sqrt{nh}O_p\left(\sqrt{\frac{\ln n}{nh}}\right)\left(O_p\left(\sqrt{\frac{\ln n}{nh}}\right) + h^{q+1}\right) = O_p\left(\frac{\ln n}{\sqrt{nh}} + h^{q+1}\sqrt{\ln n}\right).$$

From Assumption B(a) and (i)–(iv), $R_n = o_p(1)$.

Step 2. To prove Term $3 = o_p(1)$, we will use the U and V-statistic projection. First, note that

$$\begin{aligned}\frac{1}{\sqrt{nh}}\sum_{l=1}^{n} L_{i-1}(x_l)\varepsilon_l &= \frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\overline{S}^{-1}(x_l)\left(\delta_{n,i-1}^{+}(x_l) - \overline{\delta}_{i-1}^{+}(x_l)\right)\varepsilon_l \\ &+ \frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\overline{S}^{-1}(x_l)\left(\delta_{n,i-1}^{-}(x_l) - \overline{\delta}_{i-1}^{-}(x_l)\right)\varepsilon_l \\ &- \frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\overline{S}^{-1}(x_l)\left(S_n(x_l) - \overline{S}(x_l)\right)\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}^{+}(x_l)\varepsilon_l \\ &- \frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\overline{S}^{-1}(x_l)\left(S_n(x_l) - \overline{S}(x_l)\right)\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}^{-}(x_l)\varepsilon_l \\ &\equiv T1 + T2 + T3 + T4.\end{aligned}$$

Let $z_l = (x_l, \varepsilon_l)'$. For $T1$,

$$\frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\overline{S}^{-1}(x_l)\delta_{n,i-1}^{+}(x_l)\varepsilon_l = \sqrt{\frac{n}{h}}\frac{1}{n^2}\sum_{l=1}^{n}\sum_{j=1}^{n} b_n(z_l, z_j),$$

where

$$b_n(z_l, z_j) = e_1' \overline{S}^{-1}(x_l) Z_j(x_l) k_h(x_j - x_l) \left(\left(\frac{x_l - \pi}{h} \right)^{i-1} d_l - \left(\frac{x_j - \pi}{h} \right)^{i-1} \right) d_j \varepsilon_l.$$

Note that $b_n(z_l, z_l) = 0$ so that this term is a U-statistic. Under the Assumptions in Section 3.3, it is easy, although tedious in notations, to show that $E\left[b_n(z_l, z_j)^2\right] = O(1)$. Then by standard U-statistic projection results,

$$T1 = \sqrt{\frac{n}{h}} O_p \left(\frac{\left(E\left[b_n(z_l, z_j)^2\right]\right)^{1/2}}{n} \right) = O_p\left(\frac{1}{\sqrt{nh}}\right) = o_p(1).$$

$T2$ follows similarly.

For $T3$, let

$$b_n(z_l, z_j) = e_1' \overline{S}^{-1}(x_l) \left(Z_j(x_l) Z_j'(x_l) k_h(x_j - x_l) \right) \overline{S}^{-1}(x_l) \overline{\delta}_{i-1}^{+}(x_l) \varepsilon_l.$$

Then

$$\frac{1}{\sqrt{nh}} \sum_{l=1}^{n} e_1' \overline{S}^{-1}(x_l) S_n(x_l) \overline{S}^{-1}(x_l) \overline{\delta}_{i-1}^{+}(x_l) \varepsilon_l = \sqrt{\frac{n}{h}} \frac{1}{n^2} \sum_{l=1}^{n} \sum_{j=1}^{n} b_n(z_l, z_j).$$

As above, $E\left[b_n(z_l, z_j)^2\right] = O(1)$. Also, it is easy to show that $E\left[\left|b_n(z_l, z_l)\right|\right] = O(1)$ for n large enough. By a V-statistic projection theorem; see, e.g., Lemma 8.4 of Newey and McFadden (1994),

$$T3 = \sqrt{\frac{n}{h}} O_p \left(\frac{\left(E\left[b_n(z_l, z_j)^2\right]\right)^{1/2}}{n} + \frac{E\left[\left|b_n(z_l, z_l)\right|\right]}{n} \right) = O_p\left(\frac{1}{\sqrt{nh}}\right).$$

$T4$ follows similarly.

To prove Term2 $= o_p(1)$, we will use the V-statistic projection again. First, note that

$$\text{Term2} = \begin{pmatrix} \frac{1}{\sqrt{nh}} \sum_{l=1}^{n} e_1' \overline{S}(x_l)^{-1} \overline{\delta}_{i-1}(x_l) e_1' S^{-1}(x_l) \\ \left(S_n(x_l) - \overline{S}(x_l)\right) S^{-1}(x_l) \overline{r}(\Delta m(x_l)) \\ - \frac{1}{\sqrt{nh}} \sum_{l=1}^{n} e_1' \overline{S}(x_l)^{-1} \\ \overline{\delta}_{i-1}(x_l) e_1' S^{-1}(x_l) \left(\tilde{r}(\Delta m(x_l)) - \overline{r}(\Delta m(x_l))\right) \end{pmatrix} \equiv T5 - T6.$$

For $T5$, let

$$\begin{aligned} b_n(x_l, x_j) &= e_1' \overline{S}(x_l)^{-1} \overline{\delta}_{i-1}(x_l) e_1' S^{-1}(x_l) \\ &\quad \times \left(Z_j(x_l) Z_j'(x_l) k_h(x_j - x_l) \right) S^{-1}(x_l) \overline{r}(\Delta m(x_l)). \end{aligned}$$

Then

$$\begin{aligned}&\frac{1}{\sqrt{nh}}\sum_{l=1}^{n}e_1'\overline{S}(x_l)^{-1}\overline{\delta}_{i-1}(x_l)e_1'S^{-1}(x_l)\,S_n(x_l)\,S^{-1}(x_l)\,\overline{r}(\Delta m(x_l))\\&=\sqrt{\frac{n}{h}}\frac{1}{n^2}\sum_{l=1}^{n}\sum_{j=1}^{n}b_n(x_l,x_j).\end{aligned}$$

It is easy to show that $E\left[b_n(x_l,x_j)^2\right]=O(h^{2(q+1)})$ and $E\left[\left|b_n(x_l,x_j)\right|\right]=O(h^{q+1})$, so

$$T5=\sqrt{\frac{n}{h}}O_p\left(\frac{\left(E\left[b_n(x_l,x_j)^2\right]\right)^{1/2}}{n}+\frac{E\left[\left|b_n(x_l,x_j)\right|\right]}{n}\right)=O_p\left(\frac{h^{q+1}}{\sqrt{nh}}\right)=o_p(1).$$

A similar proof can be applied to $T6$ except now

$$\begin{aligned}b_n(x_l,x_j)&=e_1'\overline{S}(x_l)^{-1}\overline{\delta}_{i-1}(x_l)e_1'S^{-1}(x_l)\,Z_j(x_l)k_h\left(x_j-x_l\right)\\&\quad\times\left\{m(x_j)-m(x_l)-\sum_{\ell=1}^{q}\frac{m^{(\ell)}(x_l)}{\ell!}\left(x_j-x_l\right)^{\ell}\right\}.\end{aligned}$$

Step 3. First, analyze the bias term $\frac{1}{\sqrt{nh}}\sum_{l=1}^{n}e_1'\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}(x_l)(-\overline{L}(\Delta m\ (x_l)))$:

$$\begin{aligned}&E\left[\frac{1}{\sqrt{nh}}\sum_{l=1}^{n}e_1'\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}(x_l)\overline{L}(\Delta m(x_l))\right]\\&\approx\sqrt{\frac{n}{h}}\int\left[\int_{\frac{\pi-x}{h}}^{1}K_p^*(u)\left(\left(\frac{x-\pi}{h}\right)^{i-1}1(x\ge\pi)-\left(\frac{x-\pi}{h}+u\right)^{i-1}\right)f(x+uh)du\right.\\&\quad\left.+\int_{-1}^{\frac{\pi-x}{h}}K_p^*(u)\left(\frac{x-\pi}{h}\right)^{i-1}1(x\ge\pi)f(x+uh)du\right]\cdot\\&\frac{e_1'\Gamma^{-1}}{f(x)}\left(\int\delta(u)f(x+uh)\left\{m(x+uh)-m(x)-\sum_{\ell=1}^{q}\frac{m^{(\ell)}(x)}{\ell!}(uh)^{\ell}\right\}du\right)dx\\&=\sqrt{nh}\int\left[\int_{-1}^{1}K_p^*(u)\,w^{i-1}1(w\ge0)\frac{f(\pi+wh+uh)}{f(\pi+wh)}du\right.\\&\quad\left.\times-\int_{-w}^{1}K_p^*(u)\,(w+u)^{i-1}\frac{f(\pi+wh+uh)}{f(\pi+wh)}du\right]\\&\left(\int K_p^*(u)\,f(\pi+wh+uh)\right.\end{aligned}$$

$$\times \left\{ m(\pi + wh + uh) - m(\pi + wh) - \sum_{\ell=1}^{q} \frac{m^{(\ell)}(\pi + wh)}{\ell!} (uh)^{\ell} \right\} du \Bigg) dw$$

$$\approx \sqrt{nh} f(\pi) \int_0^1 \left[w^{i-1} - \int_{-w}^{1} K_p^*(u)(w+u)^{i-1} du \right]$$

$$\times \left(\int_{-w}^{1} K_p^*(u) \frac{m_+^{(q+1)}(\pi)}{(q+1)!} \left(((w+u)h)^{q+1} - (wh)^{q+1} \right) du \right.$$

$$\left. + \int_{-1}^{-w} K_p^*(u) \frac{m_-^{(q+1)}(\pi)((w+u)h)^{q+1} - m_+^{(q+1)}(\pi)(wh)^{q+1}}{(q+1)!} du \right) dw$$

$$- \sqrt{nh} f(\pi) \int_{-1}^{0} \left(\int_{-w}^{1} K_p^*(u)(w+u)^{i-1} du \right)$$

$$\times \left(\int_{-w}^{1} K_p^*(u) \frac{m_+^{(q+1)}(\pi)((w+u)h)^{q+1} - m_-^{(q+1)}(\pi)(wh)^{q+1}}{(q+1)!} du \right.$$

$$\left. + \int_{-1}^{-w} K_p^*(u) \frac{m_-^{(q+1)}(\pi)}{(q+1)!} \left(((w+u)h)^{q+1} - (wh)^{q+1} \right) du \right) dw$$

$$\equiv \sqrt{nh} h^{q+1} \frac{f(\pi)}{(q+1)!} \left[m_+^{(q+1)}(\pi) Q_{pq}^+(i) + m_-^{(q+1)}(\pi) Q_{pq}^-(i) \right],$$

where the third equality is from the Taylor expansion of both $m(\pi + wh + uh)$ and $m(\pi + wh)$ at $m(\pi)$.

Second, analyze the variance term $\frac{1}{\sqrt{nh}} \sum_{l=1}^{n} e_1' \overline{S}(x_l)^{-1} \overline{\delta}_{i-1}(x_l) \left(\varepsilon_l - \mathbb{P}_{x_l}(\boldsymbol{\varepsilon}) \right)$. By the V-statistic projection, this term is statistically equivalent to

$$\frac{1}{\sqrt{nh}} \sum_{j=1}^{n} e_1' \overline{S}(x_j)^{-1} \overline{\delta}_{i-1}(x_j) \varepsilon_j - \frac{1}{\sqrt{nh}} \sum_{j=1}^{n} E_{x_l} \left[e_1' \overline{S}(x_l)^{-1} \overline{\delta}_{i-1}(x_l) \frac{K_p^*\left(\frac{x_j - x_l}{h}\right)}{hf(x_l)} \right] \varepsilon_j,$$

where E_{x_l} is taking expectation with respect to x_l. The (i,k) term of the variance matrix is

$$\frac{1}{nh} E \left\{ \left[\sum_{j=1}^{n} \left(e_1' \overline{S}(x_j)^{-1} \overline{\delta}_{i-1}(x_j) - E_{x_l} \left[e_1' \overline{S}(x_l)^{-1} \overline{\delta}_{i-1}(x_l) \frac{K_p^*\left(\frac{x_j - x_l}{h}\right)}{hf(x_l)} \right] \right) \varepsilon_j \right] \right.$$

$$\left. \cdot \left[\sum_{j=1}^{n} \left(e_1' \overline{S}(x_j)^{-1} \overline{\delta}_{k-1}(x_j) - E_{x_l} \left[e_1' \overline{S}(x_l)^{-1} \overline{\delta}_{k-1}(x_l) \frac{K_p^*\left(\frac{x_j - x_l}{h}\right)}{hf(x_l)} \right] \right) \varepsilon_j \right] \right\}$$

$$= \sigma_+^2(\pi) \int_0^1 \left(e_1' \overline{S}(\pi + wh)^{-1} \overline{\delta}_{i-1}(\pi + wh) - E_{x_l} \left[e_1' \overline{S}(x_l)^{-1} \overline{\delta}_{i-1}(x_l) \frac{K_p^*\left(\frac{\pi + wh - x_l}{h}\right)}{hf(x_l)} \right] \right)$$

$$\left(e_1'\overline{S}(\pi+wh)^{-1}\overline{\delta}_{k-1}(\pi+wh)-E_{x_l}\left[e_1'\overline{S}(x_l)^{-1}\overline{\delta}_{k-1}(x_l)\frac{K_p^*\left(\frac{\pi+wh-x_l}{h}\right)}{hf(x_l)}\right]\right)f(\pi+wh)dw$$
$$+\sigma_-^2(\pi)\int_{-1}^0\left(e_1'\overline{S}(\pi+wh)^{-1}\overline{\delta}_{i-1}(\pi+wh)-E_{x_l}\left[e_1'\overline{S}(x_l)^{-1}\overline{\delta}_{i-1}(x_l)\frac{K_p^*\left(\frac{\pi+wh-x_l}{h}\right)}{hf(x_l)}\right]\right)$$
$$\left(e_1'\overline{S}(\pi+wh)^{-1}\overline{\delta}_{k-1}(\pi+wh)-E_{x_l}\left[e_1'\overline{S}(x_l)^{-1}\overline{\delta}_{k-1}(x_l)\frac{K_p^*\left(\frac{\pi+wh-x_l}{h}\right)}{hf(x_l)}\right]\right)f(\pi+wh)dw$$
$$\approx f(\pi)\left[\sigma_+^2(\pi)\cdot\Omega_p^+(i,k)+\sigma_-^2(\pi)\cdot\Omega_p^-(i,k)\right].$$

To apply the Liapunov central limit theorem, it suffices that for some $\zeta>0$,

$$\sum_{j=1}^n E\left|\frac{1}{\sqrt{nh}}\left[e_1'\overline{S}(x_j)^{-1}\overline{\delta}_{i-1}(x_j)-E_{x_l}\left[e_1'\overline{S}(x_l)^{-1}\overline{\delta}_{i-1}(x_l)\frac{K_p^*\left(\frac{x_j-x_l}{h}\right)}{hf(x_l)}\right]\right]\varepsilon_j\right|^{2+\zeta}$$
$$=o(h^{(i-1)(2+\zeta)}).$$

The left-hand side is bounded by $C\sum_{j=1}^n\left[E\left|\frac{1}{\sqrt{nh}}e_1'\overline{S}(x_j)^{-1}\overline{\delta}_{i-1}(x_j)\varepsilon_j\right|^{2+\zeta}+E\left|\frac{1}{\sqrt{nh}}E_{x_l}\left[e_1'\overline{S}(x_l)^{-1}\overline{\delta}_{i-1}(x_l)\frac{K_p^*(\frac{x_j-x_l}{h})}{hf(x_l)}\right]\varepsilon_j\right|^{2+\zeta}\right]$ for some $C>0$. Now,

$$\sum_{j=1}^n E\left|\frac{1}{\sqrt{nh}}e_1'\overline{S}(x_j)^{-1}\overline{\delta}_{i-1}(x_j)\varepsilon_j\right|^{2+\zeta}$$
$$\le\frac{1}{(nh)^{\zeta/2}}\sup_{x\in N_0}E\left[|\varepsilon|^{2+\zeta}\,\middle|\,x\right]\sup_{x\in N_0}\left|e_1'\overline{S}(x)^{-1}\overline{\delta}_{i-1}(x)\right|^{2+\zeta}\frac{1}{h}E\left[1\left(\pi-h\le x\le\pi+h\right)\right]$$
$$\le O\left(\frac{1}{(nh)^{\zeta/2}}\right)=o(1).$$

Another term can be bounded similarly, so the Liapunov condition is satisfied.

Under Assumption M(b)

Under Assumption 2(b), redefine

$$\tilde{r}(\Delta m(x))=\frac{1}{n}\sum_{j=1}^n Z_j(x)k_h\left(x_j-x\right)\left\{m(x_j)-m(x)-\sum_{\ell=1}^p\frac{m^{(\ell)}(x)}{\ell!}\left(x_j-x\right)^\ell\right\},$$
$$\overline{r}(\Delta m(x))=\int\delta(u)f(x+uh)\left\{m(x+uh)-m(x)-\sum_{\ell=1}^p\frac{m^{(\ell)}(x)}{\ell!}(uh)^\ell\right\}du.$$

When p is odd, there is no essential change in the proof above except that q is replaced by p in a few places. The asymptotic variance remains the same, but the form of the bias changes:

$$
\begin{aligned}
&E\left[\frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}(x_l)\overline{L}(\Delta m(x_l))\right]\\
&\quad\approx\sqrt{\frac{n}{h}}\int\left[\int_{\frac{\pi-x}{h}}^{1} K_p^*(u)\right.\\
&\quad\times\left(\left(\frac{x-\pi}{h}\right)^{i-1}1(x\geq\pi)-\left(\frac{x-\pi}{h}+u\right)^{i-1}\right)f(x+uh)du\\
&\quad\left.+\int_{-1}^{\frac{\pi-x}{h}} K_p^*(u)\left(\frac{x-\pi}{h}\right)^{i-1}1(x\geq\pi)f(x+uh)du\right]\cdot\\
&\frac{e_1'\Gamma^{-1}}{f(x)}\left(\int\delta(u)f(x+uh)\left\{m(x+uh)-m(x)-\sum_{\ell=1}^{p}\frac{m^{(\ell)}(x)}{\ell!}(uh)^{\ell}\right\}du\right)dx\\
&\quad=\sqrt{nh}\int\left[\int_{-1}^{1} K_p^*(u)\,w^{i-1}1(w\geq 0)\frac{f(\pi+wh+uh)}{f(\pi+wh)}du\right.\\
&\quad\left.-\int_{-w}^{1} K_p^*(u)\,(w+u)^{i-1}\frac{f(\pi+wh+uh)}{f(\pi+wh)}du\right]\\
&e_1'\Gamma^{-1}\left(\int\delta(u)f(\pi+wh+uh)\right.\\
&\quad\times\left.\left\{m(\pi+wh+uh)-m(\pi+wh)-\sum_{\ell=1}^{p}\frac{m^{(\ell)}(\pi+wh)}{\ell!}(uh)^{\ell}\right\}du\right)dw\\
&\quad\approx\sqrt{nh}f(\pi)\int_0^1\left[w^{i-1}-\int_{-w}^{1}K_p^*(u)\,(w+u)^{i-1}du\right]\\
&\quad\times\left(\int_{-1}^{1}K_p^*(u)\,\frac{m^{(p+1)}(\pi)}{(p+1)!}(uh)^{p+1}du\right)dw\\
&\quad-\sqrt{nh}f(\pi)\int_{-1}^{0}\left(\int_{-w}^{1}K_p^*(u)\,(w+u)^{i-1}du\right)\\
&\quad\times\left(\int_{-1}^{1}K_p^*(u)\,\frac{m^{(p+1)}(\pi)}{(p+1)!}(uh)^{p+1}du\right)dw\\
&\quad=\sqrt{nh}h^{p+1}\frac{f(\pi)m^{(p+1)}(\pi)}{(p+1)!}\int_{-1}^{1}K_p^*(u)\,u^{p+1}du\\
&\quad\times\left[\int_0^1 K_p^*(\overline{\delta}_{i-1}^{+}(w))dw+\int_{-1}^{0}K_p^*(\overline{\delta}_{i-1}^{-}(w))dw\right]\\
&\quad=\sqrt{nh}h^{p+1}\frac{f(\pi)m^{(p+1)}(\pi)}{(p+1)!}\int_{-1}^{1}K_p^*(u)\,u^{p+1}duQ_p(i),
\end{aligned}
$$

where note that $m^{(p+1)}(\pi)=m_0^{(p+1)}(\pi)$.

When p is even, there are some changes in Steps 1 and 3. In Step 1(i),

$$\frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}(x_l)\cdot e_1' S^{-1}(x_l)$$

$$\times\left(S_n(x_l)-S(x_l)\right)S^{-1}(x_l)\left(S_n(x_l)-S(x_l)\right)S_n^{-1}(x_l)\overline{r}(\Delta m(x_l))$$

$$\approx\frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\Gamma^{-1}\left(S_n(x_l)-S(x_l)\right)\Gamma^{-1}\left(S_n(x_l)-S(x_l)\right)\Gamma^{-1}\overline{r}(\Delta m(x_l))$$

$$\approx\frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\Gamma^{-1}\left(O_p\left(\sqrt{\frac{\ln n}{nh}}\right)+O(h)\right)\Gamma^{-1}$$

$$\times\left(O_p\left(\sqrt{\frac{\ln n}{nh}}\right)+O(h)\right)\Gamma^{-1}\int\delta(u)u^{p+1}du O(h^{p+1})$$

$$=O_p\left(\left(\frac{\ln n}{\sqrt{nh}}+h\sqrt{\ln n}+h^2\sqrt{nh}\right)h^{p+1}\right).$$

In Step 3, the bias changes:

$$E\left[\frac{1}{\sqrt{nh}}\sum_{l=1}^{n} e_1'\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}(x_l)\overline{L}(\Delta m(x_l))\right]$$

$$\approx\sqrt{nh}\int\left[\int_{-1}^{1}K_p^*(u)\,w^{i-1}1(w\geq 0)du-\int_{-w}^{1}K_p^*(u)\,(w+u)^{i-1}du\right]$$

$$e_1'\Gamma^{-1}\left(\int\delta(u)\left\{f(\pi)+f'(\pi)\,(w+u)\,h\right\}\right.$$

$$\left.\times\left\{\frac{m^{(p+1)}(\pi)}{(p+1)!}\,(uh)^{p+1}+\frac{m^{(p+2)}(\pi)}{(p+2)!}\,(uh)^{p+2}\right\}du\right)dw$$

$$=\sqrt{nh}h^{p+2}\int\left[\int_{-1}^{1}K_p^*(u)\,w^{i-1}1(w\geq 0)du-\int_{-w}^{1}K_p^*(u)\,(w+u)^{i-1}du\right]$$

$$\left(\int_{-1}^{1}K_p^*(u)\left\{f(\pi)\frac{m^{(p+2)}(\pi)}{(p+2)!}u^{p+2}+f'(\pi)\frac{m^{(p+1)}(\pi)}{(p+1)!}\,(w+u)\,u^{p+1}\right\}du\right)dw$$

$$=\sqrt{nh}h^{p+2}\left[\int_0^1 K_p^*(\overline{\delta}_{i-1}^{+}(w))dw+\int_{-1}^{0}K_p^*(\overline{\delta}_{i-1}^{-}(w))dw\right]$$

$$\left(\int_{-1}^{1}K_p^*(u)\,u^{p+2}du\right)\left[f(\pi)\frac{m^{(p+2)}(\pi)}{(p+2)!}+f'(\pi)\frac{m^{(p+1)}(\pi)}{(p+1)!}\right],$$

where note that $m^{(p+2)}(\pi) = m_0^{(p+2)}(\pi)$ and $m^{(p+1)}(\pi) = m_0^{(p+1)}(\pi)$. Note also that $e_1' S^{-1}(x)\left(\overline{S}(x) - S(x)\right) S^{-1}(x)\,\overline{r}(m(x))$ in $\overline{L}(\Delta m(x))$ does not contribute to the bias regardless of whether p is odd or even since it only contributes a higher-order term in both cases.

Denominator

We get the asymptotic limit of $\frac{1}{nh}\underline{Z}^{d'}\underline{Z}^{d}$ here. Note that the (i,j) term of $\frac{1}{nh}\underline{Z}^{d'}\underline{Z}^{d}$ is

$$\frac{1}{nh}\sum_{l=1}^{n} e_1' S_n^{-1}(x_l)\delta_{n,i-1}(x_l) e_1' S_n^{-1}(x_l)\delta_{n,j-1}(x_l),$$

which, by a similar argument as in the numerator, is asymptotically equivalent to

$$\frac{1}{nh}\sum_{l=1}^{n} e_1'\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}(x_l)\cdot e_1'\overline{S}^{-1}\overline{\delta}_{j-1}(x_l). \tag{24}$$

It is easy, although tedious, to show that its variance converges to zero. By Markov's inequality, (24) converges in probability to

$$\begin{aligned}
&\frac{1}{h}E\left[e_1'\overline{S}^{-1}(x_l)\overline{\delta}_{i-1}(x_l)\cdot e_1'\overline{S}^{-1}(x_l)\overline{\delta}_{j-1}(x_l)\right]\\
&\quad\approx f(\pi)\int\left[w^{i-1}1(w\geq 0) - \int_{-w}^{1} K_p^*(u)\,(w+u)^{i-1}du\right]\\
&\qquad\times\left[w^{j-1}1(w\geq 0) - \int_{-w}^{1} K_p^*(u)\,(w+u)^{j-1}du\right]dw\\
&\quad= f(\pi)\left[\int_0^1 K_p^*(\overline{\delta}_{i-1}^{+}(w))K_p^*(\overline{\delta}_{j-1}^{+}(w))dw\right.\\
&\qquad\left.+\int_{-1}^{0} K_p^*(\overline{\delta}_{i-1}^{-}(w))K_p^*(\overline{\delta}_{j-1}^{-}(w))dw\right]\\
&\quad= f(\pi)\,N_p(i,j).
\end{aligned}$$

By continuity of the matrix inversion,

$$e_1'\left(\frac{1}{nh}\underline{Z}^{d'}\underline{Z}^{d}\right)^{-1} \xrightarrow{p} f(\pi)^{-1}e_1'N_p^{-1}.$$

Based on the analysis on the numerator and denominator, the results in Theorem 1 follow.

APPENDIX B: PROOFS OF THEOREMS 2–5

Proof of Theorem 2. First,

$$\begin{aligned}\hat{\sigma}_+^2(\pi) &= e_1' S_n^+(\pi)^{-1} \tilde{r}\left(\hat{\varepsilon}_+^2(\pi)\right) \\ &= e_1' S_n^+(\pi)^{-1} \tilde{r}\left(\varepsilon_+^2(\pi)\right) + e_1' S_n^+(\pi)^{-1} \tilde{r}\left(\hat{\varepsilon}_+^2(\pi) - \varepsilon_+^2(\pi)\right),\end{aligned}$$

where $S_n^+(\pi)$ is defined in (23), $\tilde{r}\left(\hat{\varepsilon}_+^2(\pi)\right)$ is defined in (20) but now y_i is changed to $d_i \cdot \hat{\varepsilon}_i^2$, and $\tilde{r}\left(\varepsilon_+^2(\pi)\right)$ and $\tilde{r}\left(\hat{\varepsilon}_+^2(\pi) - \varepsilon_+^2(\pi)\right)$ are similarly defined with $\varepsilon_+^2(\pi) = d \cdot \varepsilon^2$. We can decompose $\hat{\sigma}_+^2(\pi)$ as the summation of two terms because the LPE is a linear functional of the dependent variable as mentioned at the end of the introduction. Second,

$$\begin{aligned}\hat{\varepsilon}_i &= y_i - X_i^{d\prime}\hat{\theta} - e_1' S_n^{-1}(x_i)\tilde{r}\left(\left(y - X^{d\prime}\hat{\theta}\right)(x_i)\right) \\ &= \varepsilon_i - X_i^{d\prime}\left(\hat{\theta} - \theta\right) - \left[e_1' S_n^{-1}(x_i)\tilde{r}\left(\left(y - X^{d\prime}\hat{\theta}\right)(x_i)\right) - m(x_i)\right],\end{aligned}$$

where $\tilde{r}((y - X^{d\prime}\hat{\theta})(x_i))$ is similarly defined as $\tilde{r}(y(x))$ in (20) but now y_j is replaced by $y_j - X_j^{d\prime}\hat{\theta}$, and we still use $m(\cdot)$ to represent $m_q(\cdot)$. To prove the consistency of $\hat{\sigma}_+^2(\pi)$, we need to show

$$e_1' S_n^+(\pi)^{-1} \tilde{r}\left(\varepsilon_+^2(\pi)\right) \xrightarrow{p} \sigma_+^2(\pi) \quad \text{and} \quad e_1' S_n^+(\pi)^{-1} \tilde{r}\left(\hat{\varepsilon}_+^2(\pi) - \varepsilon_+^2(\pi)\right) \xrightarrow{p} 0.$$

where

$$\begin{aligned}\hat{\varepsilon}_i^2 - \varepsilon_i^2 &= -2\Big[X_i^{d\prime}\left(\hat{\theta} - \theta\right) + \left(e_1' S_n^{-1}(x_i)\tilde{r}\left(\left(\tilde{y} - X^{d\prime}\left(\hat{\theta} - \theta\right)\right)(x_i)\right) - m(x_i)\right)\Big]\varepsilon_i \\ &\quad + 2\Big[X_i^{d\prime}\left(\hat{\theta} - \theta\right)\Big]\Big[e_1' S_n^{-1}(x_i)\tilde{r}\left(\left(\left(\tilde{y} - X^{d\prime}\left(\hat{\theta} - \theta\right)\right)\right)(x_i)\right) - m(x_i)\Big] \\ &\quad + \Big[X_i^{d\prime}\left(\hat{\theta} - \theta\right)\Big]^2 + \Big[e_1' S_n^{-1}(x_i)\tilde{r}\left(\left(\left(\tilde{y} - X^{d\prime}\left(\hat{\theta} - \theta\right)\right)\right)(x_i)\right) - m(x_i)\Big]^2 \\ &= -2\left\{X_i^{d\prime}\left(\hat{\theta} - \theta\right) - e_1' S_n^{-1}(x_i)\tilde{r}\left(\left(X^{d\prime}\left(\hat{\theta} - \theta\right)\right)(x_i)\right)\right. \\ &\quad \left. + \left[e_1' S_n^{-1}(x_i)\tilde{r}\left(\tilde{y}(x_i)\right) - m(x_i)\right]\right\}\varepsilon_i \\ &\quad + 2X_i^{d\prime}\left(\hat{\theta} - \theta\right)\left\{\left[e_1' S_n^{-1}(x_i)\tilde{r}\left(\tilde{y}(x_i)\right) - m(x_i)\right]\right. \\ &\quad \left. - e_1' S_n^{-1}(x_i)\tilde{r}\left(\left(X^{d\prime}\left(\hat{\theta} - \theta\right)\right)(x_i)\right)\right\} \\ &\quad + \left[X_i^{d\prime}\left(\hat{\theta} - \theta\right)\right]^2 - 2\left[e_1' S_n^{-1}(x_i)\tilde{r}\left(\left(X^{d\prime}\left(\hat{\theta} - \theta\right)\right)(x_i)\right)\right] \\ &\qquad \times \left[e_1' S_n^{-1}(x_i)\tilde{r}\left(\tilde{y}(x_i)\right) - m(x_i)\right] \\ &\quad + \left[e_1' S_n^{-1}(x_i)\tilde{r}\left(\left(X^{d\prime}\left(\hat{\theta} - \theta\right)\right)(x_i)\right)\right]^2 + \left[e_1' S_n^{-1}(x_i)\tilde{r}\left(\tilde{y}(x_i)\right) - m(x_i)\right]^2.\end{aligned}$$

So the proof is divided into the following two steps.

Step 1. $S_n^+(\pi) \xrightarrow{p} \Gamma_+ f(\pi)$, and $\tilde{r}\left(\varepsilon_+^2(\pi)\right) \xrightarrow{p} \sigma^2(\pi+)f(\pi)\mu_{0,p}^+$, where $\mu_{0,p}^+ = \left(\gamma_0^+, \gamma_1^+, \ldots, \gamma_p^+\right)'$ with $\gamma_j^+ = \int_0^1 u^j k(u)du$, $j = 0, 1, \ldots, p$. Both results can be proved by calculating the mean, showing the variance converging to zero, and applying Markov's inequality. Only note that when calculating the variance of the jth term of $\tilde{r}\left(\varepsilon_+^2(\pi)\right)$, we need $\sup_{x \in N} E\left[\varepsilon^4|x\right] < \infty$:

$$
\begin{aligned}
&\frac{1}{n}E\left[d\left(\frac{x-\pi}{h}\right)^{2j} k\left(\frac{x-\pi}{h}\right)^2 \varepsilon^4\right] \\
&\quad = \frac{1}{nh^2}\int_\pi^\infty \left(\frac{x-\pi}{h}\right)^{2j} k\left(\frac{x-\pi}{h}\right)^2 \varepsilon^4 f(\varepsilon|x)d\varepsilon f(x)dx \\
&\quad = \frac{1}{nh}\int_0^1 u^{2j}k\,(u)^2\, E\left[\varepsilon^4|\pi+uh\right] f(\pi+uh)dx = O\left(\frac{1}{nh}\right)
\end{aligned}
$$

if $\sup_{x \in N} E\left[\varepsilon^4|x\right] < \infty$. Therefore, $e_1' S_n^+(\pi)^{-1}\tilde{r}\left(\varepsilon_+^2(\pi)\right) \xrightarrow{p} e_1'(\Gamma_+ f(\pi))^{-1}\mu_{0,p}^+\ f(\pi)\sigma_+^2(\pi) = \sigma_+^2(\pi)$.

Since $S_n^+(\pi)^{-1} = O_p(1)$, we need only to show that each term of $\tilde{r}\left(\hat{\varepsilon}_+^2(\pi) - \varepsilon_+^2(\pi)\right)$ is $o_p(1)$. In other words, $\frac{1}{n}\sum_{l=1}^n \left(\frac{x_l-\pi}{h}\right)^j k_h(x_l-\pi)d_i\left(\hat{\varepsilon}_i^2 - \varepsilon_i^2\right) = o_p(1)$, $j = 0, 1, \ldots, p$.

Step 2. $\sup_{x_i \in N_0} |X_i^{d\prime}(\hat{\theta}-\theta)| = o_p(1)$, $\sup_{x_i \in N_0} |e_1' S_n^{-1}(x_i)\tilde{r}((X^{d\prime}(\hat{\theta}-\theta))\ (x_i))| = o_p(1)$, and $\sup_{x_i \in N_0} |e_1' S_n^{-1}(x_i)\tilde{r}(\tilde{y}(x_i)) - m(x_i)| = o_p(1)$.

First, $\sup_{x_i \in N_0} |X_i^{d\prime}(\hat{\theta}-\theta)| = o_p(1)$ since $\sup_{x_i \in N_0} |X_i^d| = O_p(1)$, and $\hat{\theta}-\theta = o_p(1)$.

Second, $\sup_{x_i \in N_0} S_n^{-1}(x_i) = O_p(1)$, so we need only to show that $\sup_{x_i \in N_0}\left|\frac{1}{n}\sum_{l=1}^n \left(\frac{x_l-x_i}{h}\right)^j k_h(x_l-x_i)X_l^{d\prime}\left(\hat{\theta}-\theta\right)\right| = o_p(1)$, $j = 0, 1, \ldots, p$. Notice that $\sup_{x_i \in N_0}\left|\frac{1}{n}\sum_{l=1}^n \left(\frac{x_l-x_i}{h}\right)^j k_h(x_l-x_i)X_l^d\right| = O_p(1)$ and $\hat{\theta}-\theta = o_p(1)$, so the result follows.

Third, this result is from Lemma 3 of Appendix C. Since we need $\zeta \geq 2$ in Step 1, the bandwidth is required to satisfy $\frac{\sqrt{nh^2}}{\ln n} \to \infty$.

Given these three results, we know $\frac{1}{n}\sum_{l=1}^n \left(\frac{x_l-\pi}{h}\right)^j k_h(x_l-\pi)d_i\ \left(\hat{\varepsilon}_i^2 - \varepsilon_i^2\right) = o_p(1)$, since $\frac{1}{n}\sum_{l=1}^n \left(\frac{x_l-\pi}{h}\right)^j k_h(x_l-\pi)d_i = O_p(1)$, and $\frac{1}{n}\sum_{l=1}^n \left(\frac{x_l-\pi}{h}\right)^j k_h(x_l-\pi)d_i\varepsilon_i = O_p(1)$, $j = 0, 1, \ldots, p$. Combining these two steps, we have shown $\hat{\sigma}_+^2(\pi) \xrightarrow{p} \sigma_+^2(\pi)$. Similarly, we can show $\hat{\sigma}_-^2(\pi) \xrightarrow{p} \sigma_-^2(\pi)$. □

Proof of Theorem 3. The proof is a simple application of the delta method; see Proposition 1 of Porter (2003). It is easy to show that if

$$
\sqrt{nh}\begin{pmatrix}\widehat{\Delta}-\Delta \\ \hat{\beta}-\beta\end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix}B^\Delta \\ B^\beta\end{pmatrix}, \begin{pmatrix}V_\Delta & C_{\Delta\beta} \\ C_{\Delta\beta} & V_\beta\end{pmatrix}\right),
$$

then

$$\sqrt{nh}\left(\frac{\widehat{\Delta}}{\hat{\beta}}-\frac{\Delta}{\beta}\right)\xrightarrow{d}\frac{1}{\beta}N\left(B^{\Delta}-\alpha B^{\beta},V_{\Delta}-2\alpha C_{\Delta\beta}+\alpha^{2}V_{\beta}\right).$$

Substituting the biases B^{Δ} and B^{β}, the variances V_{Δ} and V_{β}, and covariance $C_{\Delta\beta}$ in each case to the formula above, we can get the results in the theorem. B^{β} and V_{β} can be derived in a similar way as in the proof of Theorem 1. As to $C_{\Delta\beta}$, we can write out the influence function of $\hat{\beta}$, and find that

$$C_{\Delta\beta}=e_{1}'N_{p}^{-1}\left[E[R\eta\mid x=\pi+]\Omega_{p}^{+}+E[R\eta\mid x=\pi-]\Omega_{p}^{-}\right]N_{p}^{-1}e_{1}.$$

Also note that $E[\eta^{2}|x=\pi+]=(s(\pi)+\beta)(1-s(\pi)-\beta)$ and $E[\eta^{2}|x=\pi-]=s(\pi)(1-s(\pi))$ since t is a binary variable.

Proof of Theorem 4. By a similar manipulation as in (22), we can rewrite $\hat{\vartheta}$ as

$$\vartheta+\begin{pmatrix}H_{q}&0\\0&H_{p}\end{pmatrix}^{-1}\left[\frac{1}{nh}\sum_{x_{i}\in N_{0}}\begin{pmatrix}Z_{i}^{d}\\Z_{i}\end{pmatrix}\left(Z_{i}^{t\prime}\ Z_{i}^{\prime}\right)\right]^{-1}\frac{1}{nh}\sum_{x_{i}\in N_{0}}\begin{pmatrix}Z_{i}^{d}\\Z_{i}\end{pmatrix}\overleftarrow{y}_{i},$$

where Z_{i}^{t} and Z_{i}^{d} are similarly defined as $Z_{i}\equiv H_{p}^{-1}X_{i}$, $H_{q}=$diag$\{1,h,\ldots,h^{q}\}$, H_{p} is similarly defined, and

$$\begin{aligned}\overleftarrow{y}&=\left[\underline{m}(x)-\left(\phi_{0}+\phi_{1}(x-\pi)+\cdots+\phi_{p}(x-\pi)^{p}\right)\right]\\&\quad+t\left[\alpha(x)-\left(\alpha+\gamma_{1}(x-\pi)+\cdots+\gamma_{q}(x-\pi)^{q}\right)\right]+\varepsilon_{t}\\&\equiv\overleftarrow{m}(x)+t\overleftarrow{\alpha}(x)+\varepsilon_{t}=\overleftarrow{m}(x)+\overleftarrow{t}(x)\overleftarrow{\alpha}(x)+\varepsilon_{t}+\eta\overleftarrow{\alpha}(x)\end{aligned}$$

with $\overleftarrow{t}(x)=s(x)+d\beta$. So the asymptotic distribution is determined by

$$\sqrt{nh}\begin{pmatrix}H_{q}&0\\0&H_{p}\end{pmatrix}\left(\hat{\vartheta}-\vartheta\right)=\left[\frac{1}{nh}\sum_{x_{i}\in N_{0}}\begin{pmatrix}Z_{i}^{d}\\Z_{i}\end{pmatrix}\left(Z_{i}^{t\prime}\ Z_{i}^{\prime}\right)\right]^{-1}\frac{1}{\sqrt{nh}}\sum_{x_{i}\in N_{0}}\begin{pmatrix}Z_{i}^{d}\\Z_{i}\end{pmatrix}\overleftarrow{y}_{i}.$$

Numerator

First analyze the bias. Taking the jth term of $\frac{1}{\sqrt{nh}}\sum_{x_{i}\in N_{0}}Z_{i}^{d}\overleftarrow{y}_{i}$, $1\leq j\leq q+1$, we have

$$\begin{aligned}&E\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}1\left(\pi\leq x_{i}\leq\pi+h\right)\left(\frac{x_{i}-\pi}{h}\right)^{j-1}\overleftarrow{y}_{i}\right]\\&=\sqrt{\frac{n}{h}}E\left[1\left(\pi\leq x\leq\pi+h\right)\left(\frac{x-\pi}{h}\right)^{j-1}\left(\overleftarrow{m}(x)+\overleftarrow{t}(x)\overleftarrow{\alpha}(x)\right)\right]\end{aligned}$$

$$
\begin{aligned}
&= \sqrt{\frac{n}{h}} \int_{\pi}^{\pi+h} \left(\frac{x-\pi}{h}\right)^{j-1} \left(\overleftarrow{m}(x) + \overleftarrow{t}(x)\overleftarrow{\alpha}(x)\right) f(x)dx \\
&= \sqrt{nh} \int_{0}^{1} u^{j-1} \left(\overleftarrow{m}(\pi+uh) + \overleftarrow{t}(\pi+uh)\overleftarrow{\alpha}(\pi+uh)\right) f(\pi+uh)du \\
&\approx \sqrt{nh} \int_{0}^{1} u^{j-1} \left[\frac{\underline{m}^{(p+1)}(\pi)}{(p+1)!} u^{p+1}h^{p+1} + (s(\pi)+\beta)\frac{\alpha^{(q+1)}(\pi)}{(q+1)!} u^{q+1}h^{q+1}\right] f(\pi)du \\
&= f(\pi)\left[\sqrt{nh}h^{p+1}\gamma_{j+p}^{+}\frac{\underline{m}^{(p+1)}(\pi)}{(p+1)!} + \sqrt{nh}h^{q+1}\gamma_{j+q}^{+}(s(\pi)+\beta)\frac{\alpha^{(q+1)}(\pi)}{(q+1)!}\right],
\end{aligned}
$$

where the first equality is from

$$
E\left[\sqrt{\frac{n}{h}} 1(\pi \le x \le \pi+h)\left(\frac{x-\pi}{h}\right)^{j-1} \left[\left(\varepsilon_0 + (\overleftarrow{t}+\eta)(\varepsilon_1-\varepsilon_0)\right) + \eta\overleftarrow{\alpha}(x)\right]\right] = 0.
$$

Similarly, taking the jth term of $\frac{1}{\sqrt{nh}}\sum_{x_i\in N_0} Z_i \overleftarrow{y}_i$, $1 \le j \le p+1$, we have

$$
\begin{aligned}
&E\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} 1(\pi-h \le x_i \le \pi+h)\left(\frac{x_i-\pi}{h}\right)_i^{j-1} \overleftarrow{y}_i\right] \\
&= f(\pi)\left[\sqrt{nh}h^{p+1}\gamma_{j+p}\frac{\underline{m}^{(p+1)}(\pi)}{(p+1)!} + \sqrt{nh}h^{q+1}\left(\gamma_{j+q}s(\pi)+\gamma_{j+q}^{+}\beta\right)\frac{\alpha^{(q+1)}(\pi)}{(q+1)!}\right],
\end{aligned}
$$

where $\gamma_j = \int_{-1}^{1} u^j du$ is the γ_j defined at the end of the introduction with the uniform kernel. So the mean of the numerator converges to

$$
\begin{aligned}
&\sqrt{nh}h^{p+1}f(\pi)\frac{\underline{m}^{(p+1)}(\pi)}{(p+1)!}\begin{pmatrix}\mu_{p+1,p+q+1}^{+} \\ \mu_{p+1,2p+1}\end{pmatrix} \\
&+ \sqrt{nh}h^{q+1}f(\pi)\frac{\alpha^{(q+1)}(\pi)}{(q+1)!}\begin{pmatrix}(s(\pi)+\beta)\mu_{q+1,2q+1}^{+} \\ s(\pi)\mu_{q+1,p+q+1} + \beta\mu_{q+1,p+q+1}^{+}\end{pmatrix}.
\end{aligned}
$$

Under Assumption A(b), the second term of the bias is of lower order relative to the first term, so can be neglected.

Second analyze the variance of $\frac{1}{\sqrt{nh}}\sum_{x_i\in N_0}\binom{Z_i^d}{Z_i}[\varepsilon_{ti}+\eta_i\overleftarrow{\alpha}(x_i)]$. The covariance between the jth term associated with Z_i^d and lth term associated with Z_i is

$$
\frac{1}{nh}E\left[\sum_{i=1}^{n} 1(\pi \le x_i \le \pi+h)\left(\frac{x_i-\pi}{h}\right)^{j-1}\left(\frac{x_i-\pi}{h}\right)^{l-1}[\varepsilon_{ti}+\eta_i\overleftarrow{\alpha}(x_i)]^2\right]
$$

$$= \frac{1}{h}\int_{\pi}^{\pi+h}\left(\frac{x-\pi}{h}\right)^{j+l-2} E\left[\left[\varepsilon_t + \eta \overleftarrow{\alpha}(x)\right]^2 \middle| x\right] f(x)dx$$

$$\approx \int_0^1 u^{j+l-2} E\left[\varepsilon_t^2 \middle| x = \pi + uh\right] f(\pi + uh)du \to E\left[\varepsilon_t^2 \middle| x = \pi+\right] f(\pi)\gamma_{j+l-2}^{+},$$

where the second equality is from $\overleftarrow{\alpha}(x) = O(h)$ for $x \in N_0$. Similarly, the covariance between the jth term associated with Z_i^d and lth term associated with Z_i^d converges to $E\left[\varepsilon_t^2 \middle| x = \pi+\right] f(\pi)\gamma_{j+l-2}^{+}$, and the covariance between the jth term associated with Z_i and lth term associated with Z_i converges to $E\left[\varepsilon_t^2 \middle| x = \pi+\right] f(\pi)\gamma_{j+l-2}^{+} + E\left[\varepsilon_t^2 \middle| x = \pi-\right] f(\pi)\gamma_{j+l-2}^{-}$. In summary, the asymptotic variance of the numerator is

$$f(\pi)\begin{pmatrix} E\left[\varepsilon_t^2 \middle| x = \pi+\right]\Gamma_{+}^{qq} & E\left[\varepsilon_t^2 \middle| x = \pi+\right]\Gamma_{+}^{qp} \\ E\left[\varepsilon_t^2 \middle| x = \pi+\right]\Gamma_{+}^{pq} & E\left[\varepsilon_t^2 \middle| x = \pi+\right]\Gamma_{+}^{pp} + E\left[\varepsilon_t^2 \middle| x = \pi-\right]\Gamma_{-}^{pp} \end{pmatrix}.$$

Denominator

First calculate the probability limit of $\frac{1}{nh}\sum_{x_i \in N_0} Z_i^d Z_i^{t\prime}$. Its (j, l) term converges to

$$E\left[\frac{1}{nh}\sum_{i=1}^{n} 1\left(\pi \le x_i \le \pi + h\right)\left(\frac{x_i-\pi}{h}\right)^{j-1} t_i \left(\frac{x_i-\pi}{h}\right)^{l-1}\right]$$

$$= \frac{1}{h}\int_{\pi}^{\pi+h}\left(\frac{x-\pi}{h}\right)^{j+l-2} \overleftarrow{t}(x)f(x)dx$$

$$= \int_0^1 u^{j+l-2}\overleftarrow{t}(\pi+uh)f(\pi+uh)du$$

$$\approx \left(s(\pi)+\beta\right) f(\pi)\gamma_{j+l-2}^{+},$$

so

$$\frac{1}{nh}\sum_{x_i \in N_0} Z_i^d Z_i^{t\prime} \xrightarrow{p} \left(s(\pi)+\beta\right) f(\pi)\Gamma_{+}^{qq}.$$

Similarly,

$$\frac{1}{nh}\sum_{x_i \in N_0} Z_i^d Z_i' \xrightarrow{p} f(\pi)\Gamma_{+}^{qp},$$

$$\frac{1}{nh}\sum_{x_i \in N_0} Z_i Z_i^{t\prime} \xrightarrow{p} f(\pi)\left[s(\pi)\Gamma^{pq} + \beta\Gamma_{+}^{pq}\right],$$

$$\frac{1}{nh}\sum_{x_i \in N_0} Z_i Z_i' \xrightarrow{p} f(\pi)\Gamma^{pp}.$$

So the denominator converges to

$$f(\pi)\begin{pmatrix}(s(\pi)+\beta)\,\Gamma_+^{qq} & \Gamma_+^{qp}\\ s(\pi)\Gamma^{pq}+\beta\Gamma_+^{pq} & \Gamma^{pp}\end{pmatrix}.$$

Combing the analysis above, the theorem is proved. □

Proof of Theorem 5. Note that

$$nhe_1'\widehat{\Sigma}e_1 = e_1'\left[\frac{1}{nh}\sum_{x_i\in N_0}\begin{pmatrix}Z_i^d\\ Z_i\end{pmatrix}\left(Z_i^{t\prime}\ Z_i'\right)\right]^{-1}\left[\frac{1}{nh}\sum_{x_i\in N_0}\begin{pmatrix}Z_i^d\\ Z_i\end{pmatrix}\left(Z_i^{d\prime}\ Z_i'\right)\hat{\varepsilon}_{ti}^2\right]$$
$$\times\left[\frac{1}{nh}\sum_{x_i\in N_0}\begin{pmatrix}Z_i^t\\ Z_i\end{pmatrix}\left(Z_i^{d\prime}\ Z_i'\right)\right]^{-1}e_1.$$

From the proof of Theorem 4, the first and third terms converge to the targets we want, so we need only to show that

$$\frac{1}{nh}\sum_{x_i\in N_0}\begin{pmatrix}Z_i^d\\ Z_i\end{pmatrix}\left(Z_i^{d\prime}\ Z_i'\right)\hat{\varepsilon}_{ti}^2$$
$$\xrightarrow{p} f(\pi)\begin{pmatrix}E\left[\varepsilon_t^2\middle|\,x=\pi+\right]\Gamma_+^{qq} & E\left[\varepsilon_t^2\middle|\,x=\pi+\right]\Gamma_+^{qp}\\ E\left[\varepsilon_t^2\middle|\,x=\pi+\right]\Gamma_+^{pq} & E\left[\varepsilon_t^2\middle|\,x=\pi+\right]\Gamma_+^{pp}+E\left[\varepsilon_t^2\middle|\,x=\pi-\right]\Gamma_-^{pp}\end{pmatrix}. \tag{25}$$

Given that

$$\hat{\varepsilon}_{ti} = y_i-\left(X_i^{t\prime}\ X_i'\right)\hat{\vartheta} = y_i-\left(X_i^{t\prime}\ X_i'\right)\vartheta-\left(X_i^{t\prime}\ X_i'\right)\left(\hat{\vartheta}-\vartheta\right)$$
$$= \overleftarrow{m}(x_i)+t_i\overleftarrow{\alpha}(x_i)+\varepsilon_{ti}-\left(X_i^{t\prime}\ X_i'\right)\left(\hat{\vartheta}-\vartheta\right),$$

so

$$\hat{\varepsilon}_{ti}^2-\varepsilon_{ti}^2 = [\overleftarrow{m}(x_i)+t_i\overleftarrow{\alpha}(x_i)]^2+\left[\left(X_i^{t\prime}X_i'\right)\left(\hat{\vartheta}-\vartheta\right)\right]^2+2[\overleftarrow{m}(x_i)+t_i\overleftarrow{\alpha}(x_i)]\,\varepsilon_{ti}$$
$$-2[\overleftarrow{m}(x_i)+t_i\overleftarrow{\alpha}(x_i)]\left[\left(X_i^{t\prime}X_i'\right)\left(\hat{\vartheta}-\vartheta\right)\right]-2\left[\left(X_i^{t\prime}\ X_i'\right)\left(\hat{\vartheta}-\vartheta\right)\right]\varepsilon_{ti}.$$

Because $\sup_{x_i\in N_0}|\overleftarrow{m}(x_i)+t_i\overleftarrow{\alpha}(x_i)| = o_p(1)$, $\sup_{x_i\in N_0}|(X_i^{t\prime}\ \ X_i')| = O_p(1)$, $\frac{1}{nh}\sum_{x_i\in N_0}\begin{pmatrix}Z_i^d\\ Z_i\end{pmatrix}$ $\left(Z_i^{d\prime}\ Z_i'\right)\varepsilon_{ti} = O_p(1)$, and $\hat{\vartheta}-\vartheta = o_p(1)$,

$$\frac{1}{nh}\sum_{x_i\in N_0}\begin{pmatrix}Z_i^d\\ Z_i\end{pmatrix}\left(Z_i^{d\prime}\ Z_i'\right)\left(\hat{\varepsilon}_{ti}^2-\varepsilon_{ti}^2\right)\xrightarrow{p} 0.$$

As long as we show that $\frac{1}{nh}\sum_{x_i\in N_0}\binom{Z_i^d}{Z_i}\left(Z_i^{d\prime}\ Z_i'\right)\varepsilon_{ti}^2$ converges in probability to the right hand side of (25), the proof is completed. From the proof of Theorem 4, its mean matches the target. Also by a similar proof as in Step 1 of the proof of Theorem 2, we can show its variance shrinks to zero. So by Markov's inequality, the result follows.

APPENDIX C: LEMMAS

Lemma 1. $\underline{X}_l^d(\pi) = 0$ *for* $|x_l - \pi| > h,\ l = 1,\ldots,n.$

Proof. From (2.4) of Fan et al. (1997),

$$\sum_{j=1}^{n}(x_j - x)^v W_j^n(x) = \delta_{0,v}, \qquad 0 \le v \le p,$$

where $W_j^n(x)$ is defined at the beginning of the proof of Theorem 1, and $\delta_{0,v}$ equals 1 if $v = 0$, and equals 0 otherwise. Based on this result, for any x_l such that $|x_l - \pi| > h$,

$$(x_l - x)^{i-1}\,1\,(x_l > \pi) - \mathbb{P}_{x_l}^n\left(X^{i-1,d}(\pi)\right) = 0,\ 1 \le i \le q+1 \le p+1.$$

For example, if $x - \pi > h$, for $i = 1$,

$$(x - \pi)^{i-1}\,1\,(x > \pi) - \mathbb{P}_x^n\left(X^{i-1,d}(\pi)\right) = 1 - \sum_{j=1}^{n} W_j^n(x) = 0.$$

Note that the indicator function $1\left(x_j > \pi\right)$ in $X^{i-1,d}(\pi)$ does not play any role here. For $i = 2$,

$$\begin{aligned}(x - \pi) - \sum_{j=1}^{n} W_j^n(x)\left(x_j - \pi\right) &= (x - \pi) - \sum_{j=1}^{n} W_j^n(x)\left(x_j - x + x - \pi\right)\\ &= (x - \pi) - (x - \pi)\sum_{j=1}^{n} W_j^n(x) = 0.\end{aligned}$$

By induction, we can show all other terms are zero as long as $q \le p$.

Lemma 2. *Suppose* $m(x) = E\left[y_i|x_i = x\right]$ *is* q *times continuously differentiable with* $q \le p$ *for* $x \in N$*. Then*

$$\mathbb{P}_x^n(y) - m(x) = e_1' S^{-1}(x)\,\overline{r}(x) + \mathbb{P}_x^L(y) + \mathbb{P}_x^Q(y),$$

where

$$\overline{r}(x) = \int \delta(u) f(x+uh) \left(m(x+uh) - m(x) - \sum_{\ell=1}^{q} \frac{m^{(\ell)}(x)}{\ell!} (uh)^{\ell} \right) du,$$

and $\mathbb{P}_x^L(y)$ *and* $\mathbb{P}_x^Q(y)$ *are defined in (26). If* $q > p$, *then the* q *in* $\overline{r}(x)$ *is changed to* p, *and* $\mathbb{P}_x^L(y)$ *and* $\mathbb{P}_x^Q(y)$ *are adjusted correspondingly.*

Proof. Define $y_i = m(x_i) + \varepsilon_i$. Then

$$\begin{aligned}
\mathbb{P}_x^n(y) - m(x) &= e_1' \left(Z(x)' K_h(x) Z(x) \right)^{-1} Z(x)' K_h(x) (\mathbf{y} - m(x)), \\
&= e_1' \left(\frac{1}{n} \sum_{j=1}^{n} Z_j(x) Z_j'(x) k_h(x_j - x) \right)^{-1} \\
&\quad \times \frac{1}{nh} \sum_{j=1}^{n} Z_j(x) k_h(x_j - x) \left(m(x_j) - m(x) + \varepsilon_j \right) \\
&= e_1' \left(\frac{1}{n} \sum_{j=1}^{n} Z_j(x) Z_j'(x) k_h(x_j - x) \right)^{-1} \frac{1}{n} \sum_{j=1}^{n} Z_j(x) k_h(x_j - x) \\
&\quad \times \left\{ m(x_j) - m(x) - \sum_{\ell=1}^{q} \frac{m^{(\ell)}(x)}{\ell!} (x_j - x)^{\ell} + \varepsilon_j \right\} \\
&\equiv e_1' S_n^{-1}(x) \tilde{r}(x).
\end{aligned}$$

Linearize the denominator at its limit $S(x)$ and the numerator at its mean $\overline{r}(x)$. Note that $\overline{r}(x)$ converges to 0 when h goes to zero, so we cannot linearize at the limit of the numerator:

$$\begin{aligned}
& e_1' S_n^{-1}(x) \tilde{r}(x) - e_1' S^{-1}(x) \overline{r}(x) \\
&= -e_1' S^{-1}(x) (S_n(x) - S(x)) S^{-1}(x) \overline{r}(x) \\
&\quad + e_1' S^{-1}(x) (\tilde{r}(x) - \overline{r}(x)) \text{ (linear terms)} \\
&\quad + e_1' S^{-1}(x) (S_n(x) - S(x)) S^{-1}(x) (S_n(x) - S(x)) S_n^{-1}(x) \overline{r}(x) \\
&\quad - e_1' S^{-1}(x) (S_n(x) - S(x)) S_n^{-1}(x) (\tilde{r}(x) - \overline{r}(x)) \text{ (quadratic terms)} \\
&\equiv \mathbb{P}_x^L(y) + \mathbb{P}_x^Q(y).
\end{aligned} \tag{26}$$

Lemma 3. *If* $\sup_{x \in N} E\left[|\varepsilon|^{2+\zeta} \,|x \right] < \infty$ *for some* $\zeta > 0$, $n^{\zeta/(2+\zeta)} h / \ln n \to \infty$, $l_m \geq q+1$, *and* $l_f \geq 0$, *then for* $N_0 = [\pi - h, \pi + h]$, *the following statement holds:*

(i) $\sup_{x\in N_0}\left|\tilde{y}(x)-m(x)\right| = O_p\left(\sqrt{\frac{\ln n}{nh}}+h^{q+1}\right)$, *and* $\sup_{x\in N_0}\left|\tilde{r}(\varepsilon(x))\right| = O_p\left(\sqrt{\frac{\ln n}{nh}}\right)$.

If $\frac{nh}{\ln n}\to\infty$, $l_m \geq q+1$, *and* $l_f \geq 1$, *then the following statements hold:*

(ii) $\sup_{x\in N_0}\left|S_n(x)-\overline{S}(x)\right| = O_p\left(\sqrt{\frac{\ln n}{nh}}\right)$, $\sup_{x\in N_0}\left|S_n^{-1}(x)-\overline{S}^{-1}(x)\right| = O_p\left(\sqrt{\frac{\ln n}{nh}}\right)$, $\sup_{x\in N_0}$ $\left|\overline{S}(x)-S(x)\right| = O(h)$;

(iii) $\sup_{x\in N_0}\left|\tilde{r}(\Delta m(x))-\overline{r}(\Delta m(x))\right| = O_p\left(\sqrt{\frac{\ln n}{nh}}\right)$;

(iv) $\sup_{x\in N_0}\left|\delta_{n,i-1}(x)-\overline{\delta}_{i-1}(x)\right| = O_p\left(\sqrt{\frac{\ln n}{nh}}\right)$.

Here, the norm $|\cdot|$ *for a vector or matrix is the maximum absolute value among all elements.*

Proof. The proof follows from Lemma B.1 and B.2 of Newey (1994). The basic proof techniques are truncation and Bernstein's inequality. For (ii), (iii), and (iv), we do not need truncation, which is like $p=\infty$th moment of the dependent variable in Lemma B.1 is finite, so the bandwidth is only required to satisfy $\frac{nh}{\ln n}\to\infty$. Since the proof is very standard, omitted here for simplicity. See also Masry (1996) for more details. We only discuss a little about $\sup_{x\in N_0}\left|S_n^{-1}(x)-\overline{S}^{-1}(x)\right|$ and $\sup_{x\in N_0}\left|\overline{S}(x)-S(x)\right|$. First, note that

$$\begin{aligned}\sup_{x\in N_0}\left|S_n^{-1}(x)-\overline{S}^{-1}(x)\right| &\leq \sup_{x\in N_0}\left|\overline{S}^{-1}(x)\right|\sup_{x\in N_0}\left|S_n(x)-\overline{S}(x)\right|\sup_{x\in N_0}\left|S_n^{-1}(x)\right|\\ &= O(1)O_p\left(\sqrt{\frac{\ln n}{nh}}\right)O_p(1) = O_p\left(\sqrt{\frac{\ln n}{nh}}\right).\end{aligned}$$

Second, we know $S_n(x)$ plays the role of a density estimator in the NWE, but there is indeed some difference between S_n and the usual density estimator $\hat{f}(x)$; the bias of $\hat{f}(x)$ can be made to be higher order of h by using a higher order kernel, while the bias of S_n is only $O(h)$ since usually only a second order kernel is used in the LPE.

Lemma 4. *If* $\sup_{x\in N} E[.|\varepsilon|^2|x] < \infty$, *then* $\frac{1}{nh}\sum_{l=1}^{n}|\varepsilon_l|1(\pi-h\leq x_l\leq\pi+h) = O_p(1)$ *and* $\frac{1}{nh}\sum_{l=1}^{n}1(\pi-h\leq x_l\leq\pi+h) = O_p(1)$.

Proof. These are intermediate results in Porter (2003), and can be proved by Markov's inequality.

ACKNOWLEDGEMENT

I want to thank the seminar participants at the 51st NZAE Annual Conference and two referees for helpful comments. Special thanks go to Jack Porter for insightful discussions.

REFERENCES

Angrist, J. D., Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* 114:533–575.

Battistin, E., Rettore, E. (2002). Testing for programme effects in a regression discontinuity design with imperfect compliance. *Journal of the Royal Statistical Society, Series A* 165:39–57.

Black, S. (1999). Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics* 114:577–599.

Card, D., Mas, A., Rothstein, J. (2008). Tipping and the dynamics of segregation in neighborhoods and schools. *Quarterly Journal of Economics* 123:177–218.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics* 21:520–533.

Chay, K., Greenstone, M. (2005). Does air quality matter? Evidence from the housing market. *Journal of Political Economy* 113:376–424.

Chay, K., McEwan, P., Urquiola, M. (2005). The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review* 95:1237–1258.

Cook, T. D. (2008). Waiting for life to arrive: A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics* 142:636–654.

Dell, M. (2010). The persistent effects of peru's mining mita. *Econometrica* 78:1863–1903.

DesJardins, S. L., McCall, B. P. (2008). The Impact of the Gates Millennium Scholars Program on the Retention, College Finance- and Work-Related Choices, and Future Educational Aspirations of Low-Income Minority Students, unpublished manuscript, Department of Economics, University of Michigan.

DiNardo, J., Lee, D. S. (2004). Economic impacts of new unionization on private sector employers: 1984–2001. *Quarterly Journal of Economics* 119:1383–1441.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87:998–1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiency. *Annals of Statistics* 21: 196–216.

Fan, J., Gasser, T., Gijbels, I. (1997). Local polynomial regression: optimal kernels and asymptotic minimax efficiency. *Annals of the Institute of Statistical Mathematics* 49:79–99.

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

Gasser, T., Müller, H.-G., Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society: Series B* 47:238–252.

Hahn, J., Todd, P., Van de Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69:201–209.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* 575–603.

Härdle, W. (1990). *Applied Nonparametric Regression*. New York: Cambridge University Press.

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

Jacob, B. A., Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics* 86:226–244.

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. house elections. *Journal of Econometrics* 142:675–697.

Lee, D. S., Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature* 48:281–355.

Li, Q., Racine, J. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton, N.J.: Princeton University Press.

Ludwig, J., Miller, D. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity disign. *Quarterly Journal of Economics* 122:159–208.

Mammen, E., Marron, J. S., Turlach, B. A., Wand, M. P. (2001). A general projection framework for constrained smoothing. *Statistical Science* 16:232–248.

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17:571–599.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142:698–714.

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory* 10:233–253.

Newey, W. K., McFadden, D. L. (1994). Large sample estimation and hypothesis testing. In: Eagle, R. F., McFadden, D. L., eds. *Handbook of Econometrics*. Vol. 4, Elsevier Science B.V., Ch. 36, 2113–2245.

Pagan, A., Ullah, A. (1999). *Nonparametric Econometrics*. New York: Cambridge University Press.

Pence, K. (2006). Foreclosing on opportunity: State laws and mortgage credit. *The Review of Economics and Statistics* 88:177–182.

Porter, J. (2003). Estimation in the Regression Discontinuity Model, Mimeo, Department of Economics, University of Wisconsin at Madison.

Porter, J., Yu, P. (2010). Regression Discontinuity Designs with Unknown Discontinuity Points: Testing and Estimation. Mimeo, Department of Economics, University of Wisconsin at Madison.

Robinson, P. (1988). Root-N-consistent semiparametric regression. *Econometrica* 56:931–954.

Ruppert, D., Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics* 22:1346–1370.

Thistlewaite, D., Campbell, D. (1960). Regression-discontinuity Analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology* 51:309–317.

Trochim, W. (1984). *Research Design for Program Evaluation: The Regression Discontinuity Approach*. Beverly Hills: Sage Publications.

Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review* 43:1249–1287.

Van der Klaauw, W. (2008). Regression–discontinuity analysis: A survey of recent developments in economics. *Labour* 22:219–245.

Yu, P. (n.d.). Adaptive Estimation of the Threshold Point in Threshold Regression. *Journal of Econometrics*, forthcoming.

Yu, P. (2012). Likelihood estimation and inference in threshold regression. *Journal of Econometrics* 167: 274–294.