Optimal Bandwidth Selection for the Fuzzy Regression Discontinuity Estimator

Yoichi Arai
Hidehiko Ichimura

September 2015

GRIPS

NATIONAL GRADUATE INSTITUTE
FOR POLICY STUDIES

Optimal Bandwidth Selection for the Fuzzy Regression Discontinuity Estimator

Yoichi Arai[*a] and Hidehiko Ichimura[b]

[a]National Graduate Institute for Policy Studies (GRIPS), 7-22-1 Roppongi, Minato-ku, Tokyo 106-8677, Japan; yarai@grips.ac.jp

[b]Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan; ichimura@e.u-tokyo.ac.jp

Abstract

A new bandwidth selection method for the fuzzy regression discontinuity estimator is proposed. The method chooses two bandwidths simultaneously, one for each side of the cut-off point by using a criterion based on the estimated asymptotic mean square error taking into account a second-order bias term. A simulation study demonstrates the usefulness of the proposed method.

Key words: Bandwidth selection, fuzzy regression discontinuity design, local linear regression

JEL Classification: C13, C14

---

[*]Corresponding Author.

1 Introduction

The fuzzy regression discontinuity (FRD) estimator, developed by Hahn, Todd, and Van der Klaauw (2001) (hereafter HTV), has found numerous empirical applications in economics. The target parameter in the FRD design is the ratio of the difference of two conditional mean functions, which is interpreted as the local average treatment effect. The most frequently used estimation method is the nonparametric method using the local linear regression (LLR). Imbens and Kalyanaraman (2012) (hereafter IK) propose a bandwidth selection method specifically aimed at the FRD estimator, which uses a single bandwidth to estimate all conditional mean functions.

This paper proposes to choose two bandwidths simultaneously, one for each side of the cut-off point. In the context of the sharp RD (SRD) design, Arai and Ichimura (2015) (hereafter AI) show that the simultaneous selection method is theoretically superior to the existing methods and their extensive simulation experiments verify the theoretical predictions. We extend their approach to the FRD estimator. A simulation study illustrates the potential usefulness of the proposed method.[1]

2 Bandwidth Selection of The FRD Estimator

For individual i potential outcomes with and without treatment are denoted by $Y_i(1)$ and $Y_i(0)$, respectively. Let D_i be a binary variable that stands for the treatment status, 0 or 1. Then the observed outcome, Y_i, is described as $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$. Throughout the paper, we assume that (Y_1, D_1, X_1), ..., (Y_n, D_n, X_n) are i.i.d. observations and X_i has the Lebesgue density f.

To define the parameter of interest for the FRD design, denote $m_{Y+}(x) = E(Y_i | X_i = x)$ and $m_{D+}(x) = E(D_i | X_i = x)$ for $x \geq c$. Suppose that $\lim_{x \searrow c} m_{Y+}(x)$ and $\lim_{x \searrow c} m_{D+}(x)$ exist and they are denoted by $m_{Y+}(c)$ and $m_{D+}(c)$, respectively. We define $m_{Y-}(c)$ and $m_{D-}(c)$ similarly. The conditional variances and covariance, $\sigma_{Yj}^2(c) > 0$, $\sigma_{Dj}^2(c) > 0$, $\sigma_{YDj}(c)$, and the second and third derivatives $m_{Yj}^{(2)}(c)$, $m_{Yj}^{(3)}(c)$, $m_{Dj}^{(2)}(c)$, $m_{Dj}^{(3)}(c)$, for

---

$j = +, -$, are defined in the same manner. We assume all the limits exist and are bounded above.

In the FRD design, the treatment status depends on the assignment variable X_i in a stochastic manner and the propensity score function is known to have a discontinuity at the cut-off point c, implying $m_{D+}(c) \neq m_{D-}(c)$. Under the conditions of HTV, Porter (2003) or Dong and Lewbel (forthcoming), the LATE at the cut-off point is given by $\tau(c) = (m_{Y+}(c) - m_{Y-}(c))/(m_{D+}(c) - m_{D-}(c))$. This implies that estimation of $\tau(c)$ reduces to estimating the four conditional mean functions nonparametrically and the most popular method is the LLR because of its automatic boundary adaptive property (Fan, 1992).

Estimating the four conditional expectations, in principle, requires four bandwidths. IK simplifies the choice by using a single bandwidth to estimate all functions as they do for the SRD design. For the SRD design, AI proposes to choose bandwidths, one for each side of the cut-off point because the curvatures of the conditional mean functions and the sample sizes on the left and the right of the cut-off point may differ significantly. We use the same idea here, but take into account the bias and variance due to estimation of the denominator as well. For simplification, we propose to choose one bandwidth, h_+, to estimate $m_{Y+}(c)$ and $m_{D+}(c)$ and another bandwidth, h_-, to estimate $m_{Y-}(c)$ and $m_{D-}(c)$ because it is also reasonable to use the same group on each side.

2.1 Optimal Bandwidths Selection for the FRD Estimator

We consider the estimator of $\tau(c)$, denoted $\hat{\tau}(c)$, based on the LLR estimators of the four unknown conditional mean functions. We propose to choose two bandwidths simultaneously based on an asymptotic approximation of the mean squared error (AMSE). To obtain the AMSE, we assume the following:

ASSUMPTION 1 *(i) (Kernel) $K(\cdot) : \mathbb{R} \to \mathbb{R}$ is a symmetric second-order kernel function that is continuous with compact support; (ii) (Bandwidth) The positive sequence of bandwidths is such that $h_j \to 0$ and $nh_j \to \infty$ as $n \to \infty$ for $j = +, -$.*

Let $\mathcal{D}$ be an open set in $\mathbb{R}$, k be a nonnegative integer, $\mathcal{C}_k$ be the family of k times continuously differentiable functions on $\mathcal{D}$ and $g^{(k)}(\cdot)$ be the kth derivative of $g(\cdot) \in \mathcal{C}_k$. Let $\mathcal{G}_k(\mathcal{D})$ be the collection of functions g such that $g \in \mathcal{C}_k$ and $\left|g^{(k)}(x) - g^{(k)}(y)\right| \leq M_k |x - y|^\alpha$, $x, y, z \in \mathcal{D}$, for some positive M_k and some α such that $0 < \alpha \leq 1$.

ASSUMPTION 2 *The density of X, f, which is bounded above and strictly positive at c, is an element of $\mathcal{G}_1(\mathcal{D})$ where $\mathcal{D}$ is an open neighborhood of c.*

ASSUMPTION 3 *Let δ be some positive constant. The m_{Y+}, σ^2_{Y+} and σ_{YD+} are elements of $\mathcal{G}_3(\mathcal{D}_1)$, $\mathcal{G}_0(\mathcal{D}_1)$ and $\mathcal{G}_0(\mathcal{D}_1)$, respectively, where $\mathcal{D}_1$ is a one-sided open neighborhood of c, $(c, c+\delta)$. Analogous conditions hold for m_{Y-}, σ^2_{Y-} and σ_{YD-} on $\mathcal{D}_0$ where $\mathcal{D}_0$ is a one-sided open neighborhood of c, $(c-\delta, c)$.*

The following approximation holds for the MSE under the conditions stated above.

LEMMA 1 *Suppose Assumptions 1–3 hold. Then, it follows that*

$$MSE_n(h_+, h_-) = \frac{1}{(\tau_D(c))^2} \left\{ \left[\phi_1(c)h_+^2 - \phi_0(c)h_-^2\right] + \left[\psi_1(c)h_+^3 - \psi_0(c)h_-^3\right] + o\left(h_+^3 + h_-^3\right) \right\}^2$$
$$+ \frac{v}{nf(c)(\tau_D(c))^2} \left\{ \frac{\omega_1(c)}{h_+} + \frac{\omega_0(c)}{h_-} \right\} + o\left(\frac{1}{nh_+} + \frac{1}{nh_-} \right) \quad (1)$$

where, for $j = +, -$ and $k = Y, D$, $\tau_D(c) = m_{D+}(c) - m_{D-}(c)$, $\omega_j(c) = \sigma^2_{Yj}(c) + \tau(c)^2 \sigma^2_{Dj}(c) - 2\tau(c)\sigma_{YDj}(c)$, $\phi_j(c) = C_1 \left[m^{(2)}_{Yj}(c) - \tau(c)m^{(2)}_{Dj}(c) \right]$, $\psi_j(c) = \zeta_{Yj}(c) - \tau(c)\zeta_{Dj}(c)$,

$$\zeta_{kj}(c) = (-j) \left\{ \xi_1 \left[\frac{m^{(2)}_{kj}(c)}{2} \frac{f^{(1)}(c)}{f(c)} + \frac{m^{(3)}_{kj}(c)}{6} \right] - \xi_2 \frac{m^{(2)}_{kj}(c)}{2} \frac{f^{(1)}(c)}{f(c)} \right\},$$

$C_1 = (\mu_2^2 - \mu_1\mu_3)/2(\mu_0\mu_2 - \mu_1^2)$, $v = (\mu_2^2\nu_0 - 2\mu_1\mu_2\nu_1 + \mu_1^2\nu_2)/(\mu_0\mu_2 - \mu_1^2)^2$, $\xi_1 = (\mu_2\mu_3 - \mu_1\mu_4)/(\mu_0\mu_2 - \mu_1^2)$, $\xi_2 = (\mu_2^2 - \mu_1\mu_3)(\mu_0\mu_3 - \mu_1\mu_2)/(\mu_0\mu_2 - \mu_1^2)^2$, $\mu_j = \int_0^\infty u^j K(u)du$, $\nu_j = \int_0^\infty u^j K^2(u)du$.

A standard approach applied in this context is to minimize the following AMSE, ignoring higher order terms:

$$AMSE(h_+, h_-) = \frac{1}{(\tau_D(c))^2} \left\{ \phi_+(c)h_+^2 - \phi_-(c)h_-^2 \right\}^2 + \frac{v}{nf(c)(\tau_D(c))^2} \left\{ \frac{\omega_1(c)}{h_+} + \frac{\omega_0(c)}{h_-} \right\}.$$
$$(2)$$

4

As AI observed, (i) while the optimal bandwidths that minimize the AMSE (2) are well-defined when $\phi_+(c) \cdot \phi_-(c) < 0$, they are not well-defined when $\phi_+(c) \cdot \phi_-(c) > 0$ because the bias term can be removed by a suitable choice of bandwidths and the bias-variance trade-off breaks down.[2] (ii) When the trade-off breaks down, a new optimality criterion becomes necessary in order to take higher-order bias terms into consideration. We define the asymptotically first-order optimal (AFO) bandwidths, following AI.

DEFINITION 1 *The AFO bandwidths for the FRD estimator* minimize the AMSE defined by

$$AMSE_{1n}(h_+, h_-) = \frac{1}{(\tau_D(c))^2}\left\{\phi_+(c)h_+^2 - \phi_-(c)h_-^2\right\}^2 + \frac{v}{nf(c)(\tau_D(c))^2}\left\{\frac{\omega_+(c)}{h_+} + \frac{\omega_-(c)}{h_-}\right\},$$

when $\phi_+(c) \cdot \phi_-(c) < 0$. When $\phi_+(c) \cdot \phi_-(c) > 0$, *the AFO bandwidths for the FRD estimator* minimize the AMSE defined by

$$AMSE_{2n}(h_+, h_-) = \frac{1}{(\tau_D(c))^2}\left\{\psi_+(c)h_1^3 - \psi_-(c)h_-^3\right\}^2 + \frac{v}{nf(c)(\tau_D(c))^2}\left\{\frac{\omega_+(c)}{h_+} + \frac{\omega_-(c)}{h_-}\right\}$$

subject to the restriction $\phi_+(c)h_+^2 - \phi_-(c)h_-^2 = 0$ under the assumption of $\psi_+(c) - \{\phi_+(c)/\phi_-(c)\}^{3/2}\psi_-(c) \neq 0$.

When $\phi_+(c) \cdot \phi_-(c) < 0$, the AFO bandwidths minimize the standard AMSE (2). When $\phi_+(c) \cdot \phi_-(c) > 0$, the AFO bandwidths minimize the sum of the squared second-order bias term and the variance term under the restriction that the first-order bias term be removed. Inspecting the objective function, the resulting AFO bandwidths are $O(n^{-1/5})$ when $\phi_+(c) \cdot \phi_-(c) < 0$ and $O(n^{-1/7})$ when $\phi_+(c) \cdot \phi_-(c) > 0$.[3]

When $\phi_+(c) \cdot \phi_-(c) > 0$, Definition 1 shows that MSE_n is of order $O(n^{-6/7})$, which implies that the MSE based on the AFO bandwidths converges to zero faster than $O(n^{-4/5})$, the rate attained by the single bandwidth approaches such as the IK method. When $\phi_+(c) \cdot \phi_-(c) < 0$, they are of the same order. However, it can be shown that the ratio of the AMSE based on the AFO bandwidths to that based on IK never exceeds one asymptotically (see Section 2.2 of AI).

---

[2]This is the reason why IK proceed with assuming $h_+ = h_-$.

[3]The explicit expression of the AFO bandwidths are provided in the Supplemental Material.

2.2 Feasible Automatic Bandwidth Choice

The feasible bandwidths is based on a modified version of the estimated AMSE (MMSE) as in AI. It is defined by

$$
MMSE_n^p(h_+, h_-) = \left\{ \hat{\phi}_+(c)h_1^2 - \hat{\phi}_-(c)h_-^2 \right\}^2 + \left\{ \hat{\psi}_+(c)h_1^3 - \hat{\psi}_-(c)h_-^3 \right\}^2
$$
$$
+ \frac{v}{n\hat{f}(c)} \left\{ \frac{\hat{\omega}_+(c)}{h_1} + \frac{\hat{\omega}_-(c)}{h_-} \right\} \quad (3)
$$

where $\hat{\phi}_j(c)$, $\hat{\psi}_j(c)$, $\hat{\omega}_j(c)$ and $\hat{f}(c)$ are consistent estimators of $\phi_j(c)$, $\psi_j(c)$, $\omega_j(c)$ and $f(x)$ for $j = +, -$, respectively. A key characteristic of the MMSE is that one does not need to know the sign of the product of the second derivatives a priori and that there is no need to solve the constrained minimization problem.

Let $(\hat{h}_+, \hat{h}_-)$ be a combination of bandwidths that minimizes the MMSE given in (3). The next theorem shows that $(\hat{h}_+, \hat{h}_-)$ is asymptotically as good as the AFO bandwidths.

THEOREM 1 *Suppose that the conditions stated in Lemma 1 hold. Assume further that $\hat{\phi}_j(c) \xrightarrow{p} \phi_j(c)$, $\hat{\psi}_j(c) \xrightarrow{p} \psi_j(c)$, $\hat{f}(c) \xrightarrow{p} f(c)$ and $\hat{\omega}_j(c) \xrightarrow{p} \omega_j(c)$ for $j = +, -$, respectively. Also assume $\psi_+(c) - \{\phi_+(c)/\phi_-(c)\}^{3/2}\psi_-(c) \neq 0$. Then, the following hold.*

$$
\frac{\hat{h}_+}{h_+^\dagger} \xrightarrow{p} 1, \quad \frac{\hat{h}_-}{h_-^\dagger} \xrightarrow{p} 1, \quad and \quad \frac{MMSE_n^p(\hat{h}_+, \hat{h}_-)}{MSE_n(h_+^\dagger, h_-^\dagger)} \xrightarrow{p} 1.
$$

where $(h_+^\dagger, h_-^\dagger)$ are the AFO bandwidths.

The first part of Theorem 1 shows that the bandwidths based on the plug-in version of the MMSE are asymptotically equivalent to the AFO bandwidths and the second part exhibits that the minimum value of the MMSE is asymptotically the same as the MSE evaluated at the AFO bandwidths. Theorem 1 shows that the bandwidths based on the MMSE possess the desired asymptotic properties. Theorem 1 calls for pilot estimates for $\phi_j(c)$, $\psi_j(c)$, $f(c)$ and $\omega_j(c)$ for $j = +, -$. A detailed procedure about how to obtain the pilot estimates is given in the Supplemental Material.

3 Simulation

We conduct simulation experiments that illustrate the advantage of the proposed method and a potential gain for using bandwidths tailored to the FRD design over the bandwidths tailored to the SRD design. Application of a bandwidth developed for the SRD to the FRD context seems common in practice.[4]

Simulation designs are as follows. For the treatment probability, $E[D|X = x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp[-\frac{(u+1.28)^2}{2}]du$ for $x \geq 0$ and $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp[-\frac{(u-1.28)^2}{2}]du$ for $x < 0$. This leads to the discontinuity size of 0.8. The graph is depicted in Figure 1-(a).

For the conditional expectation functions of the observed outcome, $E[Y|X = x]$, we consider two designs, which are essentially the same as Designs 2 and 4 of AI for the the SRD design. The specification for the assignment variable and the additive error are exactly the same as that considered by IK.[5] They are depicted in Figures 1-(b) and 1-(c). We use data sets of 500 observations with 10,000 repetitions.

(a) Treatment Probability, $E[D|X = x]$.   (b) Design 1. Ludwig and Miller Data I   (c) Design 2. Ludwig and Miller Data II
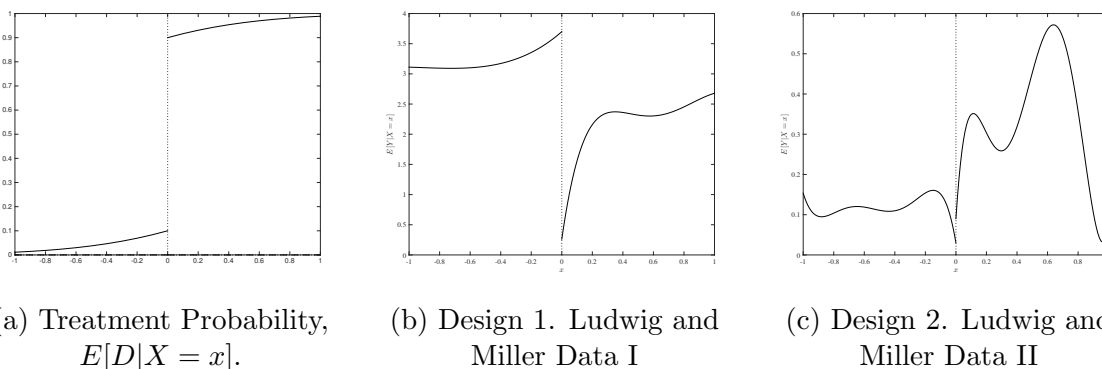
Figure 1: Simulation Designs. (a) Treatment probability, (b) $E[Y|X = x]$ for Design 1, (c) $E[Y|X = x]$ for Design 2.

The results are presented in Table 1 and Figure 2. Four bandwidth selection methods, MMSE-f, MMSE-s, IK-f and IK-s, are examined. MMSE-f is the new method proposed in the paper and MMSE-s is the one proposed for the SRD design by AI. IK-f and IK-s are the methods proposed by IK for the FRD and SRD designs, respectively.

---

[4]For example, Imbens and Kalyanaraman (2012, Section 5.1) state that the bandwidth choice for the FRD estimator is often similar to the choice for the SRD estimator of only the numerator of the FRD estimand.

[5]The exact functional form of $E[Y|X = x]$, the specification for the assignment variable and the additive error are provided in the Supplemental Material

Table 1: Bias and RMSE for the FRDD, n=500

| Design	Method	$\hat{h}_+$ Mean	SD	$\hat{h}_-$ Mean	SD	$\hat{\tau}$ Bias	RMSE	Efficiency
1	MMSE-f	0.056	0.033	0.097	0.100	0.037	0.168	1
	IK-f	0.177	0.033			0.180	0.192	0.876
	MMSE-s	0.235	0.109	0.489	0.259	0.339	0.373	0.450
	IK-s	0.325	0.068			0.443	0.451	0.373
2	MMSE-f	0.226	0.091	0.624	0.214	-0.002	0.072	1
	IK-f	0.284	0.052			0.087	0.097	0.739
	MMSE-s	0.227	0.092	0.628	0.215	-0.002	0.072	1
	IK-s	0.337	0.072			0.093	0.101	0.709

The bandwidths for MMSE-s and IK-s are computed based only on the numerator of the FRD estimator and the same bandwidths are used to estimate the denominator. Table 1 reports the mean and standard deviation of the bandwidths, the bias and root mean squared error (RMSE) for the FRD estimates, and the relative efficiency based on the RMSE.[6] Figure 2 shows the simulated CDF for the distance of the FRD estimate from the true value.

Examining Table 1 and Figure 2-(a), for Design 1, MMSE-f performs significantly better than all other methods. For Design 2, Table 1 and Figure 2-(b) indicate that MMSE-f and MMSE-s performs comparably but clearly dominate IK methods currently widely used. In cases we examined, the new method performs better than currently available methods and using methods specifically developed for the FRD dominates the method developed for the SRD.

Acknowledgement

---

[6]The bias and RMSE are 5% trimmed versions since unconditional finite sample variance is infinite.
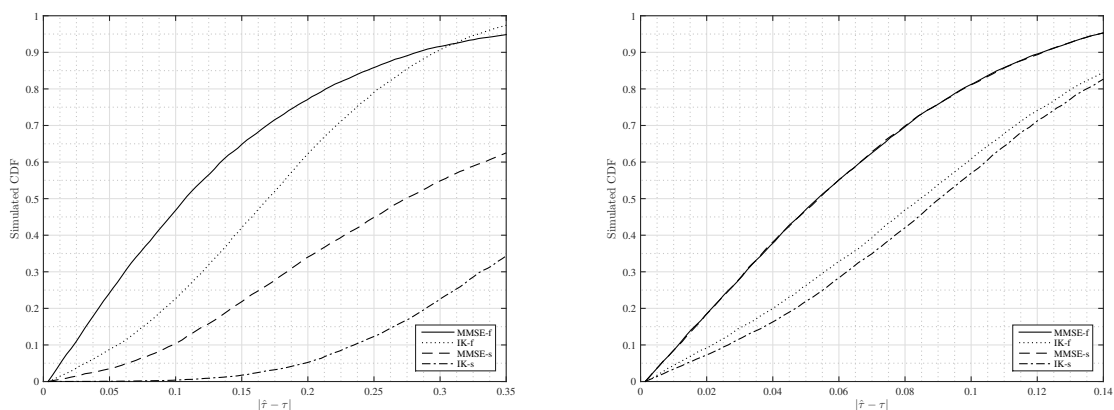
(a) Design 1. Ludwig and Miller Data I   (b) Design 2. Ludwig and Miller Data II

Figure 2: Simulated CDF of $|\hat{\tau} - \tau|$ for different bandwidth selection rules

References

ARAI, Y., AND H. ICHIMURA (2015): "Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator," CIRJE-F-889, CIRJE, Faculty of Economics, University of Tokyo.

DONG, Y., AND A. LEWBEL (forthcoming): "Identifying the effect of changing the policy threshold in regression discontinuity," *Review of Economics and Statistics*.

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

HAHN, J., P. TODD, AND W. VAN DER KLAAUW (2001): "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, 69, 201–209.

IMBENS, G. W., AND K. KALYANARAMAN (2012): "Optimal bandwidth choice for the regression discontinuity estimator," *Review of Economic Studies*, 79, 933–959.

PORTER, J. (2003): "Estimation in the regression discontinuity model," Mimeo.

Supplement to "Optimal Bandwidth Selection for the Fuzzy Regression Discontinuity Estimator"

Yoichi Arai[*]and Hidehiko Ichimura[†]

Introduction

Section 1 provides an explicit expression of the AFO bandwidths. Section 2 describes details about our simulation experiment. A sketch of the proof for Lemma 1 is provided in Section 3. Section 4 briefly describes how to obtain pilot estimates.

1 Definition of the AFO Bandwidths

DEFINITION 1 *The AFO bandwidths for the fuzzy RD estimator* minimize the AMSE defined by

$$AMSE_{1n}(h) = \frac{1}{(\tau_D(c))^2}\left\{\phi_+(c)h_+^2 - \phi_-(c)h_-^2\right\}^2 + \frac{v}{nf(c)(\tau_D(c))^2}\left\{\frac{\omega_+(c)}{h_+} + \frac{\omega_-(c)}{h_-}\right\}.$$

when $\phi_+(c) \cdot \phi_-(c) < 0$. Their explicit expressions are given by $h_+^* = \theta^* n^{-1/5}$ and $h_-^* = \lambda^* h_+^*$, where

$$\theta^* = \left\{\frac{v\omega_+(c)}{4f(c)\phi_+(c)\left[\phi_+(c) - \lambda^{*2}\phi_-(c)\right]}\right\}^{1/5} \quad \text{and} \quad \lambda^* = \left\{-\frac{\phi_+(c)\omega_-(c)}{\phi_-(c)\omega_+(c)}\right\}^{1/3}. \quad (1)$$

---
[*]National Graduate Institute for Policy Studies (GRIPS), 7-22-1 Roppongi, Minato-ku, Tokyo 106-8677, Japan; yarai@grips.ac.jp

[†]Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan; ichimura@e.u-tokyo.ac.jp

When $\phi_+(c) \cdot \phi_-(c) > 0$, *the AFO bandwidths for the fuzzy RD estimator* minimize the AMSE defined by

$$AMSE_{2n}(h) = \frac{1}{(\tau_D(c))^2}\left\{\psi_+(c)h_1^3 - \psi_-(c)h_-^3\right\}^2 + \frac{v}{nf(c)(\tau_D(c))^2}\left\{\frac{\omega_+(c)}{h_+} + \frac{\omega_-(c)}{h_-}\right\}$$

subject to the restriction $\phi_+(c)h_+^2 - \phi_-(c)h_-^2 = 0$ under the assumption of $\psi_+(c) - \{\phi_+(c)/\phi_-(c)\}^{3/2}\psi_-(c) \neq 0$. Their explicit expressions are given by $h_+^{**} = \theta^{**}n^{-1/7}$ and $h_-^{**} = \lambda^{**}h_+^{**}$, where

$$\theta^{**} = \left\{\frac{v\left[\omega_+(c) + \omega_-(c)/\lambda^{**}\right]}{6f(c)\left[\psi_+(c) - \lambda^{**3}\psi_-(c)\right]^2}\right\}^{1/7} \quad \text{and} \quad \lambda^{**} = \left\{\frac{\phi_+(c)}{\phi_-(c)}\right\}^{1/2}.$$

2   Simulation Designs

Let $\ell_1(x) = E[Y(1)|X = x]$ and $\ell_0 = E[Y(0)|X = x]$. A functional form for each design is given as follows:

(a) Design 1

$$\ell_j(x) = \begin{cases} \alpha_j + 18.49x - 54.8x^2 + 74.3x^3 - 45.02x^4 + 9.83x^5 & \text{if } x > 0, \\ \alpha_j + 2.99x + 3.28x^2 + 1.45x^3 + 0.22x^4 + 0.03x^5 & \text{if } x \leq 0, \end{cases}$$

where $(\alpha_1, \alpha_0) = (-0.17, 4.13)$.

(b) Design 2

$$\ell_j(x) = \begin{cases} \alpha_j + 5.76x - 42.56x^2 + 120.90x^3 - 139.71x^4 + 55.59x^5 & \text{if } x > 0, \\ \alpha_j - 2.26x - 13.14x^2 - 30.89x^3 - 31.98x^4 - 12.1x^5 & \text{if } x \leq 0, \end{cases}$$

where $(\alpha_1, \alpha_0) = (0.0975, 0.0225)$.

The assignment variable X_i is given by $2Z_i - 1$ for each design where Z_i have a Beta distribution with parameters $\alpha = 2$ and $\beta = 4$. We consider a normally distributed additive error term with mean zero and standard deviation 0.1295 for the outcome equation.

3   Proofs

Proof of Lemma 1: As in the proof of Lemma A2 of Calonico, Cattaneo, and Titiunik (2014), we utilize the following expansion

$$\frac{\hat{\tau}_Y(c)}{\hat{\tau}_D(c)} - \frac{\tau_Y(c)}{\tau_D(c)} = \frac{1}{\tau_D(c)}\left(\hat{\tau}_Y(c) - \tau_Y(c)\right) - \frac{\tau(c)}{\tau_D(c)}\left(\hat{\tau}_D(c) - \tau_D(c)\right)$$

$$+ \frac{\tau(c)}{\tau_D(c)\hat{\tau}_D(c)}\left(\hat{\tau}_D(c) - \tau_D(c)\right)^2 - \frac{1}{\tau_D(c)\hat{\tau}_D(c)}\left(\hat{\tau}_Y(c) - \tau_Y(c)\right)\left(\hat{\tau}_D(c) - \tau_D(c)\right). \quad (2)$$

Since the treatment of the variance component is exactly the same as that by IK, we only discuss the bias component. Observe that Lemma 1 of Arai and Ichimura (2015) implies the bias of $\hat{\tau}_Y(c)$ and $\hat{\tau}_D(c)$ are equal to

$$C_1\left[m_{Y+}^{(2)}(c)h_+^2 - m_{Y-}^{(2)}(c)h_-^2\right] + \left[\zeta_{Y1}(c)h_+^3 - \zeta_{Y0}(c)h_-^3\right] + o\left(h_+^3 + h_-^3\right)$$

and

$$C_1\left[m_{D+}^{(2)}(c)h_+^2 - m_{D-}^{(2)}(c)h_-^2\right] + \left[\zeta_{D1}(c)h_+^3 - \zeta_{D0}(c)h_-^3\right] + o\left(h_+^3 + h_-^3\right),$$

respectively. Combining these with the expansion given by (2) produces the required result. ∎

4   Procedures to Obtain Pilot Estimates

Procedures to obtain pilot estimates for $m_{Yj}^{(2)}(c)$, $m_{Yj}^{(3)}(c)$, $f(c)$, $f^{(1)}(c)$, and $\sigma_{Yj}^2(c)$, for $j = +, -$, are exactly the same as those for the sharp RD design by AI (see Appendix A of AI). Pilot estimates for $m_{Dj}^{(2)}(c)$, $m_{Dj}^{(3)}(c)$, $\sigma_{Dj}^2(c)$, for $j = +, -$, can be obtained by replacing the role of Y by D in Step 2 and 3 of Appendix A of AI. Pilot estimates for $\sigma_{YDj}(c)$, $j = +, -$ are obtained analogously to $\sigma_{Yj}^2(c)$. We obtain a pilot estimate for $\tau_D(c)$ by applying the sharp RD framework with the outcome variable D and the assignment variable X.

References

ARAI, Y., AND H. ICHIMURA (2015): "Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator," CIRJE-F-889, CIRJE, Faculty of Economics, University of Tokyo.

CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014): "Robust nonparametric bias-corrected inference in the regression discontinuity design," *Econometrica*, 6, 2295–2326.