Economics Letters 141 (2016) 103–106

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

# Optimal bandwidth selection for the fuzzy regression discontinuity estimator

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

[a] National Graduate Institute for Policy Studies (GRIPS), 7-22-1 Roppongi, Minato-ku, Tokyo 106-8677, Japan

[b] Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

## HIGHLIGHTS

- A new bandwidth selection rule for the FRD estimator based on the LLR is proposed.
- The method chooses two bandwidths, one for each side of the cut-off point.
- The method yields convergence rate improvement over currently popular methods.
- The theoretical advantage is shown to realize in our simulation study.

ARTICLE INFO

*Article history:*
Received 1 September 2015
Received in revised form
20 January 2016
Accepted 29 January 2016
Available online 12 February 2016

*JEL classification:*
C13
C14

*Keywords:*
Bandwidth selection
Fuzzy regression discontinuity design
Local linear regression

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.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

The fuzzy regression discontinuity (FRD) estimator, developed by Hahn et al. (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, and the refined version of their bandwidth is proposed by Calonico et al. (2014).

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. While the SRD estimator is constructed by the difference of the nonparametric estimators, the FRD estimator consists of the ratio of the difference of the nonparametric estimators. Hence the formula for the asymptotic approximation of the mean squared error (MSE) is more involved. In principle, we need to choose four bandwidths for four nonparametric estimators to obtain the FRD estimator. We extend the approach by Arai and Ichimura (2015) to the FRD estimator by reducing the problem to selection of two bandwidths. A simulation study illustrates the potential usefulness of the proposed method.[1]

* Corresponding author. Tel.: +81 3 6439 6177.
*E-mail address:* yarai@grips.ac.jp (Y. Arai).

[1] Matlab and Stata codes to implement the proposed method are available at http://www3.grips.ac.jp/~yarai/.

http://dx.doi.org/10.1016/j.econlet.2016.01.024
0165-1765/© 2016 Elsevier B.V. All rights reserved.

## 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 $Y_i = D_iY_i(1)+(1-D_i)Y_i(0)$. Throughout the paper, we assume that $(Y_1, D_1, X_1), \dots, (Y_n, D_n, X_n)$ 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, $\sigma^2_{Yj}(c) > 0$, $\sigma^2_{Dj}(c) > 0$, $\sigma_{YDj}(c)$, and the second and third derivatives $m^{(2)}_{Yj}(c)$, $m^{(3)}_{Yj}(c)$, $m^{(2)}_{Dj}(c)$, $m^{(3)}_{Dj}(c)$, 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 (2015), the local average treatment effect 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 simplify 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 $m_{D-}(c)$. It is possible that choosing four bandwidths improves the performance of the FRD estimator. There are two reasons why we do not pursue this direction. One is that using bandwidths, one for each side of the cut-off point, seems reasonable. That is, we use observations of the same individuals to estimate the conditional mean functions on the denominator and the numerator for each side. Another is that choosing four bandwidths makes the problem complicated fundamentally. See Section 2.1 of Arai and Ichimura (2015) about how the bandwidth selection problem can be ill-behaved.

### 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 $k$th 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 $\alpha$ 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 2 implies that individuals do not manipulate the assignment variable. Violation of this can result in the invalidity of the RD design as pointed out by McCrary (2008).

**Assumption 3.** Let $\delta$ be some positive constant. The $m_{Y+}$, $\sigma^2_{Y+}$ and $\sigma_{YD+}$ 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 $m_{Y-}$, $\sigma^2_{Y-}$ and $\sigma_{YD-}$ 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*

$$\begin{aligned} MSE_n(h_+, h_-) = \frac{1}{(\tau_D(c))^2}\Big\{\Big[\phi_+(c)h_+^2 - \phi_-(c)h_-^2\Big] \\ + \Big[\psi_+(c)h_+^3 - \psi_-(c)h_-^3\Big] + o\left(h_+^3 + h_-^3\right)\Big\}^2 \\ + \frac{v}{nf(c)(\tau_D(c))^2}\Big\{\frac{\omega_+(c)}{h_+} \\ + \frac{\omega_-(c)}{h_-}\Big\} + o\left(\frac{1}{nh_+} + \frac{1}{nh_-}\right) \end{aligned} \qquad (1)$$

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

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

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

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

$$\begin{aligned} AMSE(h_+, h_-) = \frac{1}{(\tau_D(c))^2}\left\{\phi_+(c)h_+^2 - \phi_-(c)h_-^2\right\}^2 \\ + \frac{v}{nf(c)(\tau_D(c))^2}\left\{\frac{\omega_+(c)}{h_+} + \frac{\omega_-(c)}{h_-}\right\}. \end{aligned} \qquad (2)$$

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.

[2] This is the reason why IK proceed by assuming $h_+ = h_-$ (see Section 3.1 of IK).

Download English Version:

https://daneshyari.com/en/article/5058298

Download Persian Version:

https://daneshyari.com/article/5058298

Daneshyari.com