Economics Letters 153 (2017) 23–27

Contents lists available at ScienceDirect

Economics Letters

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

Optimal bandwidth selection for local linear estimation of discontinuity in density

Hugo Jales [a], Jun Ma [b,*], Zhengfei Yu [c]

[a] Department of Economics, Center for Policy Research, Syracuse University, United States
[b] School of Economics, Renmin University of China, Beijing, China
[c] Faculty of Humanities and Social Sciences, University of Tsukuba, Tsukuba, Japan

HIGHLIGHTS

- A new bandwidth selection rule for estimating functions of the discontinuity in the density is proposed.
- The rule selects different bandwidths for both sides of the discontinuity and yields smaller asymptotic MSE than currently popular methods.
- A simulation study demonstrates the advantage of the proposed bandwidth selection rule.
- An empirical example illustrates the practical usefulness of our bandwidth selection rule.

ARTICLE INFO

Article history:
Received 19 November 2016
Received in revised form 17 January 2017
Accepted 23 January 2017
Available online 26 January 2017

JEL classification:
C13
C14

Keywords:
Bandwidth selection
Density discontinuity
Local linear estimation

ABSTRACT

We derive an optimal bandwidth selection rule for estimating a function of two one-sided limits of the probability density function at two cut-off points by the local linear estimation method. Our rule follows Arai and Ichimura (2015, AI, hereafter)'s principle and selects two bandwidths simultaneously, one for each cut-off point. Simulation results are presented. We also illustrate the usefulness of the method in an empirical application.

1. Introduction

Economic researchers are often interested in estimating the size of the discontinuity of the density of a random variable at a point, in order to measure the agent's response to a discrete change in policy or to recover structural parameters. See Jales and Yu (forthcoming) for a recent review of the literature. The nonparametric estimation of (a function of) a discontinuity in a density function calls for a sophisticated bandwidth selector. There is no reason to believe that the standard (pointwise or integrated) MSE-optimal bandwidths will also be optimal for the ratio of the one-sided limits of the density. Thus, at this point, there is no formal guidance to the applied researcher about how to choose the tuning parameters in this setup. This paper fills this gap in the literature: We provide explicit formulas for the asymptotic first-order optimal (AFO, hereafter) bandwidths for known non-linear functions of the discontinuity in the density. To do that, we extend the results from AI, which provided AFO bandwidths for local linear estimation for the regression discontinuity design.

Suppose our observations $\{X_i : i = 1, \ldots, n\}$ are i.i.d. random variables with Lebesgue density f. We assume that f is compactly supported on $[\underline{x}, \overline{x}]$. For some known cut-off points (c_-, c_+) satisfying $\underline{x} < c_- \leqslant c_+ < \overline{x}$, define $f_-(z) \equiv f(z)$ for $z < c_-$ and $f_+(z) \equiv f(z)$ for $z > c_+$. Denote $\varphi_-^{(k)} \equiv \lim_{z \uparrow c_-} f_-^{(k)}(z)$ and $\varphi_+^{(k)} \equiv \lim_{z \downarrow c_+} f_+^{(k)}(z)$, where $f_s^{(k)}$ denotes the kth order derivative of f_s, for $s \in \{-, +\}$. For $k = 0$, we denote φ_+ and φ_- for simplicity. Suppose that we are interested in estimating $\rho(\varphi_-, \varphi_+)$,[1] for some known nonlinear function $\rho : \mathbb{R}^2 \to \mathbb{R}$. Examples include

* Correspondence to: School of Economics, Renmin University of China, 59 Zhongguancun Street, Beijing 100872, China.
E-mail address: jun.ma@ruc.edu.cn (J. Ma).

[1] Notice that the estimation problem we consider in this paper is different from that of Arai and Ichimura (2013), which considers estimating the *difference* between the values of density functions at two *interior* points.

http://dx.doi.org/10.1016/j.econlet.2017.01.024

Doyle (2007), Gerard et al. (2015), Jales (2015) in labor economics, Bajari et al. (2011) in health economics, and Saez (2010), Kopczuk and Munroe (2015) in public economics.[2] Let K denote a kernel function assumed to be supported on $[-1, 1]$, h denote a bandwidth and $K_h \equiv h^{-1}K(\cdot/h)$. Denote

$$\mathrm{m}_{j,-} \equiv \int_{-1}^{0} u^j K(u)\,du, \quad \mathrm{m}_{j,+} \equiv \int_{0}^{1} u^j K(u)\,du,$$
$$\mathrm{v}_{j,-} \equiv \int_{-1}^{0} z^j K(z)^2 dz, \quad \mathrm{v}_{j,+} \equiv \int_{0}^{1} z^j K(z)^2 dz.$$

Let $\mathbb{1}(\cdot)$ denote the indicator function. We consider the following local linear estimators[3] for φ_+ and φ_-:

$$\widehat{\varphi_+} \equiv \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathrm{m}_{2,+}}{\mathrm{m}_{0,+}\mathrm{m}_{2,+}-\mathrm{m}_{1,+}^2} - \frac{\mathrm{m}_{1,+}}{\mathrm{m}_{0,+}\mathrm{m}_{2,+}-\mathrm{m}_{1,+}^2}\cdot\frac{X_i-c_+}{h_+}\right\} \times \mathbb{1}(X_i > c_+)K_{h_+}(X_i-c_+)$$
$$\widehat{\varphi_-} \equiv \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathrm{m}_{2,-}}{\mathrm{m}_{0,-}\mathrm{m}_{2,-}-\mathrm{m}_{1,-}^2} - \frac{\mathrm{m}_{1,-}}{\mathrm{m}_{0,-}\mathrm{m}_{2,-}-\mathrm{m}_{1,-}^2}\cdot\frac{X_i-c_-}{h_-}\right\} \times \mathbb{1}(X_i < c_-)K_{h_-}(X_i-c_-),$$

where two different bandwidths (h_-, h_+) are used. Bickel and Doksum (2015) referred to such estimators as minimum contrast estimators (MCE, hereafter). It is noted by Bickel and Doksum (2015) that the MCE is asymptotically unbiased when x belongs to the boundary region $[\underline{x}, \underline{x}+h) \cup (\overline{x}-h, \overline{x}]$ and coincides with the ordinary kernel density estimator when x is an interior point.

2. Optimal bandwidth selection

Assumption 1 (*Kernel and Bandwidths*). (i) $K : \mathbb{R} \to \mathbb{R}$ is a symmetric and continuous probability density function that is supported on $[-1, 1]$. (ii) The bandwidth h_s is positive and satisfies $h_s \to 0$ and $nh_s \to \infty$ as $n \to \infty$, for $s \in \{-, +\}$.

Assumption 2 (*Data Generating Process*). Let δ denote some positive constant. On a one-sided neighborhood $(c_- - \delta, c_-)$, f_- has a Lipschitz fourth-order derivative. An analogous condition holds for f_+ with some one-sided neighborhood $(c_+, c_+ + \delta)$.

Assumption 3 (*Parameter of Interest*). ρ is twice continuously differentiable on a neighborhood around (φ_-, φ_+).

Assumption 1 is identical to Assumptions 1 and 2 of AI. Assumption 2 and the continuous extension theorem guarantee the existence of $\varphi_s^{(k)}$ for $(s, k) \in \{-, +\} \times \{0, 1, \ldots, 4\}$. Assumption 3 is satisfied in all of existing economic applications we are aware of.

Denote

$$\rho_-^{(1)} \equiv \left.\frac{\partial\rho(z_1,z_2)}{\partial z_1}\right|_{(z_1,z_2)=(\varphi_-,\varphi_+)} \quad \text{and}$$
$$\rho_+^{(1)} \equiv \left.\frac{\partial\rho(z_1,z_2)}{\partial z_2}\right|_{(z_1,z_2)=(\varphi_-,\varphi_+)}.$$

As in Arai and Ichimura (2016), to deal with the nonlinearity of ρ, we seek a tractable stochastic approximation of $\rho(\widehat{\varphi_-}, \widehat{\varphi_+})$.

We consider a linearization[4] around $\rho(\varphi_-, \varphi_+)$: $\rho(\varphi_-, \varphi_+) + \sum_{s\in\{-,+\}}\rho_s^{(1)}(\widehat{\varphi_s} - \varphi_s)$. Let $\mathrm{MSE}_n(h_-, h_+)$ denote the mean square error (MSE) of the linearization of $\rho(\widehat{\varphi_-}, \widehat{\varphi_+})$. Denote

$$\gamma_1 \equiv \frac{1}{2}\frac{\mathrm{m}_{2,-}^2 - \mathrm{m}_{1,-}\mathrm{m}_{3,-}}{\mathrm{m}_{0,-}\mathrm{m}_{2,-}-\mathrm{m}_{1,-}^2}, \quad \gamma_2 \equiv \frac{1}{6}\frac{\mathrm{m}_{2,-}\mathrm{m}_{3,-} - \mathrm{m}_{1,-}\mathrm{m}_{4,-}}{\mathrm{m}_{0,-}\mathrm{m}_{2,-}-\mathrm{m}_{1,-}^2},$$
$$\theta \equiv \frac{\mathrm{m}_{2,-}^2\mathrm{v}_{0,-} - 2\mathrm{m}_{1,-}\mathrm{m}_{2,-}\mathrm{v}_{1,-} + \mathrm{m}_{1,-}^2\mathrm{v}_{2,-}}{\left(\mathrm{m}_{0,-}\mathrm{m}_{2,-}-\mathrm{m}_{1,-}^2\right)^2}.$$

A straightforward extension of Jiang and Doksum (2003, Theorem 2.3) gives

$$\mathrm{MSE}_n(h_-, h_+) = \left\{\gamma_1\left(\rho_-^{(1)}\varphi_-^{(2)}h_-^2 + \rho_+^{(1)}\varphi_+^{(2)}h_+^2\right) + O\left(h_+^3 + h_-^3\right)\right\}^2 + \frac{\theta}{n}\left\{\frac{\left(\rho_-^{(1)}\right)^2\varphi_-}{h_-} + \frac{\left(\rho_+^{(1)}\right)^2\varphi_+}{h_+}\right\} + o\left((nh_-)^{-1} + (nh_+)^{-1}\right). \quad (1)$$

The argument in Section 2.1 of AI shows that when $\left(\rho_-^{(1)}\varphi_-^{(2)}\right)\left(\rho_+^{(1)}\varphi_+^{(2)}\right) < 0$, minimization of the leading terms of the right hand side of (1) does not yield a meaningful bandwidth selection rule. The following result is analogous to Lemma 1 of AI and is an improvement on (1).

Lemma 1. *Suppose that Assumptions 1–3 hold. Then we have*

$$\mathrm{MSE}_n(h_-, h_+) = \left\{\gamma_1\left(\rho_-^{(1)}\varphi_-^{(2)}h_-^2 + \rho_+^{(1)}\varphi_+^{(2)}h_+^2\right) + \gamma_2\left(\rho_-^{(1)}\varphi_-^{(3)}h_-^3 - \rho_+^{(1)}\varphi_+^{(3)}h_+^3\right) + o\left(h_+^3 + h_-^3\right)\right\}^2 + \frac{\theta}{n}\left\{\frac{\left(\rho_-^{(1)}\right)^2\varphi_-}{h_-} + \frac{\left(\rho_+^{(1)}\right)^2\varphi_+}{h_+}\right\} + o\left((nh_-)^{-1} + (nh_+)^{-1}\right). \quad (2)$$

Remark 1. Now define the asymptotic MSE to be the leading terms of the right hand side of (2).

$$\mathrm{AMSE}_n(h_-, h_+) \equiv \left\{\gamma_1\left(\rho_-^{(1)}\varphi_-^{(2)}h_-^2 + \rho_+^{(1)}\varphi_+^{(2)}h_+^2\right) + \gamma_2\left(\rho_-^{(1)}\varphi_-^{(3)}h_-^3 - \rho_+^{(1)}\varphi_+^{(3)}h_+^3\right)\right\}^2 + \frac{\theta}{n}\left\{\frac{\left(\rho_-^{(1)}\right)^2\varphi_-}{h_-} + \frac{\left(\rho_+^{(1)}\right)^2\varphi_+}{h_+}\right\}.$$

Following the same argument as that in Section 2.1 of AI, we can see that when $\left(\rho_-^{(1)}\varphi_-^{(2)}\right)\left(\rho_+^{(1)}\varphi_+^{(2)}\right) < 0$, minimization of $\mathrm{AMSE}_n(h_-, h_+)$ still does not yield a meaningful bandwidth selection rule. Following AI, we define an optimal bandwidth selection rule under the AFO criterion.

Remark 2. It is shown in the supplement that the conclusion of Lemma 1 still holds for the popular "binning" local linear estimator (see, e.g., McCrary (2008) and references therein) provided that the "bin width" selected by the researcher is of order $o\left(h_s^{5/2}\right)$, for $s \in \{-, +\}$. The AFO bandwidths provided below can still be interpreted as certain optimizers of leading terms of the MSE. Therefore we recommend using the same bandwidth selection rule if the researcher chooses the binning estimator instead.

[2] The parameter of interest in Doyle (2007), Jales (2015) is of the functional form $\rho(z_1, z_2) = z_2/z_1$. Gerard et al. (2015) considers estimating a parameter taking the functional form $\rho(z_1, z_2) = 1 - z_2/z_1$. Bajari et al. (2011) considers $\rho(z_1, z_2) = 1/z_2 - 1/z_1$. The economic parameter to be estimated could be an implicit function of φ_- and φ_+, whose partial derivatives can be computed using the implicit function theorem. See, e.g., Saez (2010, Equation 5).

[3] See Bickel and Doksum (2015, Chapter 11.3) for derivation of such estimators and references therein on earlier development of this estimation method. See Remark 2 below for discussion of an alternative local linear estimator via binning.

[4] The conclusion of Lemma 1 still holds for a quadratic (or even higher-order) approximation.