Economics Letters 140 (2016) 19–22

Contents lists available at ScienceDirect

Economics Letters

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

Gradient-based bandwidth selection for estimating average derivatives

Cong Li [a,b], Yanfei Wang [c,*]

[a] School of Economics, Shanghai University of Finance & Economics, Shanghai, 200433, China
[b] Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai, 200433, China
[c] International School of Economics and Management, Capital University of Economics and Business, Beijing, 100070, China

HIGHLIGHTS

- We propose a new average derivative estimator based on pointwise derivative estimate.
- Our estimator reaches the optimal convergence rate.
- The superiority of our proposed estimator is illustrated by simulation.

ARTICLE INFO

Article history:
Received 21 September 2015
Received in revised form
23 November 2015
Accepted 7 December 2015
Available online 4 January 2016

JEL classification:
C14
C18
C21

Keywords:
Average derivatives estimation
Kernel smoothing
Optimal bandwidth selection

ABSTRACT

In this paper, we propose to estimate average derivatives of a function by averaging the sample pointwise local linear derivatives with the bandwidth being selected optimally. Our estimator has better finite sample performance than that of Li, Lu & Ullah (2003) because our pointwise derivative estimate reaches the optimal convergence rate. Simulations confirm our theoretical analysis.

1. Introduction

In recent years, much nonparametric literature have studied the estimation of marginal effect which is of crucial interest to applied economists. Early work on kernel estimation of the pointwise derivative includes Casser and Muller (1984) and Rilstone and Ullah (1989). Then, Hardle and Stoker (1989), Rilstone (1991), Powell et al. (1989) and Hardle et al. (1992), and Newey and Stoker (1993) propose $\sqrt{n}$-consistent average derivative estimators, and these estimators are based on Nadaraya and Watson's local constant estimators.

Local linear estimator is first studied by Stone (1977) and Cleveland (1979). Fan (1992), Fan (1993), and Fan and Gijbels (1992) show that, compared to Nadaraya–Watson estimator, local linear estimator has better asymptotic properties: it has better boundary behavior, and is unbiased when the true model is linear. Therefore, many researchers construct their nonparametric conditional mean and its derivative estimators based on the local linear method. See Cleveland and Devlin (1988), Ruppert and Wand (1994), Kniesner and Li (2002), and Li et al. (2003), among others.

Li et al. (2003) propose a local linear average derivative estimator. They derive its asymptotic normal distribution and compare it with the alternative estimators proposed by Hardle and Stoker (1989) and Rilstone (1991). Their simulation results show that the local linear average derivative estimator compares well with the Hardle and Stoker (1989) and Rilstone (1991) estimators.

One shortcoming of Li et al. (2003) is that it relies on the traditional least squares cross-validation (LSCV) of condition mean for bandwidth selection. This method is not optimal for derivative estimation. The bandwidth selection is of crucial importance

* Corresponding author. Tel.: +86 10 8395 2463.
E-mail address: yanfei330@gmail.com (Y. Wang).

http://dx.doi.org/10.1016/j.econlet.2015.12.005
0165-1765/

in nonparametric estimation. Different bandwidths can generate dramatically different nonparametric estimation results. Given such importance, many studies focus on how to appropriately select the bandwidth for local linear estimation of conditional mean function. However, the optimal smoothness parameter for estimation of conditional mean may not be appropriate when estimating derivative of conditional mean function is of interest. There is an extensive literature on the bandwidth selection for derivative function estimation. Rice (1986) uses a differencing operator and an unbiased estimator of mean integrated squared error (MISE) between the estimated and the true derivative to select a optimal bandwidth for derivative function. Subsequently, Muller et al. (1987) formally propose a differencing operator for calculating gradients and naturally extend LSCV method to select optimal bandwidth. Other papers in this field include Fan and Gijbels (1995), Fan et al. (1996), Ruppert (1997), Charnigo et al. (2011) and Henderson et al. (2015), among others.

In this paper, we propose a new average derivatives estimator based on Li et al. (2003) in which the bandwidth is selected by Gradient-Based least squares Cross-Validation (GBCV) method suggested by Henderson et al. (2015). Moreover, we extend the asymptotic normality for average derivative estimator with deterministic bandwidth (e.g., Li et al. (2003)) to allow for optimally selected stochastic bandwidth.

The remaining part of the paper is organized as follows. In Section 2, we introduce our proposed average derivatives estimator with fully data-driven selected bandwidth and its asymptotic properties. Section 3 reports simulation results on estimating average derivatives by Li et al. (2003) with three different bandwidth selection methods, and evaluates their finite sample performances. Conclusion appears in Section 4.

2. Local linear estimation of average derivatives and gradient based cross-validation of bandwidth selection

We consider the following nonparametric regression:

$$Y_i = m(X_i) + U_i \tag{2.1}$$

where $X_i \in \mathbb{D}$, $\mathbb{D}$ is a compact subset of $\mathbb{R}^d$ and the unknown conditional mean function $m(\cdot)$ has continuous derivatives of total order $p+1$. Denote $m^{(1)}(x) = (\partial m(x)/\partial x_1, \ldots, \partial m(x)/\partial x_d)'$ as the $d \times 1$ partial derivative functions of $m(x)$. The local linear estimator solves the following minimization problem:

$$\min_{\{a_{ll}, b_{ll}\}} \sum_{i=1}^{n} \left(Y_i - a_{ll}(x) - b_{ll}(x)'(X_i - x)\right)^2 K\left(\frac{X_i - x}{h}\right) \tag{2.2}$$

where $K(\cdot)$ is the product kernel function and h is the bandwidth vector of dimension d. Let $\hat{a}_{ll}(x)$ and $\hat{b}_{ll}(x)$ be the solutions to (2.2), and they are consistent estimators of $m(x)$ and $m^{(1)}(x)$ respectively. Define $\delta_{ll}(x) \equiv (a_{ll}(x), b_{ll}(x)')'$, and the local linear estimator can be written explicitly as follows:

$$\delta_{ll}(x; h) = \left[\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)\begin{pmatrix}1 \\ X_i - x\end{pmatrix}\begin{pmatrix}1 & (X_i - x)'\end{pmatrix}\right]^{-1} \times \sum_{i=1}^{n}\left(\frac{X_i - x}{h}\right)\begin{pmatrix}1 \\ X_i - x\end{pmatrix} Y_i. \tag{2.3}$$

Li et al. (2003) suggest obtaining the average derivative estimator by averaging the estimated pointwise derivatives from (2.3):

$$b_{ll}(h) = \frac{1}{n}\sum_{i=1}^{n} b_{ll}\left(X_i; h\right). \tag{2.4}$$

However, in Eq. (2.4), Li et al. (2003) use Least Squares Cross-Validation (LSCV) method to select h, which is not optimal for average derivative estimation. Theoretically, the optimal bandwidth for derivative estimator should be chosen so that it minimizes the expected mean squared error $E\{[\hat{\beta}_{LL}(x) - \beta_{\text{true}}(x)]^2\}$, and the gradient-based least squares cross-validation (GBCV) function is the sample analog of the expected mean squared error. Therefore, we propose to use the bandwidth selection method by Henderson et al. (2015) which has the oracle property: their selected bandwidth is asymptotically equivalent to the *optimal* bandwidth which minimizes the oracle GBCV function assuming the true gradient were known in the local linear estimator in Muller et al. (1987). The idea of Henderson et al. (2015) is to use the local cubic estimator of the derivative function as the true unknown derivative function in the GBCV to select h.

Although we consider a nonparametric regression with multiple regressors $x = (x_1, \ldots, x_d)$, without loss of generality, we only focus on estimating the derivative of the first regressor, x_1. If the true partial derivative function $m_1^{(1)}(x)$ were known, the optimal bandwidth $h_{0,opt}$, with deterministic order $O(n^{-1/(d+6)})$, minimizes the leading term of population gradient-based least squares cross-validation (GBCV) function for the first order derivative with respect to x_1:

$$CV_0(h) = \int E\left[b_{1,ll}(x; h) - m_1^{(1)}(x)\right]^2 M(x) f(x) dx. \tag{2.5}$$

It is known that $h_{0,opt}$ can be consistently estimated by $\hat{h}_{0,opt}$ which minimizes the following oracle sample GBCV function, which is infeasible:

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\left[b_{1,ll}(X_i; h) - m_1^{(1)}(X_i)\right]^2 M\left(X_i\right), \tag{2.6}$$

with $\hat{h}_{0,opt}/h_{0,opt} = 1 + o_P(1)$, where $b_{1,ll}(x; h)$ is the local linear estimator and $M(\cdot)$ is a weight function with bounded support that trims out the data near the support boundary of x. It is also known that a good proxy for the unknown true derivative function $m_1^{(1)}(x)$ is its local cubic estimator $b_{1,lcb}(x; h)$. Let $\hat{h}_{cubic}$ minimize the following sample GBCV function based on the local cubic estimator, which is feasible:

$$CV_{cubic}(h) = \frac{1}{n}\sum_{i=1}^{n}\left[b_{1,ll}(X_i; h) - b_{1,lcb}(X_i; h)\right]^2 M\left(X_i\right). \tag{2.7}$$

It is shown in theory that, by a simple rescaling, $\tilde{h}_{0,opt} \equiv C_K^{1/(d+6)}\hat{h}_{cubic} = O_P(n^{-1/(d+6)})$ also asymptotically minimizes the infeasible cross-validation function (2.6), i.e. $\tilde{h}_{0,opt}/h_{0,opt} = 1 + o_P(1)$, under some regularity conditions that are similar to those in Li and Li (2010). The constant C_K depends only on the kernel function. For Epanechnikov and Gaussian kernels, the corresponding scaling constants are $C_{Epan} = 44/135$ and $C_{Gaussian} = 16/15$ respectively. It is easy to verify that the assumptions (A1)–(A4) in Li et al. (2003) are satisfied for our model, and that the average derivative estimator with $h_{0,opt}$ is asymptotically normally distributed. Assumptions (A1), (A2), and (A4) are standard assumptions for nonparametric estimation. Assumption (A3) requires that $nh^{2p+2} \to 0$ and $nh^{d+2}/\ln(n) \to \infty$. Substituting the rate of optimal bandwidth $h_{0,opt} \sim n^{-1/(n+6)}$ into the two restrictions in Assumption (A3), we obtain a sufficient condition for Assumption (A3), $1 - (2p+2)/(d+6) < 0$. It is equivalent to $2p > d + 4$, where p is the order of polynomial and d is the dimension of regressors. Since $\tilde{h}_{0,opt}/h_{0,opt} = 1 + o_P(1)$, according to Theorem 3.1 in Li and Li (2010), the average derivative estimators with both stochastic $\tilde{h}_{0,opt}$ and deterministic $h_{0,opt}$ have the same asymptotic normal distribution.