Economics Letters 171 (2018) 102–105

Contents lists available at ScienceDirect

Economics Letters

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

Averaging estimators for kernel regressions

Chu-An Liu

Institute of Economics, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan

HIGHLIGHTS

- We propose an averaging estimator for kernel regressions.
- We construct a weighted average of the local constant and local linear estimators at each point of estimation.
- We propose a data-driven criterion for bandwidths and weights selection with theoretical justification.

ARTICLE INFO

Article history:
Received 19 March 2018
Received in revised form 15 May 2018
Accepted 7 July 2018
Available online xxxx

JEL classification:
C14
C51
C52

Keywords:
Cross-validation
Local constant estimator
Local linear estimator
Model averaging

ABSTRACT

This paper considers model averaging for kernel regressions. We construct a weighted average of the local constant and local linear estimators at each point of estimation. We propose a two-step cross-validation method for bandwidths and weights selection, and derive the rate of convergence of the cross-validation weights to their optimal benchmark values.

1. Introduction

In recent years, there has been increasing interest in model averaging from the frequentist perspective. Model averaging aims to achieve the best trade-off between bias and variance, and has been widely applied to various models; see Claeskens and Hjort (2008) for a literature review. However, the existing literature on model averaging for kernel regressions is comparatively small. Choi and Hall (1998) suggest a convex combination of local linear estimators to reduce the bias without changing the variance. Seifert and Gasser (1996, 2000) propose a local linear ridge regression to address the problem of the unbounded conditional variance in sparse data. Cheng et al. (2007) propose a linear combination of local linear estimators to reduce the variance without affecting the bias.

In this paper, we propose a model averaging approach to reduce the weighted integrated mean squared error (WIMSE) of the kernel regression estimator. At each point of estimation, the proposed estimator is an affine combination of the local constant and local linear estimators. We first derive the WIMSE of the averaging estimator, which allows us to characterize the optimal global weights. We then propose a two-step cross-validation method for bandwidths and weights selection and provide the theoretical justification. Our simulations show that the kernel averaging estimator can achieve significant efficiency gains over the local constant and local linear estimators.

Our paper is closely related to the local ridge estimator proposed by Seifert and Gasser (2000). The main difference is that we allow the averaging estimator to have different bandwidths for the local constant and local linear estimators instead of the equal bandwidth constraint on two kernel estimators. Furthermore, Seifert and Gasser (2000) investigate the optimal local weight under the non-negative weight constraint, while we study both the optimal local and global weights and allow the weights to take on positive and negative values. In a recent paper, Henderson and Parmeter (2016) consider a kernel regression estimator that averages over local-polynomial order, kernel function, and the bandwidth selection mechanism. Their simulations provide the finite sample gain of their proposed method, but theoretical properties need to be further investigated.

E-mail address: caliu@econ.sinica.edu.tw.

https://doi.org/10.1016/j.econlet.2018.07.016

2. Model and estimation

Let $(x_1, y_1), \ldots, (x_n, y_n)$ be pairs of independent and identically distributed random variables from a joint density $f(x, y)$. The regression model is

$$y_i = m(x_i) + e_i, \quad (1)$$

where $m(x) = \mathrm{E}(y_i|x_i = x)$ is the regression function, e_i is an unobservable random error, and $\sigma^2(x) = \mathrm{E}(e_i^2|x_i = x)$ is the conditional variance function. Our goal is to estimate the unknown regression function $m(x)$ nonparametrically without imposing any assumptions on the structure of the relationship between the dependent variable y_i and the regressor x_i.

Let $k(u)$ be the kernel function and h the bandwidth. The local constant estimator is

$$\hat{m}_{\mathrm{LC}}(x) = \left(\sum_{i=1}^{n} k\left(\frac{x_i - x}{h_{\mathrm{LC}}}\right)\right)^{-1} \left(\sum_{i=1}^{n} k\left(\frac{x_i - x}{h_{\mathrm{LC}}}\right) y_i\right), \quad (2)$$

and the local linear estimator is

$$\hat{m}_{\mathrm{LL}}(x) = \left(\sum_{i=1}^{n} s_i\right)^{-1} \left(\sum_{i=1}^{n} s_i y_i\right), \quad (3)$$

$$s_i = k\left(\frac{x_i - x}{h_{\mathrm{LL}}}\right) \Big(T_{n,2}(x) - (x_i - x) T_{n,1}(x)\Big), \quad (4)$$

$$T_{n,\ell}(x) = \sum_{i=1}^{n} k\left(\frac{x_i - x}{h_{\mathrm{LL}}}\right) (x_i - x)^{\ell}, \quad \ell = 1, 2. \quad (5)$$

We now consider an averaging estimator, the local weighted estimator, for the nonparametric regression model. For each point x, the local weighted estimator is an affine combination of the local constant and local linear estimators. Let λ be the weight for the local linear estimator and $1-\lambda$ be the weight for the local constant estimator. The local weighted estimator is defined as

$$\hat{m}_{\mathrm{LW}}(x) = \lambda \hat{m}_{\mathrm{LL}}(x) + (1 - \lambda)\hat{m}_{\mathrm{LC}}(x). \quad (6)$$

Note that the weights for the averaging estimator are allowed to take on positive and negative values, that is, $\lambda \in \mathcal{L}_n$ where $\mathcal{L}_n = \{\lambda : \lambda \in (-\infty, \infty)\}$. Furthermore, we allow the averaging estimator to have different bandwidths for the local constant and local linear estimators. The local ridge estimator proposed by Seifert and Gasser (2000) corresponds to the local weighted estimator with $h_{\mathrm{LC}} = h_{\mathrm{LL}}$ and $\lambda \in [0, 1]$.

3. Asymptotic theory

Throughout the paper, we denote $\kappa_2 = \int_{-\infty}^{\infty} u^2 k(u) du$ and $\upsilon = \int_{-\infty}^{\infty} k(u)^2 du$. We now state the assumptions and main results, and leave the technical proofs to the online supplemental appendix.

Assumption 1. The regression function $m(\cdot)$ has a bounded second derivative.

Assumption 2. The marginal density function $f(\cdot)$ of x satisfies $f(x) > 0$ and $|f(x) - f(y)| \leq c|x - y|^a$ for some $0 < a < 1$.

Assumption 3. The conditional variance function $\sigma^2(\cdot)$ is bounded and continuous.

Assumption 4. The kernel function $k(\cdot)$ is a symmetric density function with compact support.

Assumption 5. The support of x is a compact set $\mathcal{X}$. The weight function $w(\cdot)$ is a nonnegative function with compact support $\mathcal{W}$, which is contained in the interior of $\mathcal{X}$.

Assumptions 1–4 are quite standard; see Conditions 1(i)–(iv) of Fan (1993). Assumption 5 is imposed to reduce the bias of the local weighted estimator at the boundary points.

Theorem 1. *Under Assumptions 1–5, if $h_{\mathrm{LC}} \to 0$, $h_{\mathrm{LL}} \to 0$, $nh_{\mathrm{LC}} \to \infty$, and $nh_{\mathrm{LL}} \to \infty$ as $n \to \infty$, and $\lim_{n\to\infty}(h_{\mathrm{LL}}/h_{\mathrm{LC}}) \to \gamma$, $0 < \gamma < \infty$, then the WIMSE of the local weighted estimator is*

$$\begin{aligned} E\int_{-\infty}^{\infty} \left(\hat{m}_{\mathrm{LW}}(x) - m(x)\right)^2 w(x) dx &= \lambda^2 \zeta_1(h_{\mathrm{LL}}) + (1-\lambda)^2 \zeta_2(h_{\mathrm{LC}}) + 2\lambda(1-\lambda)\zeta_{12}(h_{\mathrm{LL}}, h_{\mathrm{LC}}) \\ &\quad + o\left(h_{\mathrm{LL}}^4 + h_{\mathrm{LC}}^4 + h_{\mathrm{LL}}^2 h_{\mathrm{LC}}^2 + (nh_{\mathrm{LL}})^{-1} + (nh_{\mathrm{LC}})^{-1}\right), \end{aligned}$$

where

$$\begin{aligned} \zeta_1(h_{\mathrm{LL}}) &= \kappa_2^2 h_{\mathrm{LL}}^4 \int_{-\infty}^{\infty} \left(\frac{m''(x)}{2}\right)^2 w(x) dx + \frac{\upsilon}{nh_{\mathrm{LL}}} \int_{-\infty}^{\infty} \frac{\sigma^2(x)}{f(x)} w(x) dx, \\ \zeta_2(h_{\mathrm{LC}}) &= \kappa_2^2 h_{\mathrm{LC}}^4 \int_{-\infty}^{\infty} \left(\frac{m''(x)}{2} + \frac{m'(x)f'(x)}{f(x)}\right)^2 w(x) dx + \frac{\upsilon}{nh_{\mathrm{LC}}} \int_{-\infty}^{\infty} \frac{\sigma^2(x)}{f(x)} w(x) dx, \\ \zeta_{12}(h_{\mathrm{LL}}, h_{\mathrm{LC}}) &= \kappa_2^2 h_{\mathrm{LL}}^2 h_{\mathrm{LC}}^2 \int_{-\infty}^{\infty} \left(\frac{m''(x)}{2}\right) \times \left(\frac{m''(x)}{2} + \frac{m'(x)f'(x)}{f(x)}\right) w(x) dx \\ &\quad + \frac{\tilde{\upsilon}}{nh_{\mathrm{LC}}} \int_{-\infty}^{\infty} \frac{\sigma^2(x)}{f(x)} w(x) dx, \end{aligned}$$

and $\tilde{\upsilon} = \int_{-\infty}^{\infty} k(\gamma u) k(u) du$.

Theorem 1 presents the WIMSE of the local weighted estimator, and $\zeta_1(h_{\mathrm{LL}})$ and $\zeta_2(h_{\mathrm{LC}})$ represent the leading terms of WIMSE of the local linear and local constant estimators, respectively. The mean squared error of the local weighted estimator and the optimal local weights are provided in the supplemental appendix.

As shown in the appendix, the leading term of the covariance between the local linear and local constant estimators is

$$Cov\left(\hat{m}_{\mathrm{LL}}(x), \hat{m}_{\mathrm{LC}}(x)\right) = \frac{\tilde{\upsilon}}{nh_{\mathrm{LC}}} \frac{\sigma^2(x)}{f(x)}. \quad (7)$$

Note that $\tilde{\upsilon}$ is a convolution kernel function, and its value depends on the ratio of two bandwidths.[1]

The optimal global weight that minimizes the WIMSE of the local weighted estimator is

$$\lambda^o(h_{\mathrm{LL}}, h_{\mathrm{LC}}) = \frac{\zeta_2(h_{\mathrm{LC}}) - \zeta_{12}(h_{\mathrm{LL}}, h_{\mathrm{LC}})}{\zeta_1(h_{\mathrm{LL}}) + \zeta_2(h_{\mathrm{LC}}) - 2\zeta_{12}(h_{\mathrm{LL}}, h_{\mathrm{LC}})}, \quad (8)$$

and the minimized WIMSE is

$$\frac{\zeta_1(h_{\mathrm{LL}})\zeta_2(h_{\mathrm{LC}}) - \zeta_{12}(h_{\mathrm{LL}}, h_{\mathrm{LC}})^2}{\zeta_1(h_{\mathrm{LL}}) + \zeta_2(h_{\mathrm{LC}}) - 2\zeta_{12}(h_{\mathrm{LL}}, h_{\mathrm{LC}})}. \quad (9)$$

For any given h_{LC} and h_{LL}, the WIMSE given in (9) is strictly smaller than the WIMSE of any linear combination of the local linear and local constant estimators as long as $\zeta_1(h_{\mathrm{LL}}) \neq \zeta_{12}(h_{\mathrm{LL}}, h_{\mathrm{LC}})$ and $\zeta_2(h_{\mathrm{LC}}) \neq \zeta_{12}(h_{\mathrm{LL}}, h_{\mathrm{LC}})$. Note that the optimal global weight is

1 In general, the local linear estimator tends to choose a larger bandwidth than the local constant estimator. When $h_{\mathrm{LL}} > h_{\mathrm{LC}}$, γ is greater than 1, and $\tilde{\upsilon}$ is always smaller than υ. When $h_{\mathrm{LL}} = h_{\mathrm{LC}}$, we have $\tilde{\upsilon} = \upsilon$, and the covariance term degenerates into the variance term of the local constant/linear estimator.