Statistica Sinica **28** (2018), 2749-2770
doi:https://doi.org/10.5705/ss.202017.0058

BIAS REDUCTION FOR NONPARAMETRIC AND SEMIPARAMETRIC REGRESSION MODELS

Ming-Yen Cheng[1,2], Tao Huang[3], Peng Liu[4,5], and Heng Peng[1]

[1]*Hong Kong Baptist University and* [2]*National Taiwan University,*
[3]*Shanghai University of Finance and Economics,* [4]*University of Washington*
and [5]*Fred Hutchinson Cancer Research Center*

Abstract: Nonparametric and semiparametric regression models are useful statistical regression models to discover nonlinear relationships between the response variable and predictor variables. However, optimal efficient estimators for the nonparametric components in the models are biased which hinders the development of methods for further statistical inference. In this paper, based on the local linear fitting, we propose a simple bias reduction approach for the estimation of the nonparametric regression model. Our approach does not need to use higher-order local polynomial regression to estimate the bias, and hence avoids the double bandwidth selection and design sparsity problems suffered by higher-order local polynomial fitting. It also does not inflate the variance. Hence it can be easily applied to complex statistical inference problems. We extend our approach to varying coefficient models, to estimate the variance function, and to construct simultaneous confidence band for the nonparametric regression function. Simulations are carried out for comparisons with existing methods, and a data example is used to investigate the performance of the proposed method.

Key words and phrases: Simultaneous confidence band, undersmoothing, variance function estimation.

1. Introduction

Nonparametric and semiparametric regression models have been widely used to discover nonlinear relationships between response and predictor variables, and to reduce the model bias to avoid model misspecification. Efficient model estimation to make statistical inference is a necessary step for data interpretation and for finding the insufficiency of the original statistical model. Classical efficient estimators of nonparametric components in the nonparametric or semiparametric regression models are always biased. For nonparametric or semiparametric regression models, finding an unbiased estimator for the nonparametric components or reducing the bias for classical efficient nonparametric estimators is very important for further inference.

As an important technique and methodology for nonparametric regression modelling, local linear estimators enjoy numerical and theoretical advantages (see Fan (1993); Fan and Gijbels (1996); Hastie and Loader (1993), etc.). For example, they achieve the minimax efficiency among all linear estimators. Compared with higher-order local polynomial modelling, they involve less parameters and are less subject to design sparseness (see Choi, Hall and Rousson (2000)). On the other hand, when the nonparametric function shows a high degree of smoothness, local linear estimators are less appealing than those from higher-order local polynomial modelling. That is, with n as the sample size and h as the bandwidth used to estimate the nonparametric components in the model, the local cubic estimator has a bias of order $O(h^4)$ and variance of order $O((nh)^{-1})$, while the local linear estimator has a bias of order $O(h^2)$ and variance of order $O((nh)^{-1})$ even though the local cubic estimator would suffer from the sparseness of the data design. To develop an approach to reducing the bias of the local linear estimator to combine the advantages of both has attracted many statisticians, e.g. Choi and Hall (1998), Choi, Hall and Rousson (2000), He and Huang (2009), Xia (1998), and Fan and Zhang (2000), among others.

In this literature, there are two bias reduction approaches. The first is to use higher-order local polynomial fitting, such as cubic, to estimate the bias (See Xia (1998), Fan and Zhang (2000), etc.). The other approach uses information from the closest design points and model averaging to find an approximately unbiased estimator (see Choi and Hall (1998); Choi, Hall and Rousson (2000), He and Huang (2009), etc.). The first approach carries a high computational cost; one needs to select an appropriate bandwidth for the higher-order local polynomial fitting. The second approach avoids estimating the bias of the local linear estimator, and one can directly construct a nearly unbiased nonprarametric function estimator by model averaging. Here there is no problem in bandwidth selection, and there is less computational cost. Still there is the sparseness of data close to the boundary of the support area of the data. Though nonparametric function estimation can be nearly unbiased, the asymptotic variance structure of such estimators is complex which hinders its application for further statistical inference and extension to more complicated semiparametric or nonparametric regression models.

Fan and Yao (1998) gives a good review for estimation of variance of the nonparametric regression model error based on the sum squares of residuals approach, and suggests a two-step estimator of the variance function of model errors. Though they show the proposed estimator of variance function is optimally efficient, as discussed by Wang et al. (2008), the bias of the estimator of the regression function cannot be further reduced by the second stage smoothing of the squared residuals. With the nearly unbiased regression function estimator, using the sum of squares of the residuals, it is possible to construct a stable estimator of the variance, or variance function, of the model error for further inference. Bias reduction can be used in constructing a simultaneous confidence band (SCB) for nonparametric component in the model. Many methods have been proposed to cope with the bias term in the estimation of functions to achieve the right coverage probability for SCB. Most are based on undersmoothing or oversmoothing. For undersmoothing estimation of the nonparametric components see Chen and Qin (2002), Fan and Zhang (2000), Zhang and Peng (2010), Li et al. (2014). For oversmoothing, see Xia (1998). Hall and Horowitz (2013) used a bootstrap method to avoid the bias problem and construct confidence bands for nonparametric functions. The bootstrap is time consuming and does not provide general finite-sample guarantees. For undersmoothing/oversmoothing, effective bandwidth selection relies on empirical results and data themselves (Hall and Horowitz (2013)).

In this paper, we propose a bias reduction technique for local linear estimation that is easy to implement and extend. To be specific, consider the nonparametric regression model

$$Y_i = m(X_i) + \varepsilon_i, i = 1, \ldots, n,$$

where $m(\cdot)$ is an unknown smooth function and $\varepsilon_i, i = 1, \ldots, n$, are independent random errors. Denote the local linear estimator of $m(x_0)$ using bandwidth h as $\widehat{m}_h(x_0)$. We choose different $h_1, \ldots, h_B$ to obtain a series of estimates $\{\widehat{m}_{h_i}(x_0), i = 1, \ldots, B\}$ for $m(x_0)$, then perform a linear regression with $\widehat{m}_{h_i}(x_0)$ as the dependent variable and h_i^2 as the explanatory variable. The estimator of intercept term here can be regarded as a bias-reduced estimator for $m(x_0)$. By the coefficient estimate of the term h_i^2, we also get an estimate of bias term of $\widehat{m}_h(x_0)$ and then, based on such estimate of the bias, bias correction can be made to obtain the new estimator. This estimator of $m(x_0)$ has the same order of bias and variance as the local cubic estimator, and it retains the advantages of the local linear estimator such as suffering less from design sparseness.

The remainder of this paper is organized as follows. In Section 2 we give the details of our proposed bias reduction methods for the local linear regression, and investigate its asymptotic properties. In Section 3, we consider the extension of our bias reduction estimators to varying coefficient regression models. We also investigate the use of the bias reduction estimators to estimate the variance function of the error, and how to use them to construct simultaneous confidence bands for the nonparametric components in the classical nonparametric regression model. In Section 4, some numerical studies with comparison and data analysis are given. In Section 5, we give some conclusion and discussion of the proposed methods . The proofs of the main results can be found in the Supplementary Material.

2. Bias Reduction

Consider the nonparametric regression model for a bivariate random vector (X, Y),

$$Y = m(X) + \sigma(X)\, e\,, \tag{2.1}$$

where $m(x) = \mathrm{E}(Y|X = x)$ is the regression function, $\sigma^2(x) = \mathrm{Var}(Y|X = x)$ is the conditional variance function, and e is a random error independent of X with mean zero and variance one. Suppose the data $(X_i, Y_i), i = 1, \ldots, n$, are observed from model (2.1). The local linear regression estimator of $m(x_0)$ at a given point x_0 is obtained by minimizing the objective function

$$\sum_{i=1}^{n} \{Y_i - a - b(X_i - x_0)\}^2 K_h(X_i - x_0)\,, \tag{2.2}$$

where K is a kernel function, h is a bandwidth, and $K_h(u) = K(u/h)/h$. Letting $\widehat{a}_h(x_0)$ and $\widehat{b}_h(x_0)$ denote the minimizer of the objective function (2.2), we have

$$\widehat{a}_h(x_0) = \frac{T_{n,0}S_{n,2} - T_{n,1}S_{n,1}}{S_{n,2}S_{n,0} - S_{n,1}S_{n,1}}\,, \quad \text{and} \quad \widehat{b}_h(x_0) = \frac{T_{n,1}S_{n,0} - T_{n,0}S_{n,1}}{S_{n,2}S_{n,0} - S_{n,1}S_{n,1}}\,,$$

where $S_{n,l} = \sum_{i=1}^{n} K_h(X_i - x_0)(X_i - x_0)^l$, $l = 0, 1, 2$, and $T_{n,l} = \sum_{i=1}^{n} K_h(X_i - x_0)(X_i - x_0)^l Y_i$, $l = 0, 1$. Then the local linear estimator of $m(x_0)$, denoted by $\widehat{m}_h(x_0)$, is defined as $\widehat{m}_h(x_0) = \widehat{a}_h(x_0)$.

Given a sequence of bandwidths $h_1, \ldots, h_B$, we denote the respective local linear estimators of $m(x_0)$ by $V_1 \equiv \widehat{m}_{h_1}(x_0), \ldots, V_B \equiv \widehat{m}_{h_B}(x_0)$. From Fan and Gijbels (1996) we have the following asymptotic properties for these estimators if the bandwidths $h_i, i = 1, \ldots, B$, satisfy some general regularity conditions:

$$\mathrm{Bias}\{\,\widehat{m}_h(x_0)|\,\mathbb{X}\} = \frac{1}{2}\mu_2 m^{(2)}(x_0)h^2 + o_p(h^2),$$
$$\mathrm{Var}\{\,\widehat{m}_h(x_0)|\,\mathbb{X}\} = \nu_0 \frac{\sigma^2(x_0)}{f(x_0)nh} + o_p\left(\frac{1}{nh}\right),$$

where $\mu_j = \int_{-\infty}^{\infty} u^j K(u)du$, $\nu_0 = \int_{-\infty}^{\infty} K(u)^2 du$ and f is the density function of the predictor X. Hence for any given x_0, we can consider a linear regression model for the pairs (V_i, h_i^2), $i = 1, \ldots, B$:

$$V_i \approx \alpha + \beta\, h_i^2 + \tilde{\sigma}(h_i)\, \varepsilon_i\,, i = 1, \ldots, B\,, \tag{2.3}$$

where $\tilde{\sigma}^2(h_i) = \mathrm{Var}(V_i|h_i^2) = \mathrm{Var}\{\widehat{m}_{h_i}(x_0)\}$, ε is independent of h_i^2 with $\mathrm{E}(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = 1$. The least squares estimators of α and β in (2.3) are

$$\widehat{\alpha}_B = \sum_{i=1}^{B} \left\{ \frac{\sum_{k=1}^{B} h_k^4 - \big(\sum_{k=1}^{B} h_k^2\big)\cdot h_i^2}{B\sum_{k=1}^{B} h_k^4 - \big(\sum_{k=1}^{B} h_k^2\big)^2} \right\} V_i\,, \quad \text{and}$$

$$\widehat{\beta}_B = \sum_{i=1}^{B} \left\{ \frac{B\cdot h_i^2 - \big(\sum_{k=1}^{B} h_k^2\big)}{B\sum_{k=1}^{B} h_k^4 - \big(\sum_{k=1}^{B} h_k^2\big)^2} \right\} V_i\,.$$

Then $\widehat{\beta}_B$ is an estimator of $(1/2)\mu_2 m^{(2)}(x_0)$ and $\widehat{\beta}_B h_i^2$ is an estimator of the asymptotic bias of V_i. Therefore, $\widehat{\alpha}_B$ is a bias-reduced estimator for $m(x_0)$. Denote it by $\widetilde{m}_B(x_0)$ and write

$$\widetilde{m}_B(x_0) = \widehat{\alpha}_B = \sum_{i=1}^{B} \boldsymbol{g}_i V_i\,,$$

where

$$\boldsymbol{g}_i = \frac{\sum_{k=1}^{B} h_k^4 - \big(\sum_{k=1}^{B} h_k^2\big)\cdot h_i^2}{B\sum_{k=1}^{B} h_k^4 - \big(\sum_{k=1}^{B} h_k^2\big)^2},\ i = 1, \ldots, B.$$

Since the linear regression model (2.3) has heterogenous error variance, α and β can be estimated efficiently by weighted least squares with weights $h_i, i = 1, \ldots, B$. This yields another set of estimators for α, β and $m(x_0)$:

$$\widehat{\alpha}_{WB} = \sum_{i=1}^{B} \left\{ \frac{h_i \sum_{k=1}^{B} h_k^5 - \big(\sum_{k=1}^{B} h_k^3\big)\cdot h_i^3}{\sum_{k=1}^{B} h_k \sum_{k=1}^{B} h_k^5 - \big(\sum_{k=1}^{B} h_k^3\big)^2} \right\} V_i\,,$$

$$\widehat{\beta}_{WB} = \sum_{i=1}^{B} \left\{ \frac{\big(\sum_{k=1}^{B} h_k\big)\cdot h_i^3 - h_i\big(\sum_{k=1}^{B} h_k^3\big)}{\sum_{k=1}^{B} h_k \sum_{k=1}^{B} h_k^5 - \big(\sum_{k=1}^{B} h_k^3\big)^2} \right\} V_i\,,$$

$$\widetilde{m}_{WB}(x_0) = \widehat{\alpha}_{WB} = \sum_{i=1}^{B} \boldsymbol{g}_{wi} V_i\,,$$

where

$$\boldsymbol{g}_{wi} = \frac{h_i \sum_{k=1}^{B} h_k^5 - \big(\sum_{k=1}^{B} h_k^3\big)\cdot h_i^3}{\sum_{k=1}^{B} h_k \sum_{k=1}^{B} h_k^5 - \big(\sum_{k=1}^{B} h_k^3\big)^2},\ \ i = 1, \ldots, B.$$

Alternatively, as $\widehat{\beta}_B$ and $\widehat{\beta}_{WB}$ both estimate $(1/2)\mu_2 m^{(2)}(x_0)$, we can define two other bias-reduced estimators for $m(x_0)$ as

$$\widehat{m}_B(x_0) \equiv \widehat{m}_h(x_0) - \widehat{\beta}_B h^2\,,$$
$$\widehat{m}_{WB}(x_0) \equiv \widehat{m}_h(x_0) - \widehat{\beta}_{WB} h^2.$$

Then, independent of the choices of $h_i = C_i h_0$, $i = 1, \ldots, B$, the asymptotic bias of the new estimator $\widehat{m}_B(x_0)$ is of the order $h^2 h_0^2 + h^4$ because

$$\begin{aligned}\mathrm{E}\{\widehat{m}_B(x_0)\} - m(x_0) &= \mathrm{E}\{\widehat{m}_h(x_0)\} - \mathrm{E}(\widehat{\beta}_B)h^2 - m(x_0)\\ &= O(h^4 + h^2 h_0^2),\end{aligned}$$

In addition, if we take h_0 such that $h = o(h_0)$, the asymptotic variance is exactly the same as that of the local linear estimator $\widehat{m}_h(x_0)$ because

$$\begin{aligned}\mathrm{Var}\{\widehat{m}_B(x_0)\} &= \mathrm{Var}\{\widehat{m}_h(x_0)\} + \mathrm{Var}(\widehat{\beta}_B)h^4 - 2\mathrm{Cov}(\widehat{m}_h(x_0), \widehat{\beta}_B)h^2\\ &= \mathrm{Var}\{\widehat{m}_h(x_0)\} + O\left(\frac{h^4}{nh_0^5} + \frac{h^2}{n\sqrt{hh_0^5}}\right)\\ &= \mathrm{Var}\{\widehat{m}_h(x_0)\} + o\left(\frac{1}{nh}\right).\end{aligned}$$

Similarly, we have

$$\begin{aligned}\mathrm{E}\{\widehat{m}_{WB}(x_0)\} &= m(x_0) + O(h^4 + h^2 h_0^2),\\ \mathrm{Var}\{\widehat{m}_{WB}(x_0)\} &= \mathrm{Var}\{\widehat{m}_h(x_0)\} + o\left(\frac{1}{nh}\right).\end{aligned}$$

Let $\mathbb{X}$ denote $(X_1, \ldots, X_n)^T$. For the estimator $\widetilde{m}_B(x_0)$ we can obtain results similar to those obtained by Lin and Li (2008) and Wu, Liu and Zhou (2013).

Theorem 1. *Under Assumptions* (a)-(e) *in the Supplementary Material,* $h_i = C_i h, i = 1, \ldots, B$, *with* $h \to 0$ *and* $nh \to \infty$ *as* $n \to \infty$, *we have*

$$\mathrm{E}\{\,\widetilde{m}_B(x_0)|\,\mathbb{X}\} = m(x_0) + \boldsymbol{C}(x_0)h^4 + o_p(h^4)\,,$$
$$\mathrm{Var}\,\{\,\widetilde{m}_B(x_0)|\,\mathbb{X}\} = \frac{\sigma^2(x_0)}{nf(x_0)}\sum_{i=1}^{B}\sum_{j=1}^{B}\boldsymbol{g}_i\boldsymbol{g}_j\left\{\psi_{ij}^{(0)} + o_p(1)\right\},$$

where

$$\boldsymbol{C}(x_0) = \frac{1}{2}\boldsymbol{d}(x_0)\left\{\frac{1}{12}m^{(4)}(x_0) - m^{(2)}(x_0)\boldsymbol{b}(x_0)\right\}\mu_4,$$
$$\boldsymbol{d}(x_0) = \frac{\sum_{k=1}^B C_k^4 \sum_{i=1}^B C_i^4 - \sum_{k=1}^B C_k^2 \sum_{i=1}^B C_i^6}{B\sum_{k=1}^B C_k^4 - \left(\sum_{k=1}^B C_k^2\right)^2}, \quad \textit{and}$$
$$\psi_{ij}^{(k)} = \int K(h_i u)K(h_j u)u^k du.$$

Bias reduction estimators similar to $\widetilde{m}_B(x_0)$ and $\widetilde{m}_{WB}(x_0)$ have been investigated by Lin and Li (2008) and Wu, Liu and Zhou (2013). We give more discussion and study of the properties of $\widetilde{m}_B(x_0)$ and $\widetilde{m}_{WB}(x_0)$ and their applications to further statistical inference, and we compare them with $\widehat{m}_B(x_0)$ and $\widehat{m}_{WB}(x_0)$.

Remark 1. By Theorem 1, we know that the bias of $\widetilde{m}_B(x_0)$ is of the order h^4. Even when $B = 2$ we can get such a bias-reduced estimator for $m(x_0)$, although the variance could be larger with a smaller value of B compared to using a larger value of B. As for $\widetilde{m}_{WB}(x_0)$, it has similar properties as $\widetilde{m}_B(x_0)$. Compared with $\widehat{m}_B(x_0)$ and $\widehat{m}_{WB}(x_0)$, their asymptotic variances are slightly more complicated, making them more dependent on bootstrap methods in applications.

Remark 2. As shown in the Supplementary Material, the variance of $\widetilde{m}_B(x_0)$ is no larger than

$$\left(\sum_{i=1}^{B} \boldsymbol{g}_i^2\right)\left\{\sum_{i=1}^{n} \mathrm{Var}(V_i)\right\} = \frac{\sigma^2(x_0)}{nhf(x_0)} \frac{\sum_{i=1}^{B} C_i^4}{B\sum_{i=1}^{B} C_i^4 - (\sum_{i=1}^{B} C_i^2)^2} \left(\sum_{i=1}^{B} \frac{1}{C_i}\right).$$

If we let $B = 3$ and $C_1 = 1, C_2 = 2, C_3 = 3$, then the variance of $\widetilde{m}_B(x_0)$ is no larger than twice that of $\widehat{m}_h$, and if $B = 6$ and $C_i = i, i = 1, \ldots, 6$, the variance of $\widetilde{m}_B(x_0)$ is close to the variance of $\widehat{m}_h$. Using a larger B and selecting $C_1, \ldots, C_B$ appropriately, the variance of $\widetilde{m}_B(x_0)$ can be even smaller than that of $\widehat{m}_h$, but with smaller bias.

Remark 3. For x_0 near the boundary of the support of the density of X, our proposed method cannot directly reduce the bias from the order h^2 to h^4, but it still has smaller bias than when x_0 is in the interior region. Refer to the details of the automatic boundary carpentry property of the local linear regression in Fan and Gijbels (1996). To reduce the order of the bias when x_0 is in the boundary region, take the design density of X to have support $[0, 1]$ and $x_0 = 0$ for example. Then, consider the finer regression model for bias reduction,

$$V_i = \alpha + \beta\, \frac{\mu_{c_i}^2 - \mu_{1,c_i}\mu_{3,c_i}}{\mu_{2,c_i}\mu_{0,c_i} - \mu_{1,c_i}^2} h_i^2 + \tilde{\sigma}(h_i)\varepsilon_i\,, i = 1, \ldots, B,$$

where $\mu_{j,c_i} = \int_{-c_i}^{\infty} u^j K(u)du$, and $c_i = x_0/h_i, i = 1, \ldots, B$.

Remark 4. From Theorem 1, $\widehat{m}_B(x_0)$ and $\widehat{m}_{WB}(x_0)$ have the same asymptotic variance as the local linear estimator, but with smaller asymptotic biases. Use of larger bandwidths to estimate the bias of the local linear estimator is similar to that of Xia (1998) and Fan and Zhang (2000), but we do not use higher-order local polynomial regression, hence we avoid their design sparseness and complicated bandwidth selection problems. While the variance of $\widehat{\beta}_B$ is still somehow large, as shown by our numerical study, it is expected to be more stable than other bias estimation methods.

3. Extensions and Applications

Our bias reduction methods can be easily extended to additive models and varying coefficient models. We use the varying coefficient models as an example to illustrate how to make this extension.

In addition, under the classical nonparametric regression model, we discuss how to use the bias reduction estimators to estimate the variance function of the model error, and how to use the bias reduction estimators to construct simultaneous confidence bands for the nonparametric function in the model.

3.1. Extension to varying coefficient models

Consider the varying-coefficient model

$$Y_i = \sum_{j=1}^{p} a_j(U_i)X_{ij} + \varepsilon_i, i = 1, \ldots, n$$

with

$$\mathrm{E}(\varepsilon_i|U_i, X_{i1}, \ldots, X_{ip}) = 0,$$
$$\mathrm{Var}(\varepsilon_i|U_i, X_{i1}, \ldots, X_{ip}) = \sigma^2(U_i).$$

By local linear fitting, we obtain an estimator of $a_1(u)$ as

$$\widehat{a}_{1h}(u) = e_{1,k}^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}Y,$$

where $e_{i,j}$ denotes the unit vector of length j with 1 at position i, $k = 2p$,

$$Y = (Y_1, \ldots, Y_n)^T,\ \mathbf{W} = \mathrm{diag}(K_h(U_1 - u), \ldots, K_h(U_n - u)),$$
$$\mathbf{X} = \begin{pmatrix} X_{11} & X_{11}(U_1 - u) & \ldots & X_{1p} & X_{1p}(U_1 - u) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{n1} & X_{n1}(U_n - u) & \ldots & X_{np} & X_{np}(U_n - u) \end{pmatrix}.$$

Estimators for the other components can be obtained similarly. As shown by Fan and Zhang (1999),

$$\mathrm{Bias}\{\widehat{a}_{1h}(u_0)\} = C_1h^2 + o_p(h^2) \text{ and } \mathrm{Var}\{\widehat{a}_{1h}(u_0)\} = \frac{C_2}{nh}\{1 + o_p(1)\}$$

where C_1 and C_2 are constants that depend only on u_0. Hence given different bandwidths $h_1, \ldots, h_B$, we can construct the simple linear regression

$$\widehat{a}_{1h_i}(u_0) \approx \alpha + \beta h_i^2 + \varepsilon_i,\ i = 1, \ldots, B,$$

and obtain a bias-reduced estimator for $a_1(u_0)$ as

$$\widetilde{a}_{1B}(u_0) = \widehat{\alpha}_B, \quad \widetilde{a}_{1WB}(u_0) = \widehat{\alpha}_{BW},$$
$$\widehat{a}_{1B}(u_0) = \widehat{a}_{1h}(u_0) - \widehat{\beta}_{1B}h^2, \quad \text{and} \quad \widehat{a}_{1WB}(u_0) = \widehat{a}_{1h}(u_0) - \widehat{\beta}_{1WB}h^2.$$

3.2. Variance function estimation

For the model (2.1), $\{nhf(x)\}^{-1}\nu_0\sigma^2(x)$ is the asymptotic variance of the local linear $\widehat{m}_h(x)$, and shown in Section 2, it is also the asymptotic variance of the bias-reduced estimators $\widehat{m}_B(u)$ and $\widehat{m}_{WB}(u)$. Let

$$Y = (Y_1, \dots, Y_n)^T, \ \mathbf{W} = \operatorname{diag}\left(K_h(X_1 - x), \dots, K_h(X_n - x)\right),$$
$$\mathbf{X} = \begin{pmatrix} 1 & \dots & 1 \\ X_1 - x & \dots & X_n - x \end{pmatrix}^T.$$

With local linear fitting, $\sigma^2(x)$ can be estimated by the kernel estimator

$$\widehat{\sigma}^2(x) = \frac{1}{\operatorname{tr}\{\mathbf{W} - \mathbf{W}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\}} \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2 K_h(X_i - x)$$

where

$$\widehat{Y} = (\widehat{Y}_1, \dots, \widehat{Y}_n)^T = \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}Y.$$

Define the squared residuals as $\widehat{r}_i = \{Y_i - \widehat{m}_h(X_i)\}^2, i = 1, \dots, n$. Then the residual-based estimator, denoted by $\widehat{\sigma}_L^2(x) = \widehat{\alpha}$, with the kernel K and bandwidth h_* is obtained by

$$(\widehat{\alpha}, \widehat{\beta}) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} \{\widehat{r}_i - \alpha - \beta(X_i - x)\}^2 K\left(\frac{X_i - x}{h_*}\right).$$

Fan and Yao (1998) have shown that the estimator $\widehat{\sigma}_L^2(x)$ is more efficient than the kernel estimator $\widehat{\sigma}^2(x)$ but, as discussed by Wang et al. (2008), the bias of the local linear estimator of the regression function affects the efficiency of $\widehat{\sigma}^2(x)$ and $\widehat{\sigma}_L^2(x)$. We use the bias-reduced estimators $\widetilde{m}_B(X_i), \widetilde{m}_{WB}(X_i)$, $\widehat{m}_B(X_i)$, and $\widehat{m}_{WB}(X_i)$ to replace $\widehat{m}_h(X_i)$ in the calculation of the squared residuals $\widehat{r}_i$, and get new versions of $\widehat{\sigma}_L^2(x)$, denoted by $\tilde{\sigma}_B^2(x), \tilde{\sigma}_{WB}^2(x)$, $\widehat{\sigma}_B^2(x)$ and $\widehat{\sigma}_{WB}^2(x)$, respectively. These estimators of the error variance function remove the bias effect occurring in the first step of regression estimation, hence are more stable and efficient than $\widehat{\sigma}_L^2(x)$ and $\widehat{\sigma}^2(x)$.

3.3. Simultaneous confidence band

We consider construction of simultaneous confidence bands for $m(\cdot)$ on the interval $[0,1]$ using the bias-reduced estimators $\widehat{m}_B(x)$ and $\widetilde{m}_B(x)$.

Theorem 2. *Under Assumptions* (a)-(e) *in the Supplementary Material,* $h = n^{-b}, 1/(2q+3) \le b < 1-2/s$*,* $h = o(h_0)$ *with* $h_0 \to 0$ *and* $nh_0 \to \infty$ *as* $n \to \infty$*, and* $h_i = C_i h_0, i = 1, \ldots, B$*, we have*

$$\mathrm{P}\left((-2\log h)^{1/2}\left[\nu_0^{-1/2}\big|\left\{nh\sigma^{-2}(x)f(x)\right\}^{1/2}\{\widehat{m}_B(x)-m(x)\}\big|_\infty - d_n\right] < u\right) \to \exp(-2e^{-u})$$

where

$$d_n = (-2\log h)^{1/2} + \frac{1}{(-2\log h)^{1/2}}\left\{\log\frac{K^2(A)}{\nu_0\pi^{1/2}} + \frac{1}{2}\log\log h^{-1}\right\}$$

if Assumption (e1) *in the Supplementary Material holds, and*

$$d_n = (-2\log h)^{1/2} + \frac{1}{(-2\log h)^{1/2}}\log\left(\frac{1}{4\nu_0\pi}\int\{K'(t)\}^2 dt\right),$$

if Assumption (e2) *in the Supplementary Material is satisfied.*

The asymptotic variance of $\widehat{m}_B(x)$ (or $\widehat{m}_h(x)$) can be approximated by $\widehat{\mathrm{Var}}\{\widehat{m}_B(x)\} = \widehat{\mathrm{Var}}\{\widehat{m}_h(x)\} = e_{1,2}^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{W^2}\mathbf{X})(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}e_{1,2}\widehat{\sigma}_*^2(x)$, where $e_{i,j}$ denotes the unit vector of length j with 1 at position i, and $\widehat{\sigma}_*^2(x)$ is a consistent estimator for the error variance $\sigma^2(x)$. Combining this result with Theorem 2 gives a simultaneous confidence band for $m(x)$ on $[0,1]$:

$$\left(\widehat{m}_B(x) - \Delta_{1,\alpha}(x), \widehat{m}_B(x) + \Delta_{1,\alpha}(x)\right),\ x \in [0,1],$$

where

$$\Delta_{1,\alpha}(x) = \left(d_n + [\log 2 - \log 2\{-\log(1-\alpha)\}]\,(-2\log h)^{-1/2}\right)[\widehat{\mathrm{Var}}\{\widehat{m}_B(x)\}]^{1/2}.$$

The probability that the true curve $m(x)$ is covered by the above band is approximately $1-\alpha$.

The bandwidth h_0 can be selected by some plug-in methods for the local linear fitting; this would reduce computational time, and increase the stability of the final result.

To use $\widetilde{m}_B(x)$ to construct simultaneous confidence bands, let $h_i = C_i h, i = 1, \ldots, B$, and define

$$K_1(t) = \sum_{i=1}^{B}\frac{\boldsymbol{g}_i K(t/C_i)}{C_i} \quad \text{and} \quad \nu_{1,0} = \int K_1^2(t)dt.$$

Theorem 3. *Under Assumptions* (a)-(d) *and* (e2) *in the Supplementary Material, with* $h = n^{-b}, 1/(2q+3) \le b < 1-2/s$*, we have*

$$\mathrm{P}\left((-2\log h)^{1/2}\left[\nu_{1,0}^{-1/2}\left|\left\{nh\sigma^{-2}(x)f(x)\right\}^{1/2}\{\widetilde{m}_B(x)-m(x)\}\right|_\infty - d_n\right] < u\right)$$
$$\to \exp(-2e^{-u}),$$

where

$$d_n = (-2\log h)^{1/2} + \frac{1}{(-2\log h)^{1/2}}\log\left(\frac{1}{4\nu_{1,0}\pi}\int\{K_1'(t)\}^2 dt\right).$$

As shown in the proof of Theorem 1 given in the Supplementary Material, $\{nhf(x)\}^{-1}\nu_{1,0}\sigma(x)$ is the asymptotic variance of $\widetilde{m}_B(x)$. Let

$$\mathbf{W}_i = \mathrm{diag}\left(K_{h_i}(X_1-x),\ldots,K_{h_i}(X_n-x)\right), i=1,\ldots,B.$$

Then the asymptotic variance of the estimator $\widetilde{m}_B(x)$ can be approximated by

$$\widehat{\mathrm{Var}}\{\widetilde{m}_B(x)\} = \sum_{i,j=1}^{B} \boldsymbol{g}_i\boldsymbol{g}_j e_{1,2}^T(\mathbf{X}^T\mathbf{W}_i\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{W}_i\mathbf{W}_j\mathbf{X})(\mathbf{X}^T\mathbf{W}_j\mathbf{X})^{-1}e_{1,2}\widehat{\sigma}_*^2(x),$$

where $\widehat{\sigma}_*^2(x)$ is a consistent estimator of $\sigma^2(x)$. Then an approximate $1-\alpha$ simultaneous confidence band of $m(x)$ on $[0,1]$ can be constructed as

$$\left(\widetilde{m}_B(x)-\Delta_{1,\alpha}(x), \widetilde{m}_B(x)+\Delta_{1,\alpha}(x)\right),\ x\in[0,1],$$

where

$$\Delta_{1,\alpha}(x) = \left(d_n + [\log 2 - \log 2\{-\log(1-\alpha)\}]\,(-2\log h)^{-1/2}\right)[\widehat{\mathrm{Var}}\{\widetilde{m}_B(x)\}]^{1/2}.$$

Although $\widetilde{m}_B(x)$ is a bias-reduced estimator, its asymptotic variance is not easy to estimate stably. And the bootstrap can be used to estimate the variance of $\widetilde{m}_B(x)$. Given $\widetilde{m}_B(x)$ and $\widehat{\sigma}_*^2(x)$, simulate $\epsilon_i^*, i=1,\ldots,n$, from the standard normal distribution $\mathcal{N}(0,1)$, and construct a bootstrap sample as

$$Y_i^* = \widetilde{m}_B(X_i) + \widehat{\sigma}_*(X_i)\epsilon_i^*,\ i=1,\ldots,n.$$

Use the bias-reduction method with the same bandwidth series to estimate $m(x)$ based on the bootstrap sample. Repeat the above procedure T times to get $\widetilde{m}_{B_1}(x),\ldots,\widetilde{m}_{B_T}(x)$, and use their sample variance as an estimator of the variance of $\widetilde{m}_B(x)$. Then we have an approximate $1-\alpha$ simultaneous confidence band if we use this bootstrap variance estimator in $\Delta_{1,\alpha}(x)$.

Similarly, we can also use $\widehat{m}_{WB}(x)$ and $\widetilde{m}_{WB}(x)$ together with their variance approximations to construct simultaneous confidence bands for $m(x)$.

4. Numerical Studies

4.1. Bias reduction

In this section, the finite sample performances of our bias-reduced estimators are investigated under nonparametric regression and varying coefficient models.

Example 1. Consider univariate regressions with constant variance function $\sigma(x)$, $m(x)$ specified as in Fan and Gijbels (1996) and He and Huang (2009) and the covariate X uniform over $[-2, 2]$. For our models $n = 200$ and 1,000 replicates were simulated.

(1) $m(x) = x + 2e^{-16x^2}$ with $\sigma = 0.4$;

(2) $m(x) = \sin(2x) + 2e^{-16x^2}$ with $\sigma = 0.3$;

(3) $m(x) = 0.3e^{-4(x+1)^2} + 0.7e^{-16(x-1)^2}$ with $\sigma = 0.1$;

(4) $m(x) = 0.4x + 1$ with $\sigma = 0.15$.

Figure 1 depicts these models for the local linear estimate $\widehat{m}(x)$ and the bias-reduced estimates $\widetilde{m}_B$ and $\widehat{m}_B(x)$ on one realization.

We used the functions *thumbBw* and *locpol* in the R package *locpol* to get the rule-of-thumb bandwidth, denoted by h_{opt}, and the optimal local linear estimate $\widehat{m}(x)$ based on h_{opt}. We used the bandwidth series $h_i = (1 + C_i/10)h_{\text{opt}}, C_i = -5, \ldots, 5$ to construct $\widetilde{m}_B(x)$ and $\widetilde{m}_{BW}(x)$. The bandwidth series $h_i = (1 + C_i/10)h_{\text{opt}}n^{1/5}, C_i = 0, 1, \ldots, 10$, was used to obtain $\widehat{m}_B(x)$ and $\widehat{m}_{BW}(x)$. The mean square error (MSE), defined as the average squared error over a set of equally-spaced grid points, was used to evaluate performance.

For the first three models, Table 1 shows that the MSE of $\widetilde{m}_B(x)$ and $\widetilde{m}_{BW}(x)$ are much smaller than that of the local linear estimator (LL). From Figure 1, we can also see the biases of $\widetilde{m}_B(x)$ and $\widetilde{m}_{BW}(x)$ are much smaller in the convex or concave places of the true regression curve. The MSEs of $\widehat{m}_B(x)$ and $\widehat{m}_{BW}(x)$ are comparable to the MSE of the local linear estimator, possibly because the bandwidth series used to construct $\widehat{m}_B(x)$ and $\widehat{m}_{BW}(x)$ was relatively large. For Model 4, the local linear estimator performs better than the bias-reduced estimators; this is reasonable because the local linear estimate is almost unbiased as the true regression function is linear.

Following a referee's suggestion, we compared our bias reduction methods with the twicing local linear kernel regression smoother (TLL) and the local cubic smoother (LC). The twicing local linear kernel regression smoothers with different bandwidth selection methods, denoted by TLL1 and TLL2, were proposed and investigated by Zhang and Xia (2012). We considered Model 3 with $\sigma = 0.2$ or 0.5 and $n = 100$, 200, or 400. As before, we used the functions *thumbBw* and *locpol* in R package *locpol* to get the optimal bandwidth h_{opt} for the local linear estimate. We then used the bandwidth series

Table 1. Median (median absolute deviation) ×1,000 of mean square error for models 1-4 in Example 1.

	LL	$\widetilde{m}_B(x)$	$\widetilde{m}_{BW}(x)$	$\widehat{m}_B(x)$	$\widehat{m}_{BW}(x)$
Model 1	18.7785 (5.0014)	13.4333 (4.3978)	12.7000 (4.1420)	18.7035 (5.0656)	18.7100 (5.0854)
Model 2	11.0121 (2.9765)	9.2218 (2.6248)	8.3234 (2.5929)	10.7091 (2.8912)	10.7219 (2.8775)
Model 3	1.3469 (0.3701)	1.0199 (0.3033)	0.9377 (0.2815)	1.3291 (0.3642)	1.3290 (0.3665)
Model 4	0.5257 (0.3331)	0.8801 (0.4809)	0.8010 (0.4541)	0.5341 (0.3406)	0.5330 (0.3408)

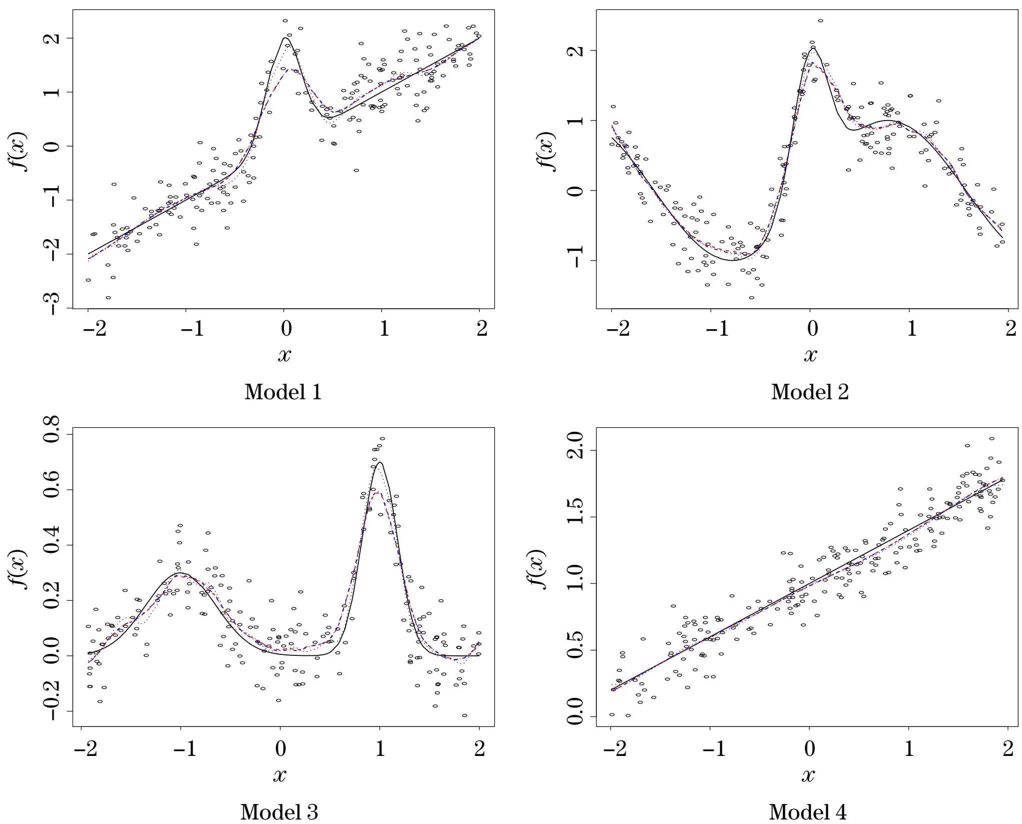


Figure 1. Example 1, Models 1-4. True regression function (solid line), and the local linear estimate $\widehat{m}(x)$ (dash line) and the bias-reduced estimate $\widetilde{m}_B(x)$ (dotted line) and $\widehat{m}_B(x)$ (dot-dash line) based on one realization (circles).

$h_i = (1 + C_i/10)h_{\text{opt}}, C_i = 0, 1, \ldots, 10$, to construct $\widetilde{m}_B(x)$ and $\widetilde{m}_{BW}(x)$. The bandwidth series $h_i = (1 + C_i/10)h_{\text{opt}}n^{1/5}, C_i = 0, 1, \ldots, 10$ was used to obtain $\widehat{m}_B(x)$ and $\widehat{m}_{BW}(x)$. The numerical results are shown in Table 2. In Table 2 our methods perform much better than the local cubic smoother. Compared to the twicing local linear estimators, $\widetilde{m}_B(x)$ and $\widetilde{m}_{BW}(x)$ are always perferable,

Table 2. Mean (standard deviation) ×1,000 of mean square error for model 2 in Example 1 with $\sigma = 0.2$ or 0.5.

σ	n	$\widetilde{m}_B(x)$	$\widetilde{m}_{BW}(x)$	$\widehat{m}_B(x)$	$\widehat{m}_{BW}(x)$	LL	TLL1	TLL2	LC
0.2	100	9.1 (3.3)	9.8 (3.5)	11.4 (3.5)	11.4 (3.5)	12.9 (13.1)	11.5 (10.8)	11.2 (5.6)	25.8 (70.3)
	200	4.8 (1.5)	5.2 (1.6)	7.3 (1.7)	7.4 (1.7)	6.1 (2.5)	5.1 (2.2)	5.3 (2.1)	7.9 (13.7)
	400	2.5 (0.7)	2.7 (0.8)	4.6 (1.0)	4.6 (1.0)	3.2 (0.9)	2.6 (0.8)	2.7 (0.9)	2.9 (0.9)
0.5	100	34.8 (13.1)	35.2 (13.1)	35.0 (12.4)	35.1 (12.4)	48.7 (24.7)	47.5 (20.9)	50.5 (22.3)	74.3 (84.2)
	200	18.5 (6.3)	18.9 (6.3)	20.4 (6.4)	20.4 (6.4)	25.5 (10.8)	24.1 (9.7)	24.6 (9.4)	32.3 (34.1)
	400	10.6 (3.7)	10.8 (3.7)	12.7 (4.0)	12.7 (4.0)	13.4 (4.7)	12.4 (4.3)	12.4 (4.3)	14.4 (8.2)

and $\widehat{m}_B(x)$ and $\widehat{m}_{BW}(x)$ are comparable and sometimes better. Our methods are more stable than the twicing local linear kernel methods as they are quite robust to the bandwidth choice. This difference may be explained by the fact that twicing kernel smoothers use only one local linear estimator (based on one bandwidth) and the bias estimator uses the same bandwidth while our estimators are based on multiple local linear estimators with different bandwidths.

Example 2. With $n = 500$ and 100 replicate samples, we considered two varying coefficient models studied by Fan and Zhang (1999, 2000).

(1) $Y = \sin(6\pi U)X_1 + \sin(2\pi U)X_2 + \varepsilon$;

(2) $Y = \sin(2\pi U)X_1 + 4U(1-U)X_2 + \varepsilon$;

where U was uniform on $[0, 1]$, and X_1 and X_2 were standard normal with correlation coefficient $2^{-1/2}$. We took ε, U, and (X_1, X_2) as independent. The random error ε was normal with mean zero and variance σ^2. The variance σ^2 was chosen so that the signal-to-noise ratio was about 5 : 1,

$$\sigma^2 = 0.2\mathrm{Var}\{m(U, X_1, X_2)\}, \text{ with } m(U, X_1, X_2) = \mathrm{E}(Y|U, X_1, X_2).$$

Given original bandwidths $h_o = 0.05, 0.075, 0.15, 0.225, 0.3$, we calculated the local linear estimates for the models (1) and (2). Then based on the bandwidth series $h_i = (1+C_i/10)h_o, C_i = -5, \ldots, 5$, we obtained the bias-reduced estimates $\tilde{a}_B(u)$, and the bandwidth series $h_i = (1 + C_i/10)h_o n^{1/5}, C_i = 0, 1, \ldots, 10$, was used to calculate the estimate $\widehat{a}_B(u)$. The numerical results are summarized by Tables 3 and 4 for the models 1 and 2 respectively.

From Figure 2, one can see that the varying coefficient functions in the first model are more occillatory than those in the second model. From Tables 3 and 4, compared to the local linear estimator the MSE of the estimator $\widetilde{a}_B(u)$ for the varying coefficient function is much smaller when the bandwidth h_o is relative large. The performance of $\widehat{a}_B(u)$ is close that of the local linear estimator, probably because the bandwidth series used for $\widehat{a}_B(u)$ is relatively large. When

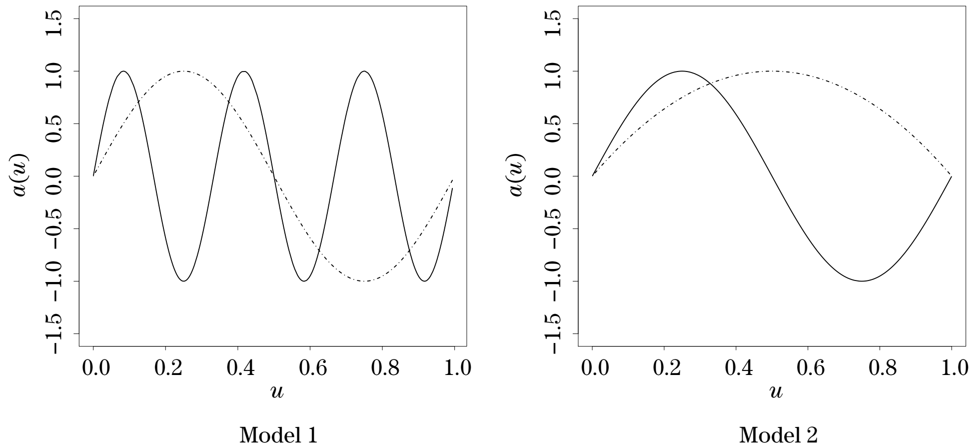


Figure 2. Varying coefficient regression models 1 (left panel) and 2 (right panel) in Example 2. Solid line: varying coefficient function for X_1; Dash line: varying coefficient function for X_2.

Table 3. Median (median absolute deviation) of mean square error for varying coefficient model 1 in Example 2.

		LL	$\tilde{a}_B(x)$	$\widehat{a}_B(x)$
$h_o = 0.300$	$\alpha_1(u)$	0.3303 (0.0362)	0.1885 (0.0246)	0.3305 (0.0362)
	$\alpha_2(u)$	0.0328 (0.0142)	0.0067 (0.0045)	0.0328 (0.0140)
$h_o = 0.225$	$\alpha_1(u)$	0.2648 (0.0223)	0.0854 (0.0125)	0.2647 (0.0213)
	$\alpha_2(u)$	0.0172 (0.0091)	0.0031 (0.0021)	0.0171 (0.0090)
$h_o = 0.150$	$\alpha_1(u)$	0.1243 (0.0171)	0.0134 (0.0181)	0.1237 (0.0169)
	$\alpha_2(u)$	0.0044 (0.0029)	0.0036 (0.0021)	0.0043 (0.0030)
$h_o = 0.075$	$\alpha_1(u)$	0.0143 (0.0049)	0.0062 (0.0022)	0.0138 (0.0045)
	$\alpha_2(u)$	0.0037 (0.0019)	0.0062 (0.0027)	0.0039 (0.0021)
$h_o = 0.050$	$\alpha_1(u)$	0.0066 (0.0032)	0.0096 (0.0040)	0.0064 (0.0031)
	$\alpha_2(u)$	0.0051 (0.0025)	0.0094 (0.0042)	0.0050 (0.0026)

the bandwidth h_0 is small, the bias of the local linear estimator is relatively small, and the bias-reduced estimators have little advantage, or even overestimate. In addition, when there is more than one varying coefficient function to estimate, bias reduction may not be simultaneously achieved. The one-step estimation procedure proposed by Fan and Zhang (1999) can be considered; this is outside the scope of this paper.

We have the following conclusions. When the signal-to-noise ratio or the variation of the regression function is relative large, $\widetilde{m}_B(x)$ and $\widetilde{a}_B(u)$ are more appropriate for the bias reduction. When the signal-to-noise ratio is small or the

Table 4. Median (median absolute deviation) ×100 of mean square error for varying coefficient model 2 in Example 2.

		LL	$\widetilde{a}_B(x)$	$\widehat{a}_B(x)$
$h_o = 0.300$	$\alpha_1(u)$	3.1044 (0.7985)	0.4716 (0.3594)	2.5782 (0.9184)
	$\alpha_2(u)$	0.3553 (0.2964)	0.1743 (0.1586)	0.2145 (0.1718)
$h_o = 0.225$	$\alpha_1(u)$	1.3474 (0.5038)	0.2398 (0.1578)	0.6000 (0.4281)
	$\alpha_2(u)$	0.2071 (0.1850)	0.2332 (0.2000)	0.1914 (0.1665)
$h_o = 0.150$	$\alpha_1(u)$	0.4527 (0.3243)	0.3177 (0.1981)	0.2546 (0.1733)
	$\alpha_2(u)$	0.2175 (0.1465)	0.3386 (0.2134)	0.2389 (0.1725)
$h_o = 0.075$	$\alpha_1(u)$	0.3630 (0.2116)	0.5602 (0.1773)	0.4091 (0.1951)
	$\alpha_2(u)$	0.3572 (0.1499)	0.5563 (0.2009)	0.4316 (0.1547)
$h_o = 0.050$	$\alpha_1(u)$	0.5609 (0.2893)	1.0152 (0.4015)	0.5552 (0.2977)
	$\alpha_2(u)$	0.5548 (0.2469)	1.1153 (0.3784)	0.5485 (0.2250)

estimated function is smooth, the performance of $\hat{m}_B(x)$ and $\hat{a}_B(u)$ are much better than the performance of $\widetilde{m}_B(x)$ and $\widetilde{a}_B(u)$, respectively.

4.2. Variance function estimation

We used Example 2 of Fan and Yao (1998) to assess the performance of our variance function estimates.

Example 3. We simulated 400 random samples of size $n = 200$ from the model

$$Y_i = a\{X_i + 2\exp(-16X_i^2)\} + \sigma(X_i)e_i,$$

with $\sigma(x) = 0.4\exp(-2x^2) + 0.2$, where $\{X_i\}$ and $\{e_i\}$ were independent, with $X_i \sim \text{Uniform}[-2, 2]$ and $e_i \sim N(0, 1)$. We took $a = 0.5, 1, 2, 4$ in the simulation. For each simulated sample, the performance of an estimator $\widehat{\sigma}(\cdot)$ was evaluated by the mean absolute deviation error,

$$\mathcal{E}_{\text{MAD}} = n_{\text{grid}}^{-1} \sum_{j=1}^{n_{\text{grid}}} |\widehat{\sigma}(x_j) - \sigma(x_j)|,$$

where $\{x_j, j = 1, \ldots, n_{\text{grid}}\}$ are grid points on $[-1.8, 1.8]$ with $n_{\text{grid}} = 101$.

The numerical results for the variance function estimator of Fan and Yao (1998) and the proposed estimators are shown in Table 5. From Table 5 one can see that $\widetilde{\sigma}_B^2(x)$, the variance function estimator based on $\widetilde{m}_B(x)$, always outperforms the local linear estimator of the variance function of Fan and Yao (1998). Especially, its performance improves as a becomes larger, which is reasonable because the sign-to-noise ratio increases as a increases, as shown by Figure 3(a). As for $\widehat{\sigma}_B^2(x)$, the variance function estimator based on $\widehat{m}_B(x)$, it outperforms their estimator when a is small. From the numerical results of Example 1, the

Table 5. Median (median absolute deviation) ×100 of mean absolute deviation error for Example 3.

	$\widetilde{\sigma}^2_B(x)$	$\widehat{\sigma}^2_B(x)$	Fan and Yao (1998)
$a = 4$	3.3732 (1.0997)	9.9722 (2.8180)	3.7587 (1.3658)
$a = 2$	3.5580 (1.2847)	3.8548 (1.3814)	3.6814 (1.2727)
$a = 1$	3.4945 (1.1850)	3.0713 (0.9096)	3.6653 (1.0994)
$a = 0.5$	3.4596 (1.0887)	3.0365 (1.0202)	3.5635 (1.15539)

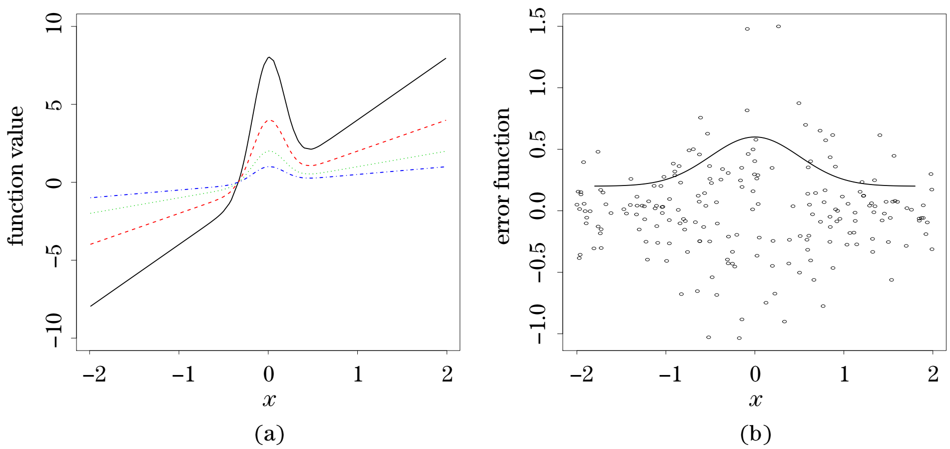


Figure 3. Regression model in Example 3. Panel (a): Regression functions. Solid line: $a = 4$, Dash line: $a = 2$, Dot line: $a = 1$, Dot-Dash line $a = 0.5$. Panel (b): Variance function.

performance of $\widehat{m}_B(x)$ is similar to that of the local linear estimator. Small bias reduction in estimating the regression function can make much improvement in estimating the variance function of the error. This supports the conclusion by Wang et al. (2008) that bias in estimation of the nonparametric regression function seriously affects the efficiency of the estimation for the variance function of the error. On the other hand, when the signal-to-noise ratio increases, the variation of the bias estimate increases so the variance function estimator based on $\widehat{m}_B(x)$ performs worse.

4.3. Simultaneous confidence band

In this section, we consider the example investigated by Eubank and Speckman (1993) and Xia (1998) to assess the performance of their simultaneous confidence bands. The sample size $n = 100$, 200, 300, 500 and 1,000 replicate samples were considered in the simulation.

Table 6. Empirical coverage of simultaneous confidence band based on $\widehat{m}_B(x)$.

σ	Norminal coverage %	$n = 100$	$n = 200$	$n = 300$	$n = 500$
0.05	90	0.895	0.887	0.869	0.870
	95	0.945	0.941	0.947	0.958
0.10	90	0.890	0.897	0.915	0.882
	95	0.954	0.951	0.955	0.961

Example 4. Let

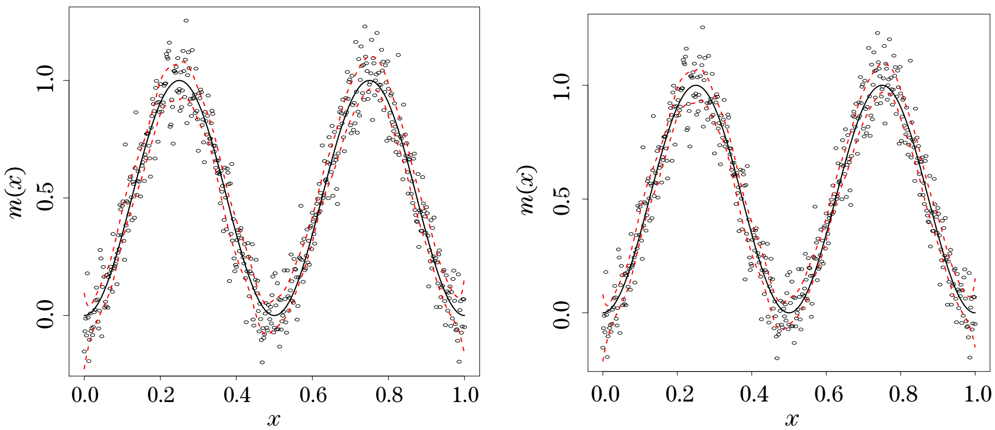
$$Y_i = \sin^2(2\pi(X_i - 0.5)) + \varepsilon_i,$$

where $\varepsilon_i, i = 1, \ldots, n$, are i.i.d. and $N(0, \sigma^2)$, with $\sigma = 0.05$ or 0.1, and $X_i, i = 1, \ldots, n$, follow a fixed design with $X_i = i/n, i = 0, 1, \ldots, n$.

Consider using $\widehat{m}_B(x)$ to construct simultaneous confidence bands. Undersmoothing is necessary to construct the simultaneous confidence bands in practice, we used $h_B = (1/2)h_{opt}$ and chose the bandwidth series $h_i = (1 + C_i/B)h_B n^{1/5}$ to construct $\widehat{m}_B(x)$, where h_{opt} is again the rule-of-thumb bandwidth and $C_i = 0, 1, \ldots, B$ with $B = 10$. Based on $\widehat{m}_B(x)$, and following the procedure given in Section 3.3, we constructed the simultaneous confidence bands for the regression function. We did not consider simultaneous confidence bands based on $\widetilde{m}_B(x)$ as the bootstrap variance estimation required much more computational time. Compared to the procedures of Xia (1998) and Fan and Zhang (2000), we did not need any further steps to reduce the bias of $\widehat{m}_B(x)$ in constructing the simultaneous confidence band. The numerical results shown in Table 6 are comparable or even better than the results given in Table 1 of Xia (1998). Figure 4 illustrates simultaneous confidence bands based on $\widehat{m}_B(x)$ when $n = 500$ and $\sigma = 0.1$. From Figure 4 one sees clearly the benefit of using $\widehat{m}_B(x)$ to construct simultaneous confidence bands.

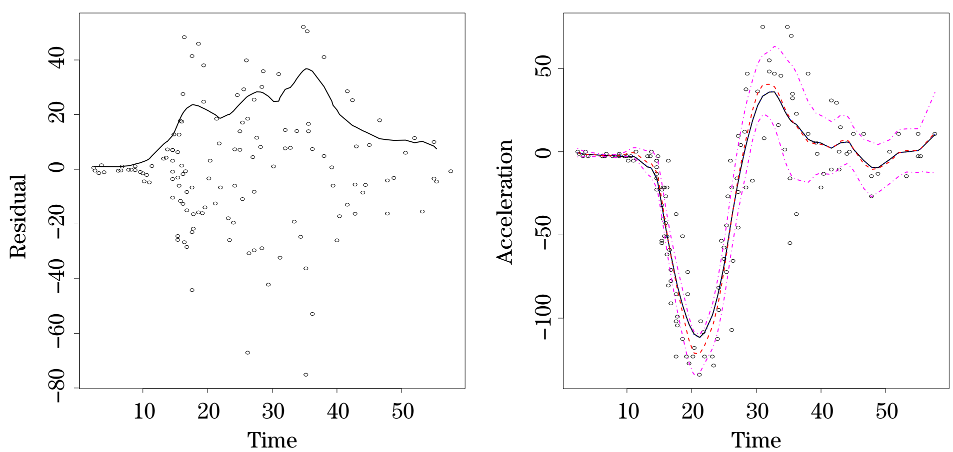
4.4. Example

The motorcycle data set given in Härdle (1990) consists of accelerometer readings taken through time in an experiment on the efficiency of crash helmets. The X-value denotes time (in milliseconds) after a simulated impact with motorcycles. The response variable Y is the head acceleration (in g) of a post mortem human test object (PTMO). The details of the experiment are in Schimdit, Mattern and Schuler (1981). We used similar procedures as before to obtain the local linear estimate, $\widetilde{m}_B$, and $\widehat{m}_B$, then followed our procedure to estimate the variance function of the error since it is obvious that the variation of the data is heterogeneous with time point X. From Figure 5, we can see that the variance



(a) Nominal coverage $1-\alpha = 0.95$ (b) Nominal coverage $1-\alpha = 0.90$

Figure 4. Simultaneous confidence bands for Example 4 with $n = 500$ and $\sigma = 0.1$, Solid line: true regression function $m(x)$; Dash line: bias-reduced estimate $\widehat{m}_B(x)$; Dotted line: simultaneous confidence band.



(a) (b)

Figure 5. Motorcycle data. Panel (a): Residuals and the curve estimate of the error variance function based on the estimate $\tilde{m}_B(x)$. Panel (b): Nonparametric regression function estimates for acceleration versus time. Solid line: the local linear estimate; Dash line: the estimate $\widetilde{m}_B(x)$; Dotted line: the estimate $\widehat{m}_B(x)$; Dot-Dash lines: the simultaneous confidence band based on $\widehat{m}_B(x)$.

function is not a constant and the bias-reduced estimate $\widetilde{m}_B$ reduces much bias when the data is at the extreme points, though $\widetilde{m}_B$ is close to the local linear estimate.

5. Summary

In this paper, in the classical nonparametric regression problem, based on the local linear regression model we investigate two simple bias-reduced estimation approaches from both theoretical insights and numerical studies, and we extend the methods to to the error variance estimation problem and the semiparametric varying coefficient regression model. Our methods avoid using higher-order local polynomial regression to estimate the bias term of the local linear estimator without reducing the efficiency as shown by our theoretical results. From the numerical results, it is obvious that our proposed estimators improve on the local linear estimator to a large extent in terms of efficiency, in particular they reduce the estimation bias by a large amount when the nonparametric functions in the nonparametric regression models or semiparametric models depict much occillation.

Asymptotically unbiased nonparametric function estimation has applications in, for example, construction of simultaneous confidence bands. In our approach we need not use complicated procedures to remove the bias effect of the nonparametric function estimation.

One can improve our bias reduction methods. For example, we know that the choice of the bandwidth series can change the efficiency of the bias-reduced estimators. This is an interesting topic for further investigation. Our bias-reduced estimators have different performances when used in estimating of the variance of the error. It would be interesting to understand their differences.

Supplementary Materials

Technical conditions and proofs of the main theoretical results are here.

Acknowledgment

The authors would like to thank the Editor, an associate editor, and anonymous referees for their helpful comments and suggestions that have helped us to substantially improve the quality of the paper. Ming-Yen Cheng's research was supported in part by the Ministry of Science and Technology grant MOST104-2118-M-002-005-MY3. Tao Huang's research was supported in part by the State Key Program in the Major Research Plan of NSFC (No. 91546202 and No. 11771268) and Program for Innovative Research Team of SHUFE. Heng Peng's research was supported part by CEGR grants of the Research Grant Council of Hong Kong (No. HKBU202012 and No. HKBU12302615) and FRG grants from the Hong Kong Baptist University (No. FRG214-15/064 and No. FRG2/16-17/042).

References

Chen, S. and Qin, Y. S. (2002). Confidence intervals based on local linear smoother. *Scandinavian Journal of Statistics* **29**, 89–99.

Choi, E. and Hall, P. (1998). On Bias reduction in local linear smoothing. *Biometrika* **85**, 333-345.

Choi, E., Hall, P. and Rousson, V. (2000). Data sharpening methods for bias reduction in nonparametric regression. *The Annals of Statistics* **28**, 1339–1355.

Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression. *Journal of the American Statistican Assocation* **88**, 1287–1301.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics* **21**, 196–216.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall: London.

Fan, J. and Yao, Q. Y. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85**, 645–660.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics* **27**, 1491–1518.

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics* **27**, 715–731.

Hall, P. and Horowitz, J. (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *The Annals of Statistics* **41**, 1892–1921.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge, UK.

Hastie, T. and Loader, C. (1993). Local regression: automatic kernel carpentry (with comments). *Statistical Science* **8**, 120–143.

He, H. and Huang, L.-S. (2009). Double-smoothing for bias reduction in local linear regression. *Journal of Statistical Planning and Inference* **139**, 1056–1072.

Li, G. R., Peng, H., Dong, K. and Tong, T. (2014). Simultaneous confidence bands and hypothesis testing for single-index models. *Statistica Sinica* **24**, 937–955.

Lin, L. and Li, F. (2008). Stable and bias-corrected estimation for nonparametric regression models. *Journal of Nonparametric Statistics* **20**, 283–303.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics* **22**, 1346–1370.

Schimdit, G., Mattern, R. and Schuler, F. (1981). Biomechanical investigation to determine physical and trumatolgoical differentiation criteria for the maximum loadcapacity of head and vertebral column with and without protective helmet under effects of impact. *EEC Research Program on Biomechanics of Impact. Final Report Phase III*, Project 65. Institute fur Rechtsmedizin, Unversitat Heidelberg, Germany.

Wang, L., Brown, L. D., Cai, T. T. and Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *The Annals of Statistics* **36**, 646–664.

Wu, T., Liu, S. Y. and Zhou, J. (2013). A semi-parametric estimation for local linear regression model. *Chinese Journal of Engineering Mathematics* **30**, 695–701.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60**, 797–811.

Zhang, W. and Peng, H. (2010). Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. *Journal of Multivariate Analysis* **101**, 1656–1680.

Zhang, W. and Xia, Y. (2012). Twicing local linear regression smoothers. *Journal of Nonparametric Statistics* **24**, 399–417.

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong.

E-mail: chengmingyen@hkbu.edu.hk

Department of Statistics, Shanghai University of Finance and Economics, 777 Guoding Road, Yangpu District, Shanghai 200433, China.

E-mail: huang.tao@mail.shufe.edu.cn

Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, WA 98109-1024, USA.

E-mail: liupeng@amss.ac.cn

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong.

E-mail: hpeng@math.hkbu.edu.hk

(Received January 2017; accepted August 2017)