Statistica Sinica

Bias Reduction for Nonparametric and Semiparametric Regression Models

Ming-Yen Cheng, Tao Huang, Peng Liu, and Heng Peng

Hong Kong Baptist University and National Taiwan University,

Shanghai University of Finance and Economics,

University of Washington and Fred Hutchinson Cancer Research Center,

Hong Kong Baptist University

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 estimates 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. The new approach does not need to use higher-order local polynomial regression to estimate the bias, and hence avoids the double bandwidth selection and the 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 new approach to varying coefficient models, and 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 real data example is used to investigate the performance of the proposed method.

Key words and phrases: Undersmoothing; Variance function estimation; Simultaneous confidence band.

1. Introduction

Nonparametric and semiparametric regression models have been widely used to discover nonlinear relationships between the response variable and predict variables, and reduce the model bias to avoid model misspecification. To give more reasonable interpretation for the model with data, based on the efficient model estimation to make statistical inference is necessary step for data interpretation and finding the insufficiency of the original statistical model. However those classical efficient estimates of those nonparametric components in the nonparametric or semiparametric regression models are always biased. It hinders the development and applications of the statistical inference methods for the nonparametric components in those models. Hence based on those nonparametric or semiparametric regression models, finding an unbiased estimate for the nonparametric components or reducing the bias for those classical efficient nonparametric estimate is very important for further inference.

As an important technique and methodology for nonparametric regression modelling, local linear estimators enjoys many numerical as well as theoretical advantages (see Fan, 1993, Fan and Gijbels, 1996, Hastie and Loader, 1993, etc). For example, it achieves the minimax efficiency among all the linear estimators. Compared with local polynomial modelling (such as local cubic or local quintic), the latter involves more parameters and are subject to design sparsity (Choi, Hall and Rouson, 2000). However, when the nonparametric function shows a high degree of smoothness, local linear approach becomes less appealing compared with high order local polynomial modelling. For example, consider the local cubic estimate, and let n be the sample size and h is the bandwidth used to estimate the nonparametric components in the model, it has a bias of order $O(h^4)$ and variance of order $O((nh)^{-1})$, while the local linear estimate has a bias of order $O(h^2)$ and variance of order $O((nh)^{-1})$ even though the local cubic estimate would suffer sparse problem of data design. To combine the advantages of local linear fitting and local cubic estimate, developing an approach to reduce the bias of the local linear estimate would be an interesting topic and 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 the above literature, there are two kind bias reduction approaches. The first approach is directly using high order local polynomial fitting, such as local cubic regression, to directly estimate the bias (See Xia, 1998, Fan and Zhang, 2000, etc). The other approach is to use information of those closest design points and the idea of the model average to find an approximately unbias estimate (see Choi and Hall, 1998, Choi, Hall and Rousson, 2000, He and Huang,2009, etc). For the first approach, the computational burden is very heavy. It need a special approach to select an appropriate bandwidth for the high order local polynomial fitting. However to improve the accuracy of the estimate of the bias of the local linear estimate, it would need undersmoothing or oversmoothing for high order polynomial fitting. The optimal bandwidth selectors are not appropriate for such high order local polynomial fitting. Furthermore because of small order bandwidth for undersmoothing, the sparsity of the data would increase the variation of the estimate of the bias. Those all hinder the applications of the first bias reduction approach in practice. For the second approach, it avoids to estimate the bias of the local linear estimate, and could directly construct a nearly unbiased nonprarametric function estimate by the idea of the model averaging. Hence there is no problem for the bandwidth selection, and does not increase much computation burden. But it is still possible suffer sparse problem of data design when data is close to the boundary of the support area of the data. More specially, such approach is based on the idea of the model average. Though the new nonparametric function estimate could be nearly unbiased, the asymptotic variance structure of such estimate would be much complex and hinder the application of such unbiased estimate for further statistical inference, and its further extension to more complicated semiparametric or nonparametric regression models.

Fan and Yao (1998) gives a good review for the variance estimate of the nonparametric regression model error based on the sum squares of residuals approach, and suggests a two steps to estimate the variance function of the error in the model. Though Fan and Yao (1998) has shown the proposed estimate of variance function is optimal efficient, as discussion by Wang, Brown, Cai and Levine (2008), the bias of the estimate of the regression function cannot be further reduced in the second stage smoothing of the squared residuals. By the nearly unbiased regression functional estimate, by sum squares of the residuals, it is possible to construct a stable estimate of the variance or variance function of the model error for further inference. Furthermore, Bias reduction estimates have wide applications in statistics. For example, it can be used in constructing a simultaneous confidence band(SCB) for nonparametric component estimates in the model.

Many bias reduction methods have been proposed to cope with the bias term in the estimate of functions to achieve the right coverage probability for SCB. Most methods are based on two simple ideas, one is undersmoothing and the other is oversmoothing. By selecting a small bandwidth to get undersmoothing estimate of the nonparametric components and make the bias small and ignorable (Chen and Qin, 2002, Fan and Zhang, 2000, Zhang and Peng, 2010, Li, Peng, Dong and Tong, 2014). For the other way, by choosing a larger bandwidth (oversmoothing) to make the bias large enough to be identified (Xia, 1998) and then get an accurate estimate for the bias. So the bias can be reduced from the original nonparametric component estimate. Recently, Hall and Horowitz (2013) used a bootstrap method to avoid the bias problem to construct confidence band for nonparametric function estimate. But it is well known that the bootstrap is time consuming and does not provide general finite sample guarantees. For undersmoothing/oversmoothing, effective empirical bandwidth selection criterion is completely unknown and relies on the empirical results and data itself (Hall and Horowitz, 2013). Moreover, when undersmoothing is employed, the confidence band will become wide and oscillate significantly, it also subjects to design sparsity. And if we use oversmoothing, in general, another local cubic estimator is needed to estimate the derivative of the unknown nonparametric function (see, for example Xia, 1998, Fan and Zhang, 2000), However such variance of the bias estimate has the same order as the bias term to be estimated, and hence the bias estimate has low accuracy compared with the function estimate itself (Zhang and Peng, 2010).

In this paper, we propose a new bias reduction technique for local linear estimate which is very easy to implement and extension in practice. To be specific, consider the simple nonparametric regression model

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

where $m(\cdot)$ is an unknown smoothing function and $\varepsilon_i, i = 1, \dots, n$ are independent random errors. Denote the local linear estimate of $m(X_0)$ as $\widehat{m}_h(X_0)$ by a bandwidth h. Our basic idea is to choose different $h_1, \cdots, h_B$ to obtain a series estimate $\{\widehat{m}_{h_i}(X_0), i = 1, \cdots, B\}$ for $m(X_0)$, and then perform a linear regression, where we treat $\widehat{m}_{h_i}(X_0)$ as the dependent variable, and h_i^2 as the explanatory variable. Then the estimate of intercept term can be regarded as one of the proposed 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 for the estimate $\widehat{m}_h(X_0)$, and then based on such estimate of the bias, bias correction can be made to obtain the new asymptotic unbias estimate for $m(X_0)$. This new asymptotic unbias estimate of $m(X_0)$ would have the same order of bias and variance as the local cubic estimate, but the new estimate retains the advantages of local linear estimate, such as avoiding sparsity problem of data design.

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. In this section, the asymptotic properties of the proposed unbias estimates are investigated. In Section 3, we firstly consider the extension of our bias reduction estimates to varying coefficient regression models. Next we investigate how to use the bias reduction estimates to estimate the variance function of the error and how to use the proposed estimates to construct simultaneously confidence bands for the estimate of nonparametric components in the classical nonparametric regression model. In Section 4, some numerical studies with comparison and real data analysis are given. In Section 5, we give some conclusion and discussion about the proposed methods . The proofs of the main results can be found in the Supplement.

2. Bias Reduction

Consider the following 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 following 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 the kernel function, h is the 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 $V_1 \equiv \widehat{m}_{h_1}(x_0), \ldots, V_B \equiv \widehat{m}_{h_B}(x_0)$. It is well known (Fan and Gijbels, 1996) that we have the following asymptotic properties for these estimators if the bandwidths $h_i, i = 1, \ldots, B$, satisfy some general regularity conditions:

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

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 the following 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 model (2.3) are

$$\begin{aligned}\widehat{\alpha}_B &= \sum_{i=1}^{B}\Big[\frac{\sum_{k=1}^{B} h_k^4 - \big(\sum_{k=1}^{B} h_k^2\big)\cdot h_i^2}{B\sum_{k=1}^{B} h_k^4 - \big(\sum_{k=1}^{B} h_k^2\big)^2}\Big]\, V_i\,, \qquad \text{and}\\ \widehat{\beta}_B &= \sum_{i=1}^{B}\Big[\frac{B\cdot h_i^2 - \big(\sum_{k=1}^{B} h_k^2\big)}{B\sum_{k=1}^{B} h_k^4 - \big(\sum_{k=1}^{B} h_k^2\big)^2}\Big]\, V_i\,.\end{aligned}$$

Then $\widehat{\beta}_B$ is an estimator of $\frac{1}{2}\mu_2 m^{(2)}(x_0)$ 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

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

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

$$\begin{aligned} \widehat{\alpha}_{WB} &= \sum_{i=1}^B \Big[\frac{h_i \sum_{k=1}^B h_k^5 - \big(\sum_{k=1}^B h_k^3\big) \cdot h_i^3}{\sum_{k=1}^B h_k \sum_{k=1}^B h_k^5 - \big(\sum_{k=1}^B h_k^3\big)^2}\Big] V_i\,, \\ \widehat{\beta}_{WB} &= \sum_{i=1}^B \Big[\frac{(\sum_{k=1}^B h_k) \cdot h_i^3 - h_i\big(\sum_{k=1}^B h_k^3\big)}{B \sum_{k=1}^B h_k^4 - \big(\sum_{k=1}^B h_k^2\big)^2}\Big] V_i\,. \end{aligned}$$

and

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

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

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

or

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

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 $h = o(h_0)$, then 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 also have

$$\mathrm{E}(\widehat{m}_{WB}(x_0)) - m(x_0) = O(h^4 + h^2 h_0^2),$$

and

$$\mathrm{Var}(\widehat{m}_{WB}(x_0)) = \mathrm{Var}(\widehat{m}_h(x_0)) + o\left(\frac{1}{nh}\right).$$

Let $\mathbb{X}$ denote $(X_1, \ldots, X_n)^T$. For the estimator $\widetilde{m}_B(x_0)$ we can obtain its asymptotic bias and variance stated in the following theorem, which are similar to the results obtained by Wu, Liu and Zhou (2013).

Theorem 1. Under assumptions (a)-(e) given in the Supplement, $h_i = C_i h, i = 1, \ldots, B$, and $h \to 0$ and $nh \to \infty$ as $n \to \infty$, we have

$$\mathrm{E}[\widetilde{m}_B(x_0) - m(x_0)|\,\mathbb{X}] = \boldsymbol{C}(x_0)h^4 + o_p(h^4)$$

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

where

$$\boldsymbol{C}(x_0) = \frac{1}{2}\boldsymbol{d}(x_0)\left[\frac{1}{12}m^{(4)}(x_0) - m^{(2)}(x_0)\boldsymbol{b}(x_0)\right]\mu_4,$$

$$\boldsymbol{d}(x_0) = \frac{\sum_{k=1}^{B} C_k^4 \sum_{i=1}^{B} C_i^4 - \sum_{k=1}^{B} C_k^2 \sum_{i=1}^{B} C_i^6}{B\sum_{k=1}^{B} C_k^4 - \left(\sum_{k=1}^{B} C_k^2\right)^2}, \text{ and}$$

$$\psi_{ij}^{(k)} = \int K(h_i u)K(h_j u)u^k du.$$

In fact, bias reduction estimators similar to $\widetilde{m}_B(x_0)$ and $\widetilde{m}_{WB}(x_0)$ have been investigated by Wu, Liu and Zhou (2013). Nonetheless, we will give more discussion and study on 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 $\widehat{m}_{WB}(x_0)$ in details.

Remark 1. By Theorem 1, we know that the bias of $\widetilde{m}_B(x_0)$ is of the order h^4. Notice that even when $B = 2$ we can also get such 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 also has similar properties as $\widetilde{m}_B(x_0)$. Compared with $\widehat{m}_B(x_0)$ and $\widehat{m}_{WB}(x_0)$, the structure of the asymptotic variances of $\widetilde{m}_B(x_0)$ or $\widetilde{m}_{WB}(x_0)$ are slightly more complicated. Thus it would be much dependent on resampling or bootstrap methods when applying them to further inference problems, and there are more difficulties in its further applications.

Remark 2. By some simple calculation and as shown in the Supplement, it is known that the variance of $\widetilde{m}_B(x_0)$ is no larger than

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

For example, if we let $B = 2$ and $C_1 = 1, C_2 = 2$, then we have the variance of $\widetilde{m}_B(x_0)$ is no larger than three times of that of the local linear estimator $\widehat{m}_h$. If we let $B = 3$ and $C_1 = 1, C_2 = 2, C_3 = 3$, then the variance of $\widetilde{m}_B(x_0)$ will be no larger than two times of that of $\widehat{m}_h$. Furthermore, if $B = 6$ and $C_i = i, i = 1, \ldots, 6$, the variance of $\widetilde{m}_B(x_0)$ will be close to the variance of $\widehat{m}_h$. Furthermore, using a larger B and by selecting $C_1, \ldots, C_B$ appropriately, the variance of $\widetilde{m}_B(x_0)$ could be even smaller than that of $\widehat{m}_h$, but with much smaller bias.

Remark 3. When x_0 is near the boundary of the support of X, although the asymptotic bias of $\widehat{m}_h(x_0)$ changes when we vary the bandwidth h and our proposed method can not directly reduce the bias from the order h^2

to h^4, it still has smaller bias than the case when x_0 is in the interior region. We refer the details to the automatic boundary carpentry property of the local linear regression discussed in Section 3.2.5 of Fan and Gijbels (1996). Hence the problem is not too serious for our method when x_0 is in the boundary region. To reduce the order of the bias when x_0 is at the boundary region, without loss of generality we assume that the design density of X has a bounded support $[0, 1]$ and x_0 is a left boundary point, then we can consider the following much 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 $\mu_{j,c_i} = \int\limits_{-c_i}^{\infty} u^j K(u) du$, 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 much smaller asymptotic biases. Hence it is expected to be easier to apply them to inference problems than using $\widetilde{m}_B(x_0)$ or $\widetilde{m}_{WB}(x_0)$. Especially, although our idea of using larger bandwidths to estimate the bias of the local linear estimator is similar to those suggested by Xia (1998) and Fan and Zhang (2000), we do not use higher-order local polynomial regression as they did. Hence we avoid the design sparse and complicated bandwidth selection problems when using higher-order local polynomial regression. Even though the variance of $\widehat{\beta}_B$ is still somehow large, as shown by our numerical study, it should be more stable than other bias estimation methods.

3. Extensions and Applications

First of all, based on the local polynomial fitting and the idea of model averaging, our bias reduction methods can be easily extended to more complex nonparametric and semiparametric models, such as additive models and varying coefficient models. In the following we use the varying coefficient models as an example to illustrate how to make extension for our bias reduction estimates.

However it would be more important how to apply the bias reduction estimates for further model estimation and inference. Normally, for the second step of the model estimation and inference, as in the discussion of Wang, *et al.* (2008), the variance of the first estimate for the nonparametric components in the model can be easily incorporated, but the bias of the estimates cannot be reduced in the second step model estimation and inference. In this section, under the classical nonparametric regression model, we further discuss how to use the bias reduction estimates to estimate the variance function of the model error, and how to use the bias reduction estimates to construct a simultaneous confidence band for the estimate of 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

$$\mathrm{E}(\varepsilon_i | U_i, X_{i1}, \ldots, X_{ip}) = 0,$$

and

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

As shown by Fan and Zhang (1999), by the local linear fitting, we obtain an estimator of $a_1(u)$ given 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, $k = 2p$,

$$Y = (Y_1, \ldots, Y_n)^T, \ \mathbf{W} = \mathrm{diag}(K_h(U_1 - u), \ldots, K_h(U_n - u)),$$

and

$$\mathbf{X} = \begin{pmatrix} X_{11} & X_{11}(U_1 - u) & \ldots & X_{1p} & X_{1p}(U_1 - u) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{n1} & X_{n1}(U_n - u) & \ldots & X_{np} & X_{np}(U_n - u) \end{pmatrix}$$

Estimator for the other components can be obtained similarly. As shown by Fan and Zhang (1999), we have that

$$\text{Bias}(\widehat{a}_{1h}(u_0)) = C_1 h^2 + o_p(h^2) \text{ and } \text{Var}(\widehat{a}_{1h}(u_0)) = \frac{C_2}{nh}(1 + o_p(1))$$

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

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

and obtain the following bias reduced estimator for $a_1(u_0)$,

$$\widetilde{a}_{1B}(u_0) = \widehat{\alpha}_B, \quad \widetilde{a}_{1WB}(u_0) = \widehat{\alpha}_{BW},$$

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

3.2 Variance Function Estimation

For the model (2.1), it is well known that $(nhf(x))^{-1}\nu_0\sigma^2(x)$ is the asymptotic variance of the local linear $\widehat{m}_h(x)$. As 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

$$Y = (Y_1, \ldots, Y_n), \ \mathbf{W} = \text{diag}\left(K_h(X_1 - x), \cdots, K_h(X_n - x)\right),$$

and

$$\mathbf{X} = \begin{pmatrix} 1 & \cdots & 1 \\ X_1 - x & \cdots & X_n - x \end{pmatrix}^T,$$

By the local linear fitting, it is known that $\sigma^2(x)$ can be estimated by the kernel estimator defined as

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

where

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

Define the squared residuals as $\widehat{r}_i = (Y_i - \widehat{m}_h(X_i))^2, i = 1, \ldots, n$. 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) has shown that the estimate $\widehat{\sigma}_L^2(x)$ is more efficient than the kernel estimator $\widehat{\sigma}^2(x)$. However, as discussed by Wang *et al.* (2008), the bias of the local linear estimate of the regression function would affect the efficiency of $\widehat{\sigma}^2(x)$ and $\widehat{\sigma}_L^2(x)$. So it is a natural idea to 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 the new version 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 bias reduced estimators of the error variance function remove the bias effect occurring in the first step of regression estimation. Hence they are more stable and efficient than $\widehat{\sigma}_L^2(x)$ and $\widehat{\sigma}^2(x)$.

3.3 Simultaneous Confidence Band

Without loss of generality, we consider construction of simultaneous confidence bands for $m(\cdot)$ on the interval $[0,1]$. For this purpose, we make use of the bias reduced estimators $\widehat{m}_B(x)$ and $\widetilde{m}_B(x)$.

Recall that $\nu_0 = \int K^2(t)dt$. We have the following theorem.

Theorem 2. Under assumptions (a)-(e) given in the Supplement, $h = n^{-b}, 1/(2q+3) \le b < 1-2/s$, $h = o(h_0)$ 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

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

where

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

if assumption (e1) given in the Supplement 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) given in the Supplement is satisfied.

As shown in Section 2, the asymptotic variance of $\widehat{m}_B(x)$ is same as that of $\widehat{m}_h(x)$. Hence it can be simply 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 the following simultaneous confidence band for $m(x)$ on $[0,1]$:

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

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 $m(x)$ will be covered by the above band is approximately $1-\alpha$.

The bandwidth h_0 can be simply selected by some plug-in methods based the optimal bandwidth selector for the local linear fitting. It would reduce much 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

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

Similar as Theorem 2, we have the following theorem.

Theorem 3. Under assumptions (a)-(d) and (e2) with $h = n^{-b}, 1/(2q+3) \leq$

$b < 1 - 2/s$, we have

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

where

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

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

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

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

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

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 somehow complicated and it is not easy to estimate it stably even though we provide a sandwich formula. 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 proposed 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 still 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, finite sample performances of the proposed bias reduced estimators are investigated under nonparametric regression and varying coefficient models.

Example 1. Consider the following univariate regressions with constant variance function $\sigma(x)$, $m(x)$ specified in the same way by in Fan and Gijbels (1996) and He and Huang (2009) and the covariate X having a uniform distribution over $[-2, 2]$. In all of the following four models $n = 200$ and 1000 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 four regression models, and the local linear estimate $\widehat{m}(x)$ and the bias reduced estimates $\widetilde{m}_B$ and $\widehat{m}_B(x)$ for 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 the bias reduced estimates $\widetilde{m}_B(x)$ and $\widetilde{m}_{BW}(x)$. The bandwidth series $h_i = (1 + C_i/10)h_{\text{opt}}n^{\frac{1}{5}}, C_i = 0, 1, \ldots, 10$, was used to obtain the bias reduced estimates $\widehat{m}_B(x)$ 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 the performance.

For the first three models, Table 1 shows that the MSE of the bias reduced estimators $\widetilde{m}_B(x)$ and $\widetilde{m}_{BW}(x)$ are much smaller than that of the local linear estimator (LL). From the 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. This is possibly because the bandwidth series used to construct $\widehat{m}_B(x)$ and $\widehat{m}_{BW}(x)$ is 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 (in other words, the h^2-order asymptotic bias is exactly zero) as the true regression function is linear.

Following one referee's suggestion, we also compared our bias reduction methods with the twicing local linear kernel regression smoother (TLL) and

Table 1: Median (Median Absolute Deviation) ×1000 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)

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. Similar 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 $h_i = (1 + C_i/10)h_{\text{opt}}, C_i = 0, 1, \ldots, 10$, to construct the bias reduced estimates $\widetilde{m}_B(x)$ and $\widetilde{m}_{BW}(x)$. And 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 the bias reduced estimates $\widehat{m}_B(x)$ and $\widehat{m}_{BW}(x)$. The numerical results are shown in Table 2. From Table 2 it is obvious that our bias reduction methods perform much better than the local cubic smoother. Compared to the twicing local linear estimators, our bias reduced estimates $\widetilde{m}_B(x)$ and $\widetilde{m}_{BW}(x)$

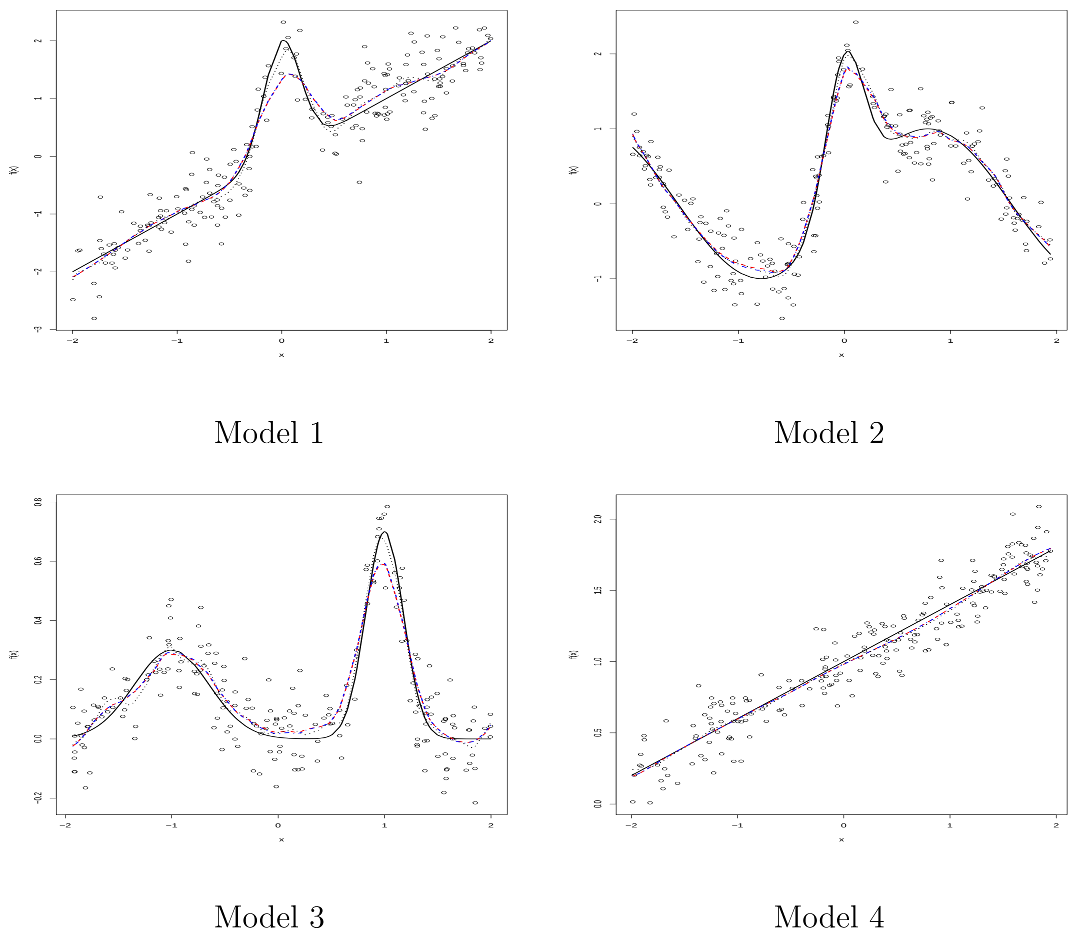


Figure 1: Example 1, Models 1-4. True regression function (solid line), and the local linear estimate $\widehat{m}(x)$ (dash line) and the bias reduced estimate $\widetilde{m}_B(x)$ (dotted line) and $\widehat{m}_B(x)$ (dot-dash line) based on one realization (circles).

are always more perferrable, and $\widehat{m}_B(x)$ and $\widehat{m}_{BW}(x)$ are comparable and sometimes better (probably because the bandwidth series used for $\widehat{m}_B(x)$ and $\widehat{m}_{BW}(x)$ is relatively large). Moreover, our methods are much 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 bias reduced estimators are based on multiple local linear estimators with different bandwidths.

Table 2: Mean (Standard Deviation) $\times 1000$ 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)

Example 2. Consider the following two varying coefficient models with $n = 500$ and 100 replicate samples. These two models were studied by Fan

and Zhang (1999, 2000).

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

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

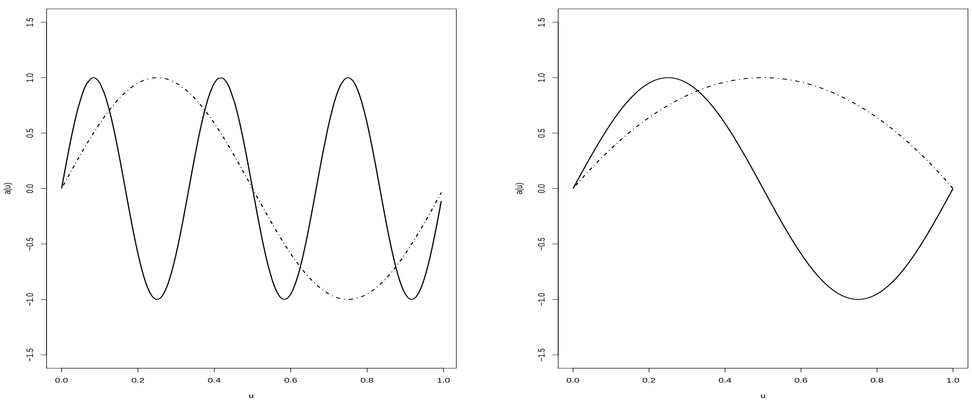
where U follows a uniform distribution on $[0, 1]$, and X_1 and X_2 are standard normal random variables with correlation coefficient $2^{-1/2}$. Furthermore, ε, U and (X_1, X_2) are independent. The random error ε follows a normal distribution with mean zero and variance σ^2. The variance σ^2 was chosen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

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

Given different original bandwidths $h_o = 0.05, 0.075, 0.15, 0.225, 0.3$, we calculate the local linear estimates for the varying coefficient 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 $h_i = (1 + C_i/10)h_o n^{\frac{1}{5}}, C_i = 0, 1, \ldots, 10$, was used to calculate the bias reduced estimate $\widehat{a}_B(u)$. The numerical results are summarized by Table 3 and Table 4 for the varying coefficient models 1 and 2 respectively.

From Figure 2, we can see that the varying coefficient functions in the first model are more occillating than those in the second model. From Tables 3 and 4, compared to the local linear estimator the MSE of the bias reduced 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 the bandwidth h_0 is small, the bias of the local linear estimator is relatively small, and then the bias reduced estimators would have little advantage or even overestimate. In addition, when there are more than one varying coefficient functions to estimate, bias reduction may not be achieved simultaneously. The one-step estimation procedure proposed by Fan and Zhang (1999) can be considered; this is outside the scope of this paper, however.

Observe from the numerical results shown in Tables 1-4 for Examples 1-2 and by the corresponding discussions given above, we have the following conclusions. When the signal-to-noise ratio or the variation of the regression function are relative large, $\widetilde{m}_B(x)$ and $\widetilde{a}_B(u)$ would be more appropriate for the bias reduction purpose. When the signal-to-noise ratio is small or the estimated function is smooth, the performance of $\hat{m}_B(x)$ and $\hat{a}_B(u)$ would be much better than the performance of $\widetilde{m}_B(x)$ and $\widetilde{a}_B(u)$, respectively.



Model 1 Model 2

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.

4.2 Variance Function Estimation

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

Example 3. We simulate 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 $\{e_i\}$ are independent

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)

with $X_i \sim \text{Unif}[-2, 2]$ and $e_i \sim N(0, 1)$. Four different values of a, namely $a = 0.5, 1, 2, 4$, were used in the simulation. For each simulated sample, the performance of the estimator is 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 $n_{\text{grid}} = 101$.

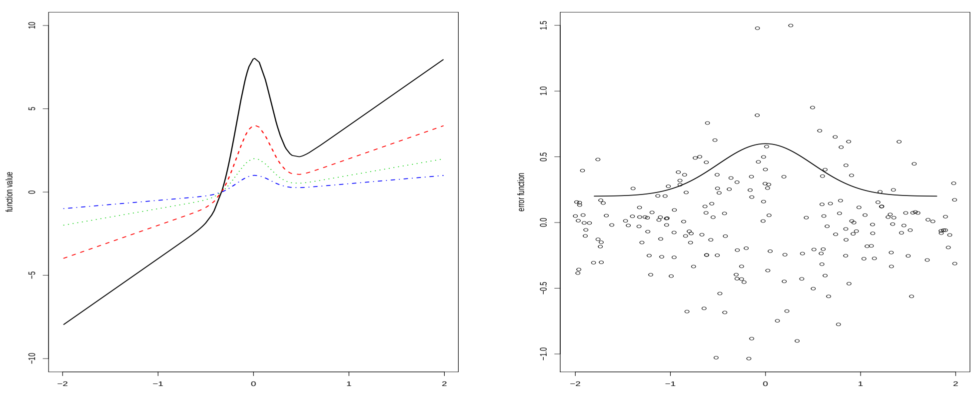
The numerical results for the variance function estimator by Fan and Yao (1998) and the proposed estimators are shown in Table 5. From Table

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)

5 we can see that $\widetilde{\sigma}^2_B(x)$, the variance function estimator based on $\widetilde{m}_B(x)$, always outperforms the local linear estimator of the variance function by Fan and Yao (1998). Especially, its performance improves as a becomes larger. This seems reasonable because the sign-to-noise ratio increases as a increases, as shown by Figure 3(a). As for $\widehat{\sigma}^2_B(x)$, the variance function estimator based on $\widehat{m}_B(x)$, it outperforms the estimator by Fan and Yao (1998) when a becomes smaller. From the numerical results of Example 1, we know the performance of $\widehat{m}_B(x)$ is similar to the local linear estima-

tor. Now, small bias reduction in estimating the regression function could make much improvement in estimating the variance function of the error. This gives strong support for the conclusion by Wang *et al.* (2008) that bias in estimation of the nonparametric regression function would seriously affect the efficiency of the estimation for the variance function of the error. On the other hand, when the signal-to-noise ratio increases (i.e. when a increases), the variation of the bias estimate increases which makes the variance function estimator based on $\widehat{m}_B(x)$ perform worse.



(a) (b)

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.

Table 5: Median (Median Absolute Deviation) ×100 of Mean Absolute Deviation Error for Models

	$\widetilde{\sigma}_B^2(x)$	$\widehat{\sigma}_B^2(x)$	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)

4.3 Simultaneous Confidence Band

In this section, we consider the following example which have been 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 1000 replicate samples were considered in the simulation.

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

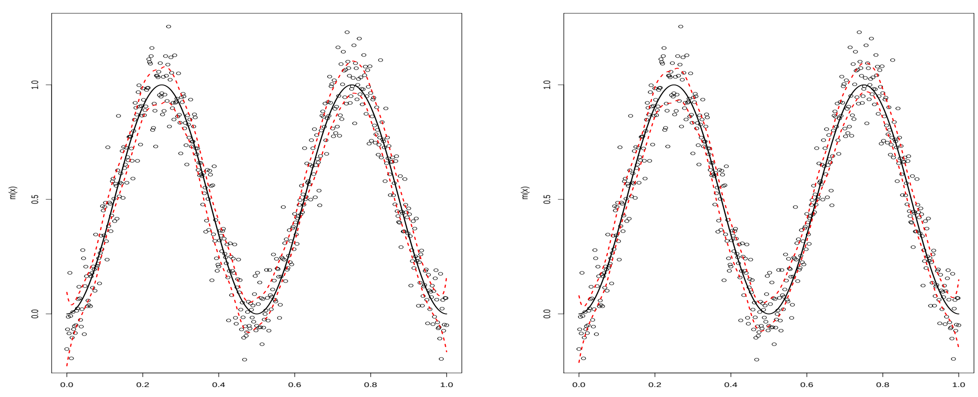
Consider using the bias reduced estimate $\widehat{m}_B(x)$ to construct simultaneous confidence band. Undersmoothing is necessary to construct the simultaneous confidence bands in practice. So we used $h_B = \frac{1}{2}h_{opt}$ and selected the bandwidth series $h_i = (1 + C_i/B)h_B n^{1/5}$ to construct the bias reduced estimate $\widehat{m}_B(x)$, where h_{opt} is again the rule-of-thumb bandwidth and $C_i = 0, 1, \ldots, B$ with $B = 10$. Then based on $\widehat{m}_B(x)$ and by following the procedure given in Section 3.3 we constructed the simultaneous confidence band for the regression function. We do not consider simultaneous confidence bands based on $\widetilde{m}_B(x)$ because the bootstrap variance estimation discussed in Section 3.3 requires much more computational time. Note that compared to the procedures suggested by Xia (1998) and Fan and Zhang (2000), we do not need to do any further steps to reduce the bias of $\widehat{m}_B(x)$ in constructing the simultaneous confidence band. Our numerical results shown in Table 6 is comparable or even better than the results given by 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 we can see clearly the benefit of using the biased reduced estimator $\widehat{m}_B(x)$ in construction of simultaneous confidence bands.

4.4 Real Example

The motorcycle data set given in Härdle (1990) consists of accelerometer reading taken through time in an experiment on the efficiency of crash hel-

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



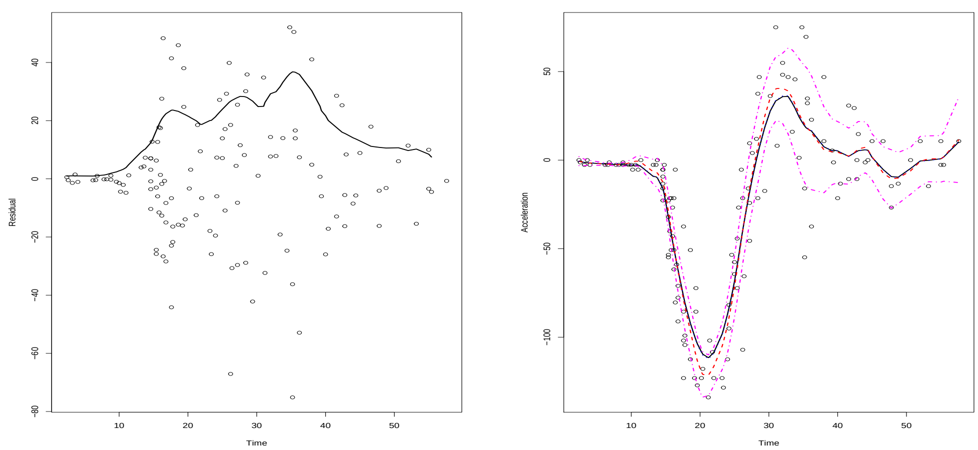
(a) Nominal coverage $1 - \alpha = 0.95$ (b) Nominal coverage $1 - \alpha = 0.90$

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.

mets. 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 is described in Schmidt *et al.* (1981). We use similar procedures as before to obtain the local linear estimate, $\widetilde{m}_B$ and $\widehat{m}_B$ respectively, and then following our suggested 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 it is obvious the variance 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 $\widetilde{m}_B$ has slimly modification with 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



(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 $\widetilde{m}_B(x)$; Dot line: the estimate $\widehat{m}_B(x)$; Dot-Dash lines: the simultaneous confidence band based on $\widehat{m}_B(x)$.

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 have wild applications in the inference, for example, construction of simultaneous confidence bands. Notably, in our approach we need not to use complicated procedures to remove the bias effect of the nonparametric function estimation in construction of simultaneous confidence band construction. This helps improve the stability of the inference and decision making.

There is still space to improve our bias reduction methods. For example, from our extensive numerical studies we know that choice of the bandwidth series could change the efficiency of the bias reduced estimators. We have not discussed thoroughly which kind of bandwidth series is optimal. This is an interesting and important topic that requires further investigation. Furthermore, as shown by Example 3 of our numerical studies, the two bias reduced estimators show much different performances when used in estimating of the variance of the error. Even though the estimate $\widehat{m}_B(x)$ does not reduce much bias compared to the local linear estimator, in some situations, based on $\widehat{m}_B(x)$ the performance of further model estimation and inference can be much improved. It is interesting to fully understand the differences between the two bias reduction approaches and it could be beneficial for the further model estimation and inference.

Supplementary Materials

Technical conditions and proofs of the main theoretical results.

Acknowledgements

The authors would like to thank the Editor, 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, pp. 89–99.*

Choi, E. and Hall, P. (1998). On Bias Reduction in Local Linear Smoothing. Biometrika 85, pp. 333-345.

Choi, E., Hall, P. and Rousson, V. (2000). Data Sharpening Methods for Bias Reduction in Nonparametric Regression. *The Annals of Statistics 28, pp. 1339–1355.*

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

Fan, J. (1993). Local Linear Regression Smoothers and Their Minimax Efficiencies. *The Annals of Statistics 21, pp. 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, pp. 645–660.

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

Fan, J. and Zhang, W. (2000). Simultaneous Confidence Bands and Hypothesis Testing in Varying-coefficient Models. Scandinavian Journal of Statistics 27, pp. 715–731.

Hall, P. and Horowitz, J. (2013). A Simple Bootstrap Method for Constructing Nonparametric Confidence Bands for Functions. *The Annals of Statistics 41, pp. 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 Com-

ments). Statistical Science 8, pp. 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, pp. 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, pp. 937–955.

Ruppert, D. and Wand, M. P. (1994). Multivariate Locally Weighted Least Squares Regression. *The Annals of Statistics 22, pp. 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, pp. 646–664.

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

Xia, Y. (1998). Bias-corrected Confidence Bands in Nonparametric Regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60, pp. 797–811.

Zhang, W. and Peng, H. (2010). Simultaneous Confidence Band and Hypothesis Test in Generalised Varying-coefficient Models. *Journal of Multivariate Analysis 101, pp. 1656–1680.*

Zhang, W. and Xia, Y. (2012). Twicing Local Linear Regression Smoothers. Journal of Nonparametric Statistics 24, pp. 399–417.

Hong Kong Baptist University and National Taiwan University

E-mail: (chengmingyen@hkbu.edu.hk)

Shanghai University of Finance and Economics

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

University of Washington and Fred Hutchinson Cancer Research Center

E-mail: (liupeng@amss.ac.cn)

Hong Kong Baptist University

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