ISSN 1440-771X

MONASH University

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Local linear multivariate regression with variable bandwidth in the presence of heteroscedasticity

Azhong Ye, Rob J Hyndman and Zinai Li

May 2006

Working Paper 8/06

Local linear multivariate regression with variable bandwidth in the presence of heteroscedasticity

Azhong Ye

College of Management

Fuzhou University, Fuzhou, 350002

China.

Email: ye2004@fzu.edu.cn

Rob J Hyndman

Department of Econometrics and Business Statistics,

Monash University, VIC 3800

Australia.

Email: Rob.Hyndman@buseco.monash.edu

Zinai Li

School of Economics and Management,

Tsinghua University, Beijing, 100084,

China.

Email: lizinai@mail.tsinghua.edu.cn

18 May 2006

JEL classification: C12,C15,C52

Local linear multivariate regression with variable bandwidth in the presence of heteroscedasticity

Abstract: We present a local linear estimator with variable bandwidth for multivariate non-parametric regression. We prove its consistency and asymptotic normality in the interior of the observed data and obtain its rates of convergence. This result is used to obtain practical direct plug-in bandwidth selectors for heteroscedastic regression in one and two dimensions. We show that the local linear estimator with variable bandwidth has better goodness-of-fit properties than the local linear estimator with constant bandwidth, in the presence of heteroscedasticity.

Keywords: heteroscedasticity; kernel smoothing; local linear regression; plug-in bandwidth, variable bandwidth.

1 Introduction

We are interested in the problem of heteroscedasticity in nonparametric regression, especially when applied to economic data. Heteroscedasticity is very common in economics, and heteroscedasticity in linear regression is covered in almost every econometrics textbook. Applications of nonparametric regression in economics are growing, and Yatchew (1998) argued that it will become an indispensable tool for every economist because it typically assumes little about the shape of the regression function. Consequently, we believe that heteroscedasticity in nonparametric regression is an important problem that has received limited attention to date. We seek to develop new estimators that have better goodness of fit than the common estimators in nonparametric econometric models. In particular, we are interested in using heteroscedasticity to improve nonparametric regression estimation.

There have been a few papers on related topics. Testing for heteroscedasticity in nonparametric regression has been discussed by Eubank and Thomas (1993) and Dette and Munk (1998). Ruppert and Wand (1994) discussed multivariate locally weighted least squares regression when the variances of the disturbances are not constant. Ruppert et al. (1997) presented the local polynomial estimator of the conditional variance function in a heteroscedastic, nonparametric regression model using linear smoothing of squared residuals. Sharp-optimal and adaptive estimators for heteroscedastic nonparametric regression using the classical trigonometric Fourier basis are given by Efromovich and Pinsker (1996).

Our approach is to exploit the heteroscedasticity by using variable bandwidths in local linear regression. Müller and Stadtmüller (1987) discussed variable bandwidth kernel estimators of regression curves. Fan and Gijbels (1992, 1995, 1996) discussed the local linear estimator with variable bandwidth for nonparametric regression models with a single covariate. In this paper, we extend these papers by presenting a local linear estimator with variable bandwidth for nonparametric multiple regression models.

We demonstrate that the local linear estimator has optimal conditional mean squared error when its variable bandwidth is a function of the density of the explanatory variables and conditional variances. Numerical simulation shows that the local linear estimator with this variable bandwidth has better goodness of fit than the local linear estimator with constant bandwidth for the heteroscedastic models.

2 Local linear regression with a variable bandwidth

Suppose we have a univariate response variable Y and a d-dimensional set of covariates $\boldsymbol{X}$, and we observe the random vectors $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$ which are independent and identically distributed. It is assumed that each variable in $\boldsymbol{X}$ has been scaled so they have similar measures of spread.

Our aim is to estimate the regression function $m(\boldsymbol{x}) = \mathrm{E}(Y \mid \boldsymbol{X} = \boldsymbol{x})$. We can regard the data as being generated from the model

$$Y = m(\boldsymbol{X}) + u,$$

where $\mathrm{E}(u \mid \boldsymbol{X}) = 0$, $\mathrm{Var}(u \mid \boldsymbol{X} = \boldsymbol{x}) = \sigma^2(\boldsymbol{x})$ and the marginal density of $\boldsymbol{X}$ is denoted by $f(\boldsymbol{x})$. We assume the second-order derivatives of $m(\boldsymbol{x})$ are continuous, $f(\boldsymbol{x})$ is bounded above 0 and $\sigma^2(\boldsymbol{x})$ is continuous and bounded.

Let K be a d-variate kernel which is symmetric, nonnegative, compactly supported, $\int K(\boldsymbol{u})\, d\boldsymbol{u} = 1$ and $\int \boldsymbol{u}\boldsymbol{u}^T K(\boldsymbol{u})\, d\boldsymbol{u} = \mu_2(K)\boldsymbol{I}$ where $\mu_2(K) \neq 0$ and $\boldsymbol{I}$ is the $d \times d$ identity matrix. In addition, all odd-order moments of K vanish, that is, $\int u_1^{l_1} \ldots u_d^{l_d} K(\boldsymbol{u})\, d\boldsymbol{u} = 0$ for all nonnegative integers $l_1, \ldots, l_d$ such that their sum is odd. Let $K_h(\boldsymbol{u}) = K(\boldsymbol{u}/h)$.

Then the local linear estimator of $m(\boldsymbol{x})$ with variable bandwidth is

$$\hat{m}_n(\boldsymbol{x}, h_n, \alpha) = \boldsymbol{e}_1^T (\boldsymbol{X}_{\boldsymbol{x}}^T \boldsymbol{W}_{\boldsymbol{x},\alpha} \boldsymbol{X}_{\boldsymbol{x}})^{-1} \boldsymbol{X}_{\boldsymbol{x}}^T \boldsymbol{W}_{\boldsymbol{x},\alpha} \boldsymbol{Y}, \tag{1}$$

where $h_n = cn^{-1/(d+4)}$, c is a constant that depends only on K, $\alpha(\boldsymbol{x})$ is the variable bandwidth function, $\boldsymbol{e}_1^T = (1, 0, \ldots, 0)$, $\boldsymbol{X}_{\boldsymbol{x}} = (\boldsymbol{X}_{\boldsymbol{x},1}, \ldots, \boldsymbol{X}_{\boldsymbol{x},n})^T$, $\boldsymbol{X}_{\boldsymbol{x},i} = (1, (\boldsymbol{X}_i - \boldsymbol{x}))^T$ and $\boldsymbol{W}_{\boldsymbol{x},\alpha} = \mathrm{diag}\left(K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1 - \boldsymbol{x}), \ldots, K_{h_n\alpha(\boldsymbol{X}_n)}(\boldsymbol{X}_n - \boldsymbol{x})\right)$. We assume $\alpha(\boldsymbol{x})$ is continuously differentiable.

We now state our main result. A proof is given in the Appendix.

Theorem 1 *Let $\boldsymbol{x}$ be a fixed point in the interior of $\{\boldsymbol{x} \mid f(\boldsymbol{x}) > 0\}$, $\boldsymbol{H}_m(\boldsymbol{x}) = \left(\frac{\partial^2 m(\boldsymbol{x})}{\partial x_i \partial x_j}\right)_{d\times d}$ and let $s(\boldsymbol{H}_m(\boldsymbol{x}))$ be the sum of the elements of $\boldsymbol{H}_m(\boldsymbol{x})$. Then*

1. $\mathrm{E}\left[\hat{m}_n(\boldsymbol{x}, h_n, \alpha) \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\right] - m(\boldsymbol{x}) = 0.5 h_n^2 \mu_2(K) \alpha^2(\boldsymbol{x}) s(\boldsymbol{H}_m(\boldsymbol{x})) + o_p(h_n^2)$;
2. $\mathrm{Var}\left[\hat{m}_n(\boldsymbol{x}, h_n, \alpha) \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\right] = n^{-1} h_n^{-d} R(K) \alpha^{-d}(\boldsymbol{x}) \sigma^2(\boldsymbol{x}) f^{-1}(\boldsymbol{x}) + o_p(n^{-1} h_n^{-d})$, *where* $R(K) = \int K^2(\boldsymbol{u}) d\boldsymbol{u}$; *and*
3. $n^{2/(d+4)}\left[\hat{m}_n(\boldsymbol{x}, h_n, \alpha) - m(\boldsymbol{x})\right] \xrightarrow{d}$
$$N\left(0.5 c^2 \mu_2(K) \alpha^2(\boldsymbol{x}) s(\boldsymbol{H}_m(\boldsymbol{x}))\,,\; c^{-d} R(K) \alpha^{-d}(\boldsymbol{x}) \sigma^2(\boldsymbol{x}) f^{-1}(\boldsymbol{x})\right).$$

When $\alpha(\boldsymbol{x}) = 1$, results 1 and 2 coincide with Theorem 2.1 of Ruppert and Wand (1994). By result 2 and the law of large numbers, we find that $\hat{m}_n(\boldsymbol{x}, h_n, \alpha)$ is consistent. From result 3 we know that the rate of convergence of $\hat{m}_n(\boldsymbol{x}, h_n, \alpha)$ in interior points is $O(n^{-2/(d+4)})$ which, according to Stone (1980, 1982), is the optimal rate of convergence for nonparametric estimation of a smooth function $m(\boldsymbol{x})$.

3 Using heteroscedasticity to improve local linear regression

Although Fan and Gijbels (1992) and Ruppert and Wand (1994) discuss the local linear estimator of $m(\boldsymbol{x})$, nobody has previously developed an improved estimator using the information of heteroscedasticity. We now show how this can be achieved.

Using Theorem 1, we can give an expression for the conditional mean squared error of the local linear estimator with variable bandwidth.

$$\begin{aligned}\mathrm{MSE} &= \mathrm{E}\left\{\left[\hat{m}_n(\boldsymbol{x}, h_n, \alpha) - m(\boldsymbol{x})\right]^2 \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\right\} \\ &= \tfrac{1}{4}[h_n \alpha(\boldsymbol{x})]^4 \mu_2^2(K) s^2(\boldsymbol{H}_m(\boldsymbol{x})) + \frac{R(K)\sigma^2(\boldsymbol{x})}{n[h_n \alpha(\boldsymbol{x})]^d f(\boldsymbol{x})} + o_p(h_n^2) + o_p(n^{-1} h_n^{-d}).\end{aligned}$$

Minimizing MSE, ignoring the higher order terms, we obtain the optimal variable bandwidth

$$\alpha_{opt}(\boldsymbol{x}) = c^{-1}\left(\frac{d R(K) \sigma^2(\boldsymbol{x})}{\mu_2^2(K) f(\boldsymbol{x}) s^2(\boldsymbol{H}_m(\boldsymbol{x}))}\right)^{1/(d+4)}.$$

Note that the constant c will cancel out in the bandwidth expression $h_n \alpha_{opt}(\boldsymbol{x})$. Therefore, without loss of generality, we can set $c = \left\{dR(K)\mu_2^{-2}(K)\right\}^{1/(d+4)}$ which simplifies the above expression to

$$\alpha_{opt}(\boldsymbol{x}) = \left(\frac{\sigma^2(\boldsymbol{x})}{f(\boldsymbol{x}) s^2(\boldsymbol{H}_m(\boldsymbol{x}))} \right)^{1/(d+4)}.$$

To apply this new estimator, we need to replace $f(\boldsymbol{x})$, $\boldsymbol{H}_m(\boldsymbol{x})$ and $\sigma^2(\boldsymbol{x})$ with estimators. There are several potential ways to do this, depending on the dimension d. Some proposals for $d = 1$ and $d = 2$, are outlined below.

3.1 Univariate regression

When $d = 1$, we first use the direct plug-in methodology of Sheather and Jones (1991) to select the bandwidth of a kernel density estimate for $f(\boldsymbol{x})$. Second, we estimate $\sigma^2(\boldsymbol{x}) = \mathrm{E}(u^2 \mid \boldsymbol{X} = \boldsymbol{x})$ using local linear regression with the model $\hat{u}_i^2 = \sigma^2(\boldsymbol{X}_i) + v_i$ where $\hat{u}_i = Y_i - \hat{m}_n(\boldsymbol{X}_i, \hat{h}_n, 1)$, v_i are iid with zero mean and $\hat{h}_n$ is chosen by the direct plug-in methodology of Ruppert et al. (1995). Third, we estimate $\ddot{m}(x)$ by fitting the quartic $m(x) = \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \alpha_4 x^3 + \alpha_5 x^4$, using ordinary least squares regression and so obtain the estimate $\hat{\ddot{m}}(x) = 2\hat{\alpha}_3 + 6\hat{\alpha}_4 x + 12\hat{\alpha}_5 x^2$. Then, our direct plug-in bandwidth for univariate regression ($d = 1$) is

$$\hat{h}(x) = \left(\frac{\hat{\sigma}^2(x)}{2n\sqrt{\pi}\hat{f}(x)\hat{\ddot{m}}^2(x)} \right)^{1/5}.$$

3.2 Bivariate regression

When $d = 2$, we use a bivariate kernel density estimator (Scott, 1992) of $f(\boldsymbol{x})$, with the direct plug-in methodology of Wand and Jones (1995) for the bandwidth. To estimate $\sigma^2(\boldsymbol{x})$, we first calculate $\hat{u}_i = Y_i - \hat{Y}_i$, where $\hat{Y}_i = \hat{m}_n(\boldsymbol{X}_i, \hat{h}_n, 1)$, $\hat{h}_n = min(\hat{h}_1, \hat{h}_2)$, and $\hat{h}_1$ and $\hat{h}_2$ are chosen by the direct plug-in methodology of Ruppert et al. (1995) for $Y = m_1(X_1) + u_1$ and $Y = m_2(X_2) + u_2$ respectively. Then we estimate $\sigma^2(\boldsymbol{x}_1)$ using local linear regression with the model $\hat{u}_i^2 = \sigma^2(\boldsymbol{X}_{1i}) + v_i$, where v_i are iid with zero mean. Again, the direct plug-in methodology of Ruppert et al. (1995) is used for bandwidth selection.

To estimate the second derivative of $m(x_1, x_2)$, we fit the model

$$m(x_1,x_2) = \alpha_1 + \alpha_2 x_1 + \alpha_3 x_2 + \alpha_4 x_1^2 + \alpha_5 x_1 x_2 + \alpha_6 x_2^2 + \alpha_7 x_1^3 + \alpha_8 x_1^2 x_2 + \alpha_9 x_1 x_2^2 + \alpha_{10} x_2^3 + \alpha_{11} x_1^4 + \alpha_{12} x_1^3 x_2 + \alpha_{13} x_1^2 x_2^2 + \alpha_{14} x_1 x_2^3 + \alpha_{15} x_2^4,$$

using ordinary least squares, and so obtain estimates of α. Hence, estimators for the second derivatives of $m(x_1, x_2)$ are obtained using

$$\frac{\partial^2 m(\boldsymbol{x})}{\partial x_1^2} = 2\alpha_4 + 6\alpha_7 x_1 + 2\alpha_8 x_2 + 12\alpha_{11} x_1^2 + 6\alpha_{12} x_1 x_2 + 2\alpha_{13} x_2^2,$$

$$\frac{\partial^2 m(\boldsymbol{x})}{\partial x_1 \partial x_2} = \alpha_5 + 2\alpha_8 x_1 + 2\alpha_9 x_2 + 3\alpha_{12} x_1^2 + 4\alpha_{13} x_1 x_2 + 3\alpha_{14} x_2^2$$

and $$\frac{\partial^2 m(\boldsymbol{x})}{\partial x_2^2} = 2\alpha_6 + 2\alpha_9 x_1 + 6\alpha_{10} x_2 + 2\alpha_{13} x_1^2 + 6\alpha_{14} x_1 x_2 + 12\alpha_{15} x_2^2.$$

Then, our direct plug-in bandwidth for bivariate regression ($d = 2$) is

$$\hat{h}(\boldsymbol{x}) = \left(\frac{\hat{\sigma}^2(\boldsymbol{x})}{n\pi \hat{f}(\boldsymbol{x}) s^2(\hat{\boldsymbol{H}}_m(\boldsymbol{x}))} \right)^{1/6}.$$

4 Numerical studies with univariate regression

This section examines the performance of the proposed variable bandwidth selection method via several data sets of univariate regression, generated from known functions. For comparison, we also compare the performance of the constant bandwidth method, based on the direct plug-in methodology described by Ruppert et al. (1995).

As the true regression function is known in each case, the performances of the bandwidth methods are measured and compared using the root mean squared error,

$$\text{RMSE} = \left(n^{-1} \sum_{i=1}^{n} [\hat{m}_n(\boldsymbol{X}_i, h_n, \alpha) - m(\boldsymbol{X}_i)]^2 \right)^{1/2}.$$

We simulate data from the following five models, each with $Y = m(X) + \sigma(X)u$ where $u \sim N(0,1)$

and the covariate X has a Uniform $(-2,2)$ distribution.

Model A: $m_A(x) = x^2 + x$

$\sigma_A^2(x) = 32x^2 + 0.04$

Model B: $m_B(x) = (1+x)\sin(1.5x)$

$\sigma_B^2(x) = 3.2x^2 + 0.04$

Model C: $m_C(x) = x + 2\exp(-2x^2)$

$\sigma_C^2(x) = 16(x^2 - 0.01)I_{(x^2>0.01)} + 0.04$

Model D: $m_D(x) = \sin(2x) + 2\exp(-2x^2)$

$\sigma_D^2(x) = 16(x^2 - 0.01)I_{(x^2>0.01)} + 0.04$

Model E: $m_E(x) = \exp(-(x+1)^2) + 2\exp(-2x^2)$

$\sigma_E^2(x) = 32(x^2 - 0.01)I_{(x^2>0.01)} + 0.04$

We draw 1000 random samples of size 200 from each model. Table 1 presents a summary of the results and shows that the variable bandwidth method has smaller RMSE than the constant bandwidth method in each case. For each model, the variable bandwidth method has smaller RMSE than the constant bandwidth method, and is better for more than 50% of samples.

We plot the true regression functions (the solid line) and four typical estimated curves in Figure 1. These correspond to the 10th, 30th, 70th and 90th percentiles. For each percentile, the variable bandwidth method (dotted line) is closer to the true regression function than the constant bandwidth method (dashed line). Therefore, we conclude that for heteroscedastic models, the local linear estimator with variable bandwidth has better goodness-of-fit than the local linear estimator with constant bandwidth.

5 Numerical studies with bivariate regression

We now examine the performance of the proposed variable bandwidth selection method via several data sets of bivariate regression, generated from known functions. For comparison, we also compare the performance of a constant bandwidth method given by

$$\hat{h} = \left(\frac{\hat{\sigma}^2}{\pi \sum_{i=1}^{n} s^2(\hat{\boldsymbol{H}}_m(\boldsymbol{X}_i))} \right)^{1/6}$$

where

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} \hat{u}_i^2,$$

and u_i and $\hat{\boldsymbol{H}}_m$ are the same as for the variable bandwidth selector.

We simulate data from four models, each with $Y = m(X_1, X_2) + \sigma(X_1)u$ where $u \sim N(0,1)$ and the covariates X_1 and X_2 are independent and have a Uniform $(-2,2)$ distribution.

Model F: $m_A(x_1, x_2) = x_1 x_2$

$\sigma_A^2(x_1, x_2) = (x_1^2 - 0.04) I_{(x_1^2 > 0.04)} + 0.01$

Model G: $m_B(x_1, x_2) = x_1 \exp(-2x_2^2)$

$\sigma_B^2(x_1, x_2) = 2.5(x_1^2 - 0.04) I_{(x_1^2 > 0.04)} + 0.025$

Model H: $m_C(x_1, x_2) = x_1 + 2\sin(1.5x_2)$

$\sigma_C^2(x_1, x_2) = (x_1^2 - 0.04) I_{(x_1^2 > 0.04)} + 0.01$

Model I: $m_D(x_1, x_2) = \sin(x_1 + x_2) + 2\exp(-2x_2^2)$

$\sigma_D^2(x_1, x_2) = 3(x_1^2 - 0.04) I_{(x_1^2 > 0.04)} + 0.03$

We draw 200 random samples of size 400 from each model.

Table 2 presents a summary of the results and shows that the variable bandwidth method has smaller RMSE than the constant bandwidth method. For each model, the variable bandwidth method has lower RMSE than the constant bandwidth method and is better for more than 50% of samples.

We plot the true regression functions with one fixed variable (solid line) and four typical estimated curves in Figures 2–5. These correspond to the 10th, 30th, 70th and 90th percentiles. For each percentile, the variable bandwidth method (dotted line) is closer to the true regression function than the constant bandwidth method (dashed line). Therefore, we conclude that for heteroscedastic models, the local linear estimator with variable bandwidth has better goodness-of-fit than the local linear estimator with constant bandwidth.

6 Summary

We have presented a local linear nonparametric estimator with variable bandwidth for multivariate regression models. We have shown that the estimator is consistent and asymptotically normal in the interior of the sample space. We have also shown that its convergence rate is optimal for nonparametric regression (Stone, 1980, 1982).

By minimizing the conditional mean squared error of the estimator, we have derived the optimal variable bandwidth as a function of the density of the explanatory variables and the conditional variance. We have also provided a plug-in algorithm for computing the estimator when $d = 1$ or $d = 2$. Numerical simulation shows that our local linear estimator with variable bandwidth has better goodness-of-fit than the local linear estimator with constant bandwidth for heteroscedastic models.

Acknowledgements

We thank Xibin Zhang for his helpful comments.

References

Dette, H. and A. Munk (1998) Testing heteroscedasticity in nonparametric regression, *Journal of the Royal Statistical Society, Series B*, **60**(4), 693–708.

Efromovich, S. and M. Pinsker (1996) Sharp-optimal and adaptive estimation for heteroscedastic nonparametric regression, *Statistica Sinica*, **6**, 925–942.

Eubank, R. L. and W. Thomas (1993) Detecting heteroscedasticity in nonparametric regression, *Journal of the Royal Statistical Society, Series B*, **55**(1), 145–155.

Fan, J. and I. Gijbels (1992) Variable bandwidth and local linear regression smoothers, *The Annals of Statistics*, **20**(4), 2008–2036.

Fan, J. and I. Gijbels (1995) Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation, *Journal of the Royal Statistical Society, Series B*, **57**(2), 371–394.

Fan, J. and I. Gijbels (1996) *Local polynomial modelling and its applications*, Chapman and Hall.

Müller, H.-G. and U. Stadtmüller (1987) Variable bandwidth kernel estimators of regression curves, *The Annals of Statistics*, **15**(1), 182–201.

Ruppert, D., S. J. Sheather and M. P. Wand (1995) An effective bandwidth selector for local least squares regression (Corr: 96V91 p1380), *Journal of the American Statistical Association*, **90**, 1257–1270.

Ruppert, D. and M. P. Wand (1994) Multivariate locally weighted least squares regression, *The Annals of Statistics*, **22**(3), 1346–1370.

Ruppert, D., M. P. Wand, U. Holst and O. Hössjer (1997) Local polynomial variance-function estimation, *Technometrics*, **39**(3), 262–273.

Scott, D. W. (1992) *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons.

Sheather, S. J. and M. C. Jones (1991) A reliable data-based bandwidth selection method for kernel density estimation, *J.R. Statist. Soc. B*, **53**(3), 683–690.

Stone, C. J. (1980) Optimal convergence rates for nonparametric estimators, *The Annals of Statistics*, **8**, 1348–1360.

Stone, C. J. (1982) Optimal global rates of convergence for nonparametric regression, *The Annals of Statistics*, **10**, 1040–1053.

Wand, M. P. and M. C. Jones (1995) *Kernel Smoothing*, Chapman and Hall, London.

White, H. (1984) *Asymptotic theory for econometricians*, Academic Press.

Yatchew, A. (1998) Nonparametric regression techniques in economics, *Journal of Economic Literature*, **34**, 669–721.

Appendix: Proof of Theorem 1

Before we state a lemma that will be used in the proof, note that

$$n^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{X}_{\boldsymbol{x}} = n^{-1}\begin{pmatrix} \sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x}) & \sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})^T \\ \sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x}) & \sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})^T \end{pmatrix}$$

and $$\boldsymbol{e}_1^T(\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{X}_{\boldsymbol{x}})^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{X}_{\boldsymbol{x}}\begin{bmatrix} m(\boldsymbol{x}) \\ \boldsymbol{D}_m(\boldsymbol{x})\end{bmatrix} = \boldsymbol{e}_1^T\begin{bmatrix} m(\boldsymbol{x}) \\ \boldsymbol{D}_m(\boldsymbol{x})\end{bmatrix} = m(\boldsymbol{x}),$$

where $\boldsymbol{D}_m(\boldsymbol{x}) = \left[\frac{\partial m(\boldsymbol{x})}{\partial x_1},\dots,\frac{\partial m(\boldsymbol{x})}{\partial x_d}\right]^T$. Therefore,

$$\hat{m}_n(\boldsymbol{x},h_n,\alpha) - m(\boldsymbol{x}) = \boldsymbol{e}_1^T(\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{X}_{\boldsymbol{x}})^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\left[0.5\boldsymbol{Q}_m(\boldsymbol{x})+\boldsymbol{U}\right],$$

where $\boldsymbol{Q}_m(\boldsymbol{x}) = \left[\boldsymbol{Q}_{m,1}(\boldsymbol{x}),\dots,\boldsymbol{Q}_{m,n}(\boldsymbol{x})\right]^T$, $\boldsymbol{Q}_{m,i}(\boldsymbol{x}) = (\boldsymbol{X}_i-\boldsymbol{x})^T\boldsymbol{H}_m(\boldsymbol{z}_i(\boldsymbol{x},\boldsymbol{X}_i))(\boldsymbol{X}_i-\boldsymbol{x})$, $\boldsymbol{H}_m(\boldsymbol{x}) = \left(\frac{\partial^2 m(\boldsymbol{x})}{\partial x_i \partial x_j}\right)_{d\times d}$, $\|\boldsymbol{z}_i(\boldsymbol{x},\boldsymbol{X}_i)-\boldsymbol{x}\| \le \|\boldsymbol{X}_i-\boldsymbol{x}\|$ and $\boldsymbol{U}=(u_1,\dots,u_n)^T$. We can deduce that $\{\boldsymbol{z}_i(\boldsymbol{x},\boldsymbol{X}_i)\}_{i=1}^n$ are independent because $\{\boldsymbol{X}_i\}_{i=1}^n$ are independent.

Now we state a lemma using the notation of (1) and Theorem 1.

Lemma 1 *Let*

$$G(\alpha,f,\boldsymbol{x}) = f(\boldsymbol{x})\int_{supp(K)}\boldsymbol{u}\boldsymbol{D}_\alpha^T(\boldsymbol{x})\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{D}_K(\boldsymbol{u})d\boldsymbol{u} + \mu_2(K)\left[df(\boldsymbol{x})\boldsymbol{D}_\alpha(\boldsymbol{x}) + \alpha^{-1}(\boldsymbol{x})\boldsymbol{D}_f(\boldsymbol{x})\right],$$

$$\boldsymbol{B}(\boldsymbol{x},\alpha) = -\mu_2(K)^{-1}f(\boldsymbol{x})^{-2}\alpha(\boldsymbol{x})G(\alpha,f,\boldsymbol{x})^T$$

and $\mathbf{1}$ *be a generic matrix having each entry equal to 1, the dimensions of which will be clear from the context. Then*

$$n^{-1}\sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x}) = f(\boldsymbol{x}) + o_p(1) \tag{2}$$

$$n^{-1}\sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x}) = h_n^2\alpha^3(\boldsymbol{x})G(\alpha,f,\boldsymbol{x}) + o_p(h_n^2\mathbf{1}) \tag{3}$$

$$n^{-1}\sum_{i=1}^{n} K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})^T = \mu_2(K)h_n^2\alpha^2(\boldsymbol{x})f(\boldsymbol{x})\boldsymbol{I} + o_p(h_n^2\boldsymbol{1}) \tag{4}$$

$$(\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{X}_{\boldsymbol{x}})^{-1} = \begin{pmatrix} f(\boldsymbol{x})^{-1}+o_p(1) & \boldsymbol{B}(\boldsymbol{x},\alpha)+o_p(\boldsymbol{1}) \\ \boldsymbol{B}^T(\boldsymbol{x},\alpha)+o_p(\boldsymbol{1}) & \mu_2(K)^{-1}h_n^{-2}\alpha^{-2}(\boldsymbol{x})f(\boldsymbol{x})^{-1}\boldsymbol{I}+o_p(h_n^{-2}\boldsymbol{1}) \end{pmatrix} \tag{5}$$

$$n^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{Q}_m(\boldsymbol{x}) = h_n^2\begin{pmatrix} f(\boldsymbol{x})\mu_2(K)\alpha^2(\boldsymbol{x})s(\boldsymbol{H}_m(\boldsymbol{x})) \\ \boldsymbol{0} \end{pmatrix} + o_p(h_n^2\boldsymbol{1}) \tag{6}$$

$$\textit{and} \qquad (nh_n^d)^{1/2}(n^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{u}) \xrightarrow{d} \begin{pmatrix} N\big(0\ ,\ R(K)\alpha^{-d}(\boldsymbol{x})\sigma^2(\boldsymbol{x})f(\boldsymbol{x})\big) \\ \boldsymbol{0} \end{pmatrix} \tag{7}$$

We only prove results (3), (6) and (7) as the other results can be proved similarly.

Proof of (3)

It is easy to show that

$$n^{-1}\sum_{i=1}^{n} K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x}) = \mathrm{E}\left[K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})(\boldsymbol{X}_1-\boldsymbol{x})\right] + O_p(\sqrt{n^{-1}\boldsymbol{\Psi}}), \tag{8}$$

where $\boldsymbol{\Psi} = \left(\mathrm{Var}(K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})(X_{11}-x_1)), \ldots, \mathrm{Var}(K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})(X_{d1}-x_d))\right)^T$.

Because $\boldsymbol{x}$ is a fixed point in the interior of $\mathrm{supp}(f) = \{\boldsymbol{x} \mid f(\boldsymbol{x}) \neq 0\}$, we have

$$\mathrm{supp}(K) \subset \{\boldsymbol{z} : (\boldsymbol{x} + h_n\alpha(\boldsymbol{x})\boldsymbol{z}) \in \mathrm{supp}(f)\},$$

provided the bandwidth h_n is small enough.

Due to the continuity of f, K and α, we have

$$\begin{aligned} &\mathrm{E}\left[K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})(\boldsymbol{X}_1-\boldsymbol{x})\right] \\ &\quad = \int_{\mathrm{supp}(f)} h_n^{-d}\alpha^{-d}(\boldsymbol{x})K(h_n^{-1}(\alpha(\boldsymbol{X}_1))^{-1}(\boldsymbol{X}_1-\boldsymbol{x}))(\boldsymbol{X}_1-\boldsymbol{x})f(\boldsymbol{X}_1)d\boldsymbol{X}_1 \end{aligned}$$

$$\begin{aligned} &= \int_{\Omega_n} (\alpha(\boldsymbol{x}+h_n\boldsymbol{Q}))^{-d} K(\boldsymbol{Q}(\alpha(\boldsymbol{x}+h_n\boldsymbol{Q}))^{-1}) f(\boldsymbol{x}+h_n\boldsymbol{Q}) h_n \boldsymbol{Q} d\boldsymbol{Q} \\ &= h_n^2 \alpha(\boldsymbol{x})^3 G(\alpha, f, \boldsymbol{x}) + o(h_n^2 \mathbf{1}), \end{aligned} \tag{9}$$

where $\Omega_n = \{\boldsymbol{Q} : \boldsymbol{x} + h_n\boldsymbol{Q} \in \text{supp}(f)\}$.

It is easy to see

$$\begin{aligned} &\text{Var}\left[K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})(\boldsymbol{X}_1-\boldsymbol{x})\right] \\ &\quad = \text{E}\left\{\left[K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})(\boldsymbol{X}_1-\boldsymbol{x})\right]\left[K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})(\boldsymbol{X}_1-\boldsymbol{x})\right]^T\right\} \\ &\qquad - \left\{\text{E}\left[K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})(\boldsymbol{X}_1-\boldsymbol{x})\right]\right\}\left\{\text{E}\left[K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})(\boldsymbol{X}_1-\boldsymbol{x})\right]\right\}^T. \end{aligned} \tag{10}$$

Again by the continuity of f, K and α, we have

$$\begin{aligned} &\text{E}\left\{\left[K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})(\boldsymbol{X}_1-\boldsymbol{x})\right]\left[K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})(\boldsymbol{X}_1-\boldsymbol{x})\right]^T\right\} \\ &= \text{E}\left[K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})^2(\boldsymbol{X}_1-\boldsymbol{x})(\boldsymbol{X}_1-\boldsymbol{x})^T\right] \\ &= \int_{\text{supp}(f)} \left[h_n^{-d}(\alpha(\boldsymbol{x}))^{-d}K(h_n^{-1}(\alpha(\boldsymbol{X}_1))^{-1}(\boldsymbol{X}_1-\boldsymbol{x}))\right]^2 (\boldsymbol{X}_1-\boldsymbol{x})(\boldsymbol{X}_1-\boldsymbol{x})^T f(\boldsymbol{X}_1) d\boldsymbol{X}_1 \\ &= h_n^{-d+2}\int_{\Omega_n} (\alpha(\boldsymbol{x}+h_n\boldsymbol{Q}))^{-d}K(\boldsymbol{Q}(\alpha(\boldsymbol{x}+h_n\boldsymbol{Q}))^{-1})^2 f(\boldsymbol{x}+h_n\boldsymbol{Q})\boldsymbol{Q}\boldsymbol{Q}^T d\boldsymbol{Q} \\ &= h_n^{-d+2}\int_{\Omega_n} (\alpha(\boldsymbol{x}))^{-d}K(\boldsymbol{Q}(\alpha(\boldsymbol{x}))^{-1})^2 f(\boldsymbol{x})\boldsymbol{Q}\boldsymbol{Q}^T d\boldsymbol{Q} + O(h_n^{-d+2}\mathbf{1}) = O(h_n^{-d+2}\mathbf{1}). \end{aligned} \tag{11}$$

Therefore we have

$$O_p(\sqrt{n^{-1}\boldsymbol{\Psi}}) = o_p(h_n^2\mathbf{1}). \tag{12}$$

Then (3) follows from (8)–(12).

Proof of (6)

It is straightforward to show that

$$n^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T \boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{Q}_m(\boldsymbol{x})$$

$$= \begin{pmatrix} n^{-1}\sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})^T\boldsymbol{H}_m(\boldsymbol{z}_i(\boldsymbol{x},\boldsymbol{X}_i))(\boldsymbol{X}_i-\boldsymbol{x}) \\ n^{-1}\sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})^T\boldsymbol{H}_m(\boldsymbol{z}_i(\boldsymbol{x},\boldsymbol{X}_i))(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x}) \end{pmatrix},$$

$$n^{-1}\sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})^T\boldsymbol{H}_m(\boldsymbol{z}_i(\boldsymbol{x},\boldsymbol{X}_i))(\boldsymbol{X}_i-\boldsymbol{x})$$

$$= h_n^2 f(\boldsymbol{x})\mu_2(K)(\alpha(\boldsymbol{x}))^2 s(\boldsymbol{H}_m(\boldsymbol{x})) + o_p(h_n^2\mathbf{1}),$$

$$\text{and}\quad n^{-1}\sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})^T\boldsymbol{H}_m(\boldsymbol{z}_i(\boldsymbol{x},\boldsymbol{X}_i))(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x}) = O_p(h_n^3\mathbf{1}).$$

Therefore (6) holds.

Proof of (7)

It is obvious that

$$\mathrm{E}\left[n^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{u}\right] = \mathrm{E}\left[n^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{u}\mid \boldsymbol{x}\right] = \mathbf{0},$$

$$\text{and}\qquad n^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{u} = \begin{pmatrix} n^{-1}\sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})u_i \\ n^{-1}\sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})u_i \end{pmatrix}.$$

By the continuity of f, K, σ^2 and α, we have

$$\begin{aligned} \mathrm{Var}&\left[n^{-1}\sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})u_i\right] = n^{-1}\mathrm{Var}\left[K_{h_n\alpha(\boldsymbol{X}_1)}(\boldsymbol{X}_1-\boldsymbol{x})u_1\right] \\ &= n^{-1}\int_{\mathrm{supp}(f)}\left[h_n^{-d}(\alpha(\boldsymbol{x}))^{-d}K(h_n^{-1}(\alpha(\boldsymbol{X}_1))^{-1}(\boldsymbol{X}_1-\boldsymbol{x}))\right]^2\sigma^2(\boldsymbol{X}_1)f(\boldsymbol{X}_1)d\boldsymbol{X}_1 \\ &= n^{-1}h_n^{-d}\int_{\Omega_n}((\alpha(\boldsymbol{x}+h_n\boldsymbol{Q}))^{-d}K(\boldsymbol{Q}(\alpha(\boldsymbol{x}+h_n\boldsymbol{Q}))^{-1}))^2\sigma^2(\boldsymbol{x}+h_n\boldsymbol{Q})f(\boldsymbol{x}+h_n\boldsymbol{Q})d\boldsymbol{Q} \\ &= n^{-1}h_n^{-d}\int_{\Omega_n}((\alpha(\boldsymbol{x}))^{-d}K(\boldsymbol{Q}(\alpha(\boldsymbol{x}))^{-1}))^2\sigma^2(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{Q} + o(n^{-1}h_n^{-d}) \\ &= n^{-1}h_n^{-d}R(K)(\alpha(\boldsymbol{x}))^{-d}\sigma^2(\boldsymbol{x})f(\boldsymbol{x}) + o(n^{-1}h_n^{-d}), \end{aligned}$$

$$\text{and Var}\left[n^{-1}\sum_{i=1}^n K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})u_i\right] = o(n^{-1}h_n^{-d+2}\mathbf{1}).$$

Then (7) holds.

Proof of Theorem 1

By (5) and (6), we have

$$\begin{aligned}\mathrm{E}\left[\hat{m}_n(\boldsymbol{x},h_n,\alpha)\mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\right]-m(\boldsymbol{x}) &= 0.5e_1^T(\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{X}_{\boldsymbol{x}})^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{Q}_m(\boldsymbol{x})\\ &= 0.5h_n^2\mu_2(K)(\alpha(\boldsymbol{x}))^2 s(\boldsymbol{H}_m(\boldsymbol{x}))+o_p(h_n^2).\end{aligned}$$

Therefore Theorem 1(1) holds.

Let $\boldsymbol{V}=\mathrm{diag}\left\{\sigma^2(\boldsymbol{X}_1),\ldots,\sigma^2(\boldsymbol{X}_n)\right\}$. Then

$$\mathrm{Var}\left[\hat{m}_n(\boldsymbol{x},h_n,\alpha)\mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\right]=\boldsymbol{e}_1^T(\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{X}_{\boldsymbol{x}})^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{V}\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{X}_{\boldsymbol{x}}(\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{X}_{\boldsymbol{x}})^{-1}\boldsymbol{e}_1, \tag{13}$$

and

$$n^{-1}\boldsymbol{X}_{\boldsymbol{x}}^T\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{V}\boldsymbol{W}_{\boldsymbol{x},\alpha}\boldsymbol{X}_{\boldsymbol{x}}=\begin{pmatrix} a_{11}(\boldsymbol{x},h_n,\alpha) & (\boldsymbol{a}_{21}(\boldsymbol{x},h_n,\alpha))^T\\ \boldsymbol{a}_{21}(\boldsymbol{x},h_n,\alpha) & \boldsymbol{a}_{22}(\boldsymbol{x},h_n,\alpha)\end{pmatrix},$$

$$\begin{aligned}\text{where}\quad a_{11}(\boldsymbol{x},h_n,\alpha) &= n^{-1}\sum_{i=1}^n(K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x}))^2\sigma^2(\boldsymbol{X}_i)\\ \boldsymbol{a}_{21}(\boldsymbol{x},h_n,\alpha) &= n^{-1}\sum_{i=1}^n(K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x}))^2(\boldsymbol{X}_i-\boldsymbol{x})\sigma^2(\boldsymbol{X}_i)\\ \text{and}\quad \boldsymbol{a}_{22}(\boldsymbol{x},h_n,\alpha) &= n^{-1}\sum_{i=1}^n(K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x}))^2(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})^T\sigma^2(\boldsymbol{X}_i).\end{aligned}$$

It is easy to prove that

$$\begin{aligned} & n^{-1}\sum_{i=1}^n(K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x}))^2\sigma^2(\boldsymbol{X}_i)=h_n^{-d}R(K)(\alpha(\boldsymbol{x}))^{-d}\sigma^2(\boldsymbol{x})f(\boldsymbol{x})o_p(h_n^{-d}),\\ & n^{-1}\sum_{i=1}^n(K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x}))^2(\boldsymbol{X}_i-\boldsymbol{x})^T\sigma^2(\boldsymbol{X}_i)=O_p(h_n^{-d+1}\boldsymbol{1})\\ \text{and}\quad & n^{-1}\sum_{i=1}^n(K_{h_n\alpha(\boldsymbol{X}_i)}(\boldsymbol{X}_i-\boldsymbol{x}))^2(\boldsymbol{X}_i-\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})^T\sigma^2(\boldsymbol{X}_i)=O_p(h_n^{-d+2}\boldsymbol{1}).\end{aligned} \tag{14}$$

By (13)–(14) and (5), we have

$$\mathrm{Var}\left[\hat{m}_n(\boldsymbol{x},h_n,\alpha)\mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\right]=n^{-1}h_n^{-d}R(K)(\alpha(\boldsymbol{x}))^{-d}\sigma^2(\boldsymbol{x})f(\boldsymbol{x})^{-1}+o_p(n^{-1}h_n^{-d}).$$

Therefore Theorem 1(2) holds.

By (6) and (7) and the central limit theorem, we have

$$n^{2/(d+4)} n^{-1} \boldsymbol{X}_{\boldsymbol{x}}^T \boldsymbol{W}_{\boldsymbol{x},\alpha} \left[0.5\boldsymbol{Q}_m(\boldsymbol{x}) + \boldsymbol{u}\right] \xrightarrow{d} \begin{pmatrix} N(0.5c^2\mu_2(K)(\alpha(\boldsymbol{x}))^2 f(\boldsymbol{x}) s(\boldsymbol{H}_m(\boldsymbol{x})), c^{-d} R(K)(\alpha(\boldsymbol{x}))^{-d} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x})) \\ \boldsymbol{0} \end{pmatrix}.$$

Applying White (1984, Proposition 2.26) and (5), we can easily deduce Theorem 1(3).

Tables

Table 1: *The percentage of 1000 samples in which the variable bandwidth method is better than the constant bandwidth method, and the RMSE of the two methods.*

Model	**Percentage better**	**Root mean squared error**	
		Constant bandwidth	**Variable bandwidth**
Model A	75.0	1.3581	1.1150
Model B	63.6	0.4991	0.4347
Model C	75.7	0.9739	0.7995
Model D	68.5	0.9737	0.8524
Model E	80.0	1.3641	1.1009

Table 2: *The percentage of 200 samples in which the variable bandwidth method is better than the constant bandwidth method, and the RMSE of the two methods.*

Model	**Percentage better**	**Root mean squared error**	
		Constant bandwidth	**Variable bandwidth**
Model F	54.6	0.0411	0.0397
Model G	54.0	0.0843	0.0816
Model H	53.0	0.0935	0.0814
Model I	52.0	0.0999	0.0907

Figures

Figure 1: *Results for the simulated univariate regression data of models A–E. The true regression functions (the solid line) and four typical estimated curves are presented. These correspond to the 10th, the 30th, the 70th, the 90th percentile. The dashed line is for the constant bandwidth method and the dotted line is for the variable bandwidth method.*

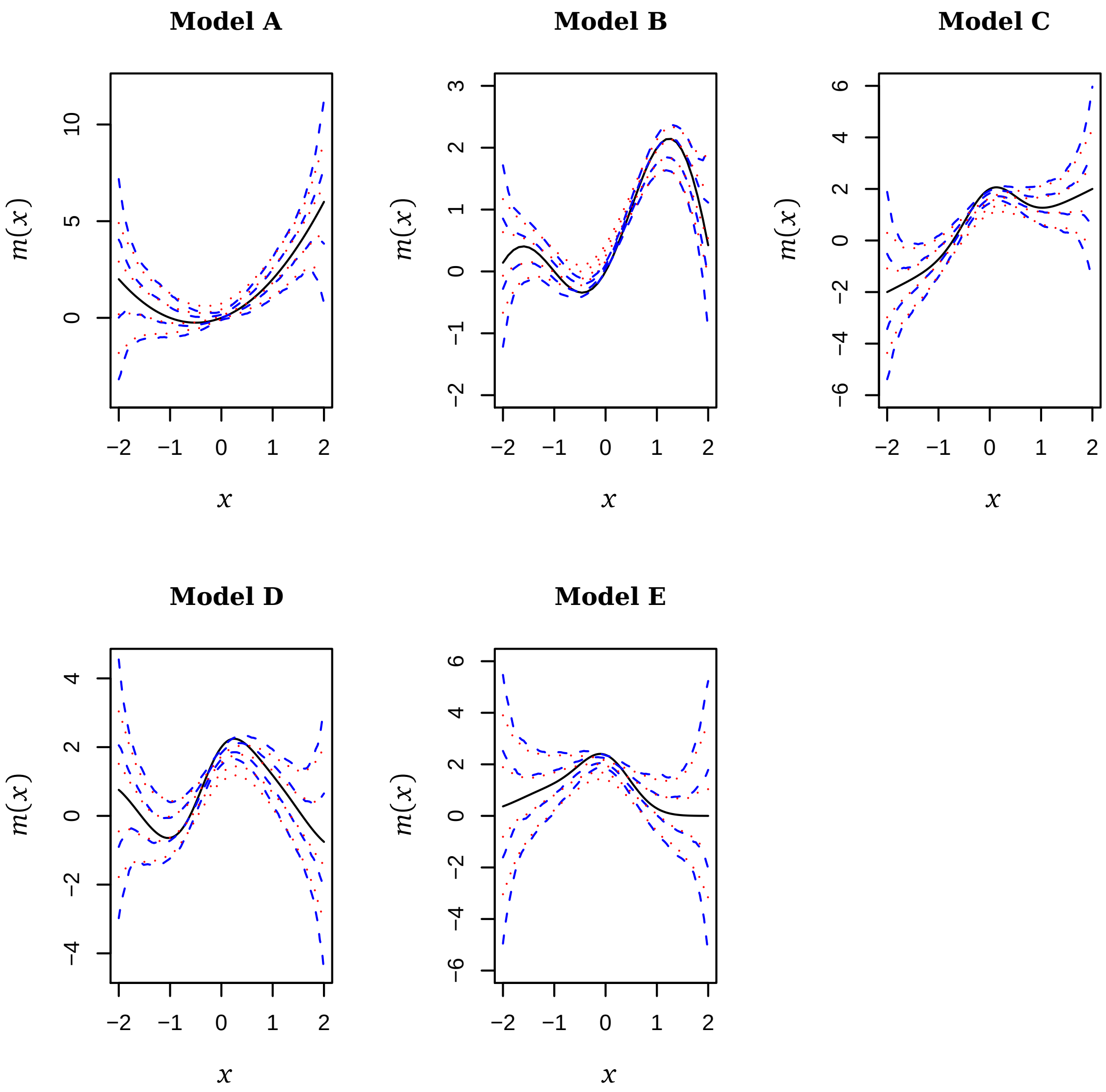

Figure 2: *Results for the simulated bivariate data of model F. The true regression functions (the solid line) and four typical estimated curves are presented. These correspond to the 10th, the 30th, the 70th, the 90th percentile. The dashed line is for the constant bandwidth method and the dotted line is for the variable bandwidth method.*

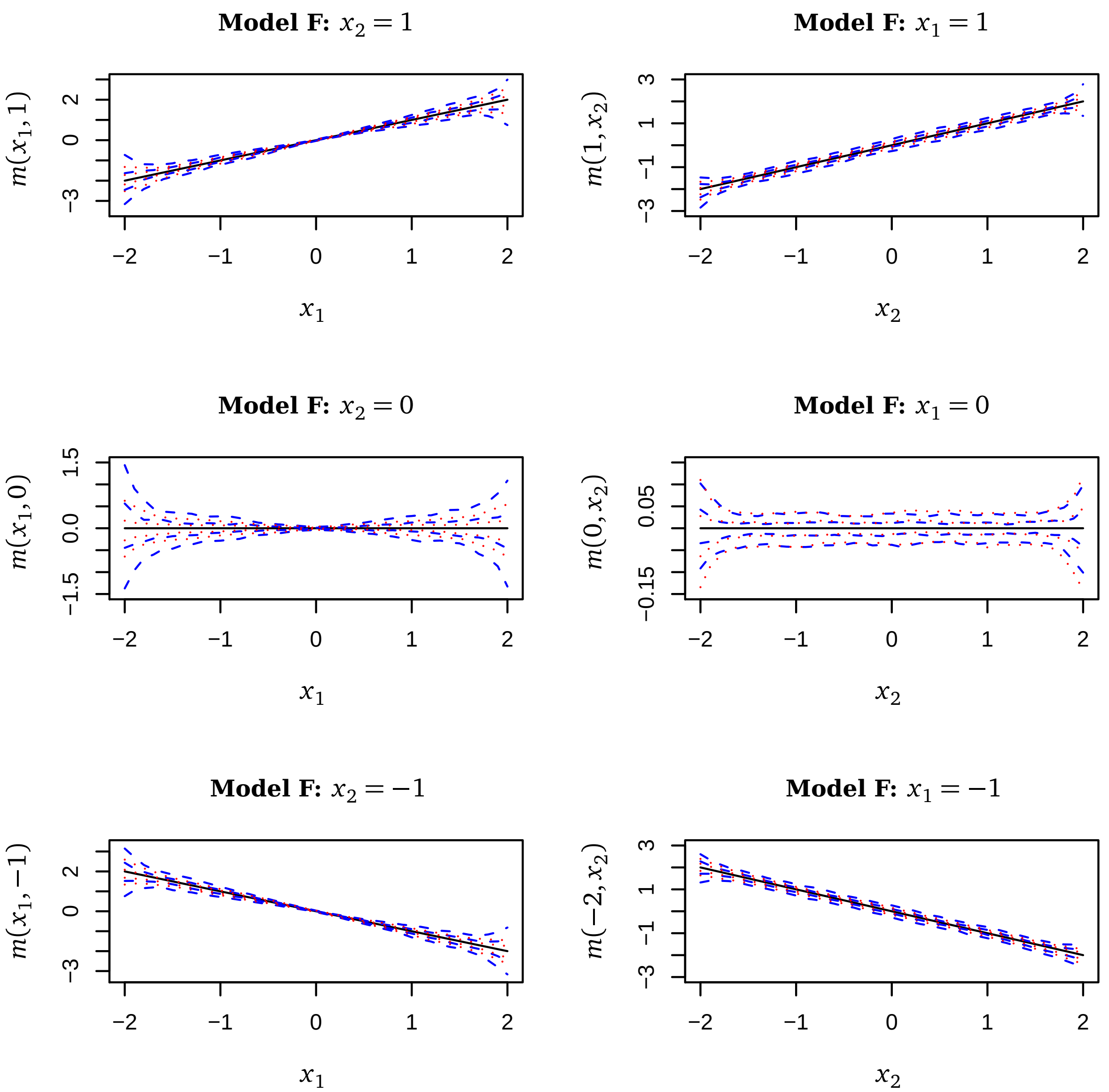

Figure 3: *Results for the simulated bivariate data of model G. The true regression functions (the solid line) and four typical estimated curves are presented. These correspond to the 10th, the 30th, the 70th, the 90th percentile. The dashed line is for the constant bandwidth method and the dotted line is for the variable bandwidth method.*

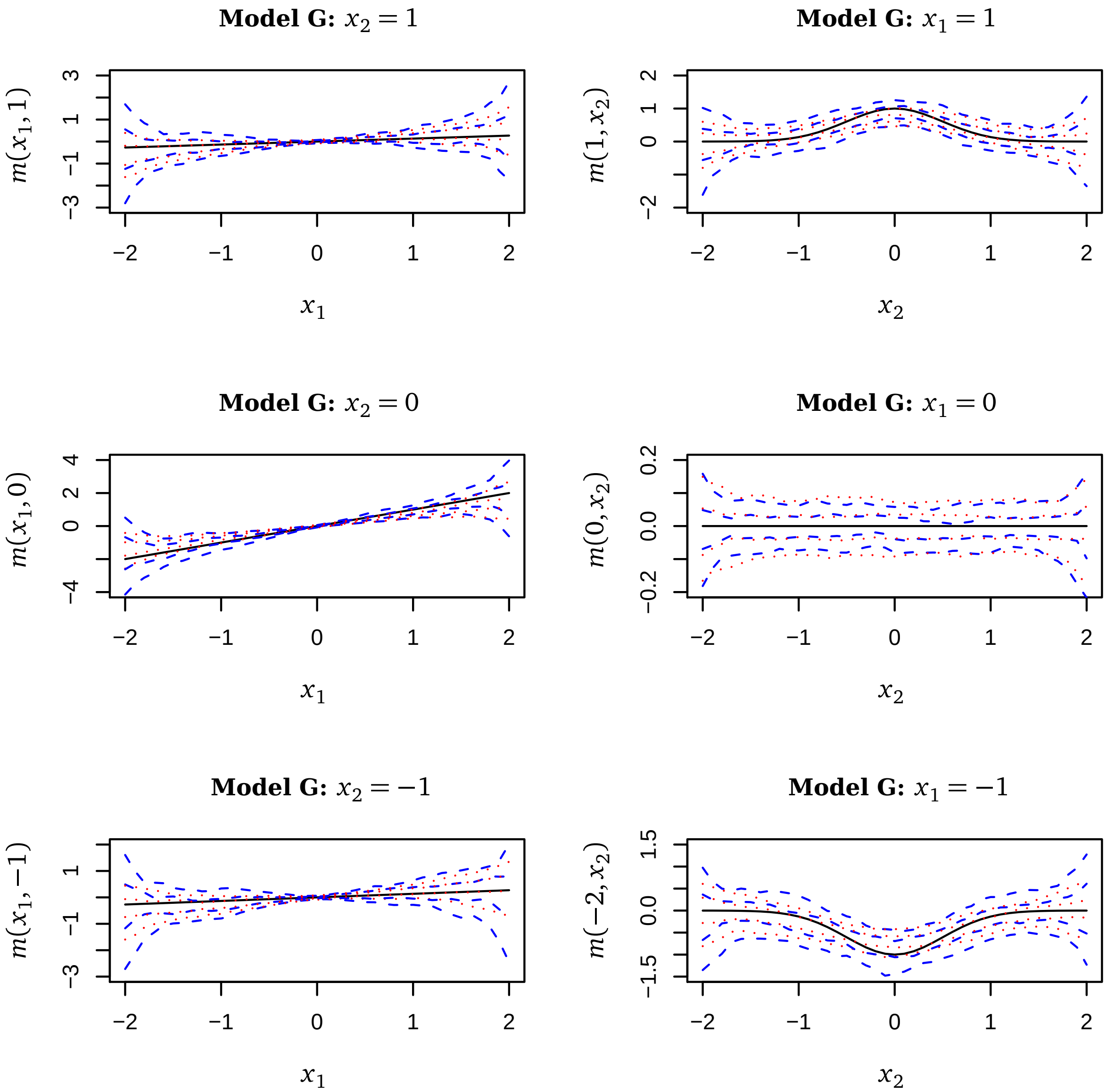

Figure 4: *Results for the simulated bivariate data of model H. The true regression functions (the solid line) and four typical estimated curves are presented. These correspond to the 10th, the 30th, the 70th, the 90th percentile. The dashed line is for the constant bandwidth method and the dotted line is for the variable bandwidth method.*

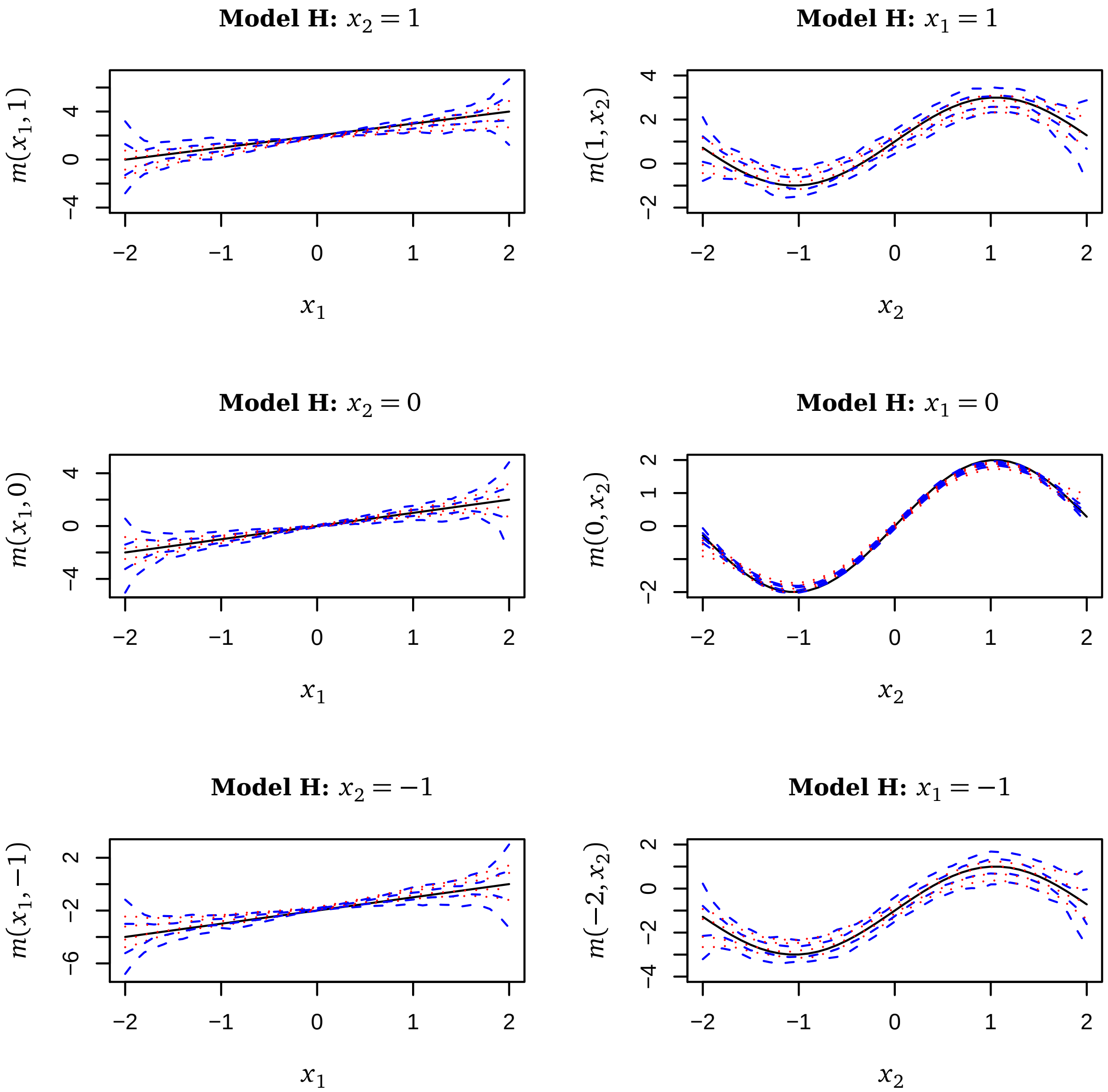

Figure 5: *Results for the simulated bivariate data of model I. The true regression functions (the solid line) and four typical estimated curves are presented. These correspond to the 10th, the 30th, the 70th, the 90th percentile. The dashed line is for the constant bandwidth method and the dotted line is for the variable bandwidth method.*