Statistica Sinica **25** (2015), 1377-1397
doi:http://dx.doi.org/10.5705/ss.2013.010

OPTIMAL DIFFERENCE-BASED VARIANCE ESTIMATION IN HETEROSCEDASTIC NONPARAMETRIC REGRESSION

Yuejin Zhou, Yebin Cheng, Lie Wang and Tiejun Tong

Anhui University of Science and Technology, Donghua University, Massachusetts Institute of Technology and Hong Kong Baptist University

Abstract: Estimating the residual variance is an important question in nonparametric regression. Among the existing estimators, the optimal difference-based variance estimation proposed in Hall, Kay, and Titterington (1990) is widely used in practice. Their method is restricted to the situation when the errors are independent and identically distributed. In this paper, we propose the optimal difference-based variance estimation in heteroscedastic nonparametric regression under settings of uncorrelated errors and correlated errors. The proposed estimators are shown to be asymptotically unbiased and their mean squared errors are derived. Simulation studies indicate that the proposed estimators perform better than existing competitors in finite sample settings. In addition, data examples are analyzed to demonstrate the practical usefulness of the proposed method. The proposed method has many applications and we apply it to the nonparametric regression model with repeated measurements and the semiparametric partially linear model.

Key words and phrases: Difference-based estimator, heteroscedasticity, homoscedasticity, nonparametric regression, repeated measurements, variance estimation.

1. Introduction

Consider the nonparametric regression model

$$Y_i = f(x_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{1.1}$$

where Y_i are observations, f is an unknown mean function, x_i are design points, and ε_i are random errors. When the errors ε_i are assumed to be independent and identically distributed (i.i.d.) with zero mean and variance σ^2. This is a homoscedastic nonparametric regression model. The estimation of σ^2 has received a lot of attention in the literature. An accurate estimate of σ^2 is needed for the purpose of constructing confidence bands, choosing the amount of smoothing, testing the goodness of fit, and for other uses (Carroll (1987), Carroll and Ruppert (1988), Kay (1988), Buckley, Eagleson, and Silverman (1988), Eubank and Spiegelman (1990), Gasser, Kneip, and Kohler (1991), Härdle and Tsybakov (1997), Kulasekera and Gallagher (2002)).

The existing estimators of σ^2 fall into two classes. The first consists of residual-based estimators and the second consists of difference-based estimators. Difference-based estimators do not require an estimate of the mean function, and so is popular in practice. Rice (1984) proposed a first-order difference-based estimator; Gasser, Sroka, and Jennen-Steinmetz (1986) constructed a second-order difference-based estimator; Hall, Kay, and Titterington (1990) proposed an mth-order difference-based estimator with $m \geq 2$ a fixed integer. Specifically, let $\{d_j,\ j = -m_1, \ldots, m_2\}$ be a sequence of real numbers that satisfy

$$\sum_{j=-m_1}^{m_2} d_j = 0 \qquad \text{and} \qquad \sum_{j=-m_1}^{m_2} d_j^2 = 1, \tag{1.2}$$

where m_1 and m_2 are non-negative integers, with $m = m_1 + m_2$ termed the order of the sequence. The estimator of Hall, Kay, and Titterington (1990) is

$$\hat{\sigma}^2_{\mathrm{H}} = \frac{1}{n-m} \sum_{k=m_1+1}^{n-m_2} \left(\sum_{j=-m_1}^{m_2} d_j Y_{j+k} \right)^2. \tag{1.3}$$

Without loss of generality, we take $m_1 = 0$ and $m_2 = m$. The optimal sequence $\{d_0, \ldots, d_m\}$ is then obtained by minimizing the asymptotic mean squared error (MSE) of the estimator $\hat{\sigma}^2_{\mathrm{H}}$. The optimal sequence $\{d_0, \ldots, d_m\}$ is unique except for the initial sign and reversal order.

Beyond the above estimators, there exists more complicated difference-based estimators in the literature. Nevertheless, most of them suffer from at least one of the following: the method applies only to restricted situations, e.g., the difference-based estimator in Dette, Munk, and Wagner (1998) only works for a small sample size or a rough mean function; the method is not easy to apply, e.g., the covariate-matched U-statistic estimator in Müller, Schick, and Wefelmeyer (2003) requires an appropriate choice of bandwidth in advance; the method requires rigorous assumptions, e.g, the least squares estimators in Tong and Wang (2005) and Tong, Ma, and Wang (2013) require that the mean function have a bounded second derivative and that the design points be equidistant.

The optimal difference-based estimator in Hall, Kay, and Titterington (1990) remains widely used in practice, and the optimal sequence idea has been used elsewhere. For instance, Levine (2006) and Brown and Levine (2007) applied the optimal sequence idea to estimate the variance function in nonparametric regression; Wang, Brown, and Cai (2011) and Zhao and You (2011) applied it in semiparametric partially linear regression models. Hall, Kay, and Titterington (1990) require a homoscedastic nonparametric regression model and, in practice, it is not uncommon that the errors have different variances or be correlated to each other. This motivates us to propose an optimal difference-based variance estimation in heteroscedastic nonparametric regression.

The rest of the paper is organized as follows. In Section 2, we propose the optimal difference-based estimator of σ^2 in heteroscedastic nonparametric regression under uncorrelated errors and under correlated errors. We derive the asymptotic MSEs of the proposed estimator, and present a simulation study to evaluate its finite sample performance. In Sections 3, we apply the proposed method to the nonparametric regression model with repeated measurements, and to the semiparametric partially linear model. Two data examples are presented and analyzed to demonstrate the practical usefulness of the proposed method. We conclude the paper in Section 4 and provide the technical proofs in Section 5.

2. Main Results

Let $\boldsymbol{Y}=(Y_1,\ldots,Y_n)^T$, $\boldsymbol{f}=(f_1,\ldots,f_n)^T$ with $f_i=f(x_i)$, and $\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_n)^T$. In matrix notation, the regression model is

$$\boldsymbol{Y} = \boldsymbol{f} + \boldsymbol{\varepsilon},$$

where $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\Sigma$. Hall, Kay, and Titterington (1990) assumed $\Sigma = I$. We extend the optimal difference-based variance estimation to a general covariance matrix Σ. We need assumptions on the design points, the mean function, and the random errors.

Assumption 1: The design points x_i satisfy $0 \le x_1 \le \cdots \le x_n \le 1$ and $x_i - x_{i-1} = 1/n + o(1/n)$.

Assumption 2: The mean function $f(x)$ satisfies the following Lipschitz condition $|f(x_j) - f(x_i)| \le M|x_j - x_i|$, where $M > 0$ is a constant.

Assumption 3: $E(\varepsilon_i^4) < \infty$.

2.1. Uncorrelated errors

We first consider uncorrelated but heteroscedastic errors. Specifically, we take

$$\Sigma = \begin{pmatrix} c_1 & 0 & \ldots & 0 \\ 0 & c_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & c_n \end{pmatrix}, \tag{2.1}$$

where the constants $c_i > 0$ are assumed to be known. Without loss of generality, we assume that $\sum_{i=1}^n c_i = n$.

When $\Sigma = I$, the expectation of the estimator at (1.3) is

$$E(\hat{\sigma}^2_{\mathrm{H}}) = \sigma^2 + \frac{1}{n-m}\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m} d_j f_{j+k}\Big)^2.$$

Under Assumptions 1 and 2, $\sum_{k=1}^{n-m}(\sum_{j=0}^{m} d_j f_{j+k})^2/(n-m) = O(m^3/n^2)$. Thus $\hat\sigma^2_{\rm H}$ is an asymptotically unbiased estimator σ^2 in homoscedastic nonparametric regression when $m = n^r$ with $0 \le r < 2/3$. When $\Sigma \ne I$, we have

$$E(\hat\sigma^2_{\rm H}) = \Big[1 + \frac{1}{n-m}\Big\{m - \sum_{j=0}^{m} d_j^2\Big(\sum_{i=1}^{j} c_i + \sum_{i=j+n-m+1}^{n} c_i\Big)\Big\}\Big]\sigma^2 + \frac{1}{n-m}\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m} d_j f_{j+k}\Big)^2,$$

where $\sum_{i=1}^{0} c_i = 0$ and $\sum_{i=n+1}^{n} c_i = 0$. Here the bias term for $\hat\sigma^2_{\rm H}$ is non-negligible when the c_i values, with i close to 1 or n, are significantly far from 1. We propose a new difference-based estimator of σ^2 that is asymptotically unbiased.

Let $\omega_j = \sum_{k=1}^{n-m} c_{j+k}/(n-m)$ for $j = 0, \ldots, m$. Define a new estimator as

$$\hat\sigma^2_{\rm new} = \frac{1}{n-m}\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m}\frac{d_j}{\sqrt{\omega_j}}Y_{j+k}\Big)^2.$$

Under mild conditions, we can verify that $\hat\sigma^2_{\rm new}$ is an asymptotically unbiased estimator of σ^2. For ease of notation, let $d'_j = d_j/\sqrt{w_j}$ for $j = 0, \ldots, m$. Then

$$\hat\sigma^2_{\rm new} = \frac{1}{n-m}\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m} d'_j Y_{j+k}\Big)^2, \tag{2.2}$$

where the difference sequence $\{d'_j\}$ satisfies

$$\sum_{j=0}^{m} d'_j = 0 \quad \text{and} \quad \sum_{j=0}^{m}\omega_j d'^2_j = 1. \tag{2.3}$$

Note that the sequence $\{d'_j\}$ in (2.3) differs from the sequence $\{d_j\}$ in (1.2) except when $\Sigma = I$. With the sequence $\{d'_j\}$, the proposed estimator $\hat\sigma^2_{\rm new}$ has the quadratic form,

$$\hat\sigma^2_{\rm new} = \frac{1}{n-m}\boldsymbol{Y}^T D^T D\boldsymbol{Y} = \frac{1}{n-m}\boldsymbol{Y}^T A\boldsymbol{Y},$$

where $A = D^T D$ and D is an $(n-m)\times n$ matrix of form

$$D = \begin{pmatrix} d'_0 & d'_1 & d'_2 & \cdots & d'_m & 0 & \cdots & \cdots & 0 \\ 0 & d'_0 & d'_1 & d'_2 & \cdots & d'_m & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & d'_0 & d'_1 & d'_2 & \cdots & d'_m & 0 \\ 0 & \cdots & \cdots & 0 & d'_0 & d'_1 & d'_2 & \cdots & d'_m \end{pmatrix}.$$

Theorem 1. *If* Σ *at* (2.1) *has* $\min\{w_0,\ldots,w_m\}\geq c_0$, *where* $c_0>0$ *is a constant independent of* m *and* n, *under Assumptions* 1–3, *for any* $m=n^r$ *with* $0<r<2/7$,

$$\text{MSE}(\hat{\sigma}^2_{\text{new}})=\frac{1}{(n-m)^2}\Big\{\Big(\sum_{j=0}^{m}{d'_j}^2\Big)^2\sum_{i=1}^{n}\text{var}(\varepsilon_i^2)+4\sigma^4\sum_{k=1}^{m}b_k\Big(\sum_{j=0}^{m-k}d'_jd'_{j+k}\Big)^2\Big\}+o\Big(\frac{1}{n}\Big), \tag{2.4}$$

$$\text{MSE}(\hat{\sigma}^2_{\text{H}})=\frac{1}{(n-m)^2}\Big[\sum_{i=1}^{n}\text{var}(\varepsilon_i^2)+4\sum_{k=1}^{m}b_k\Big(\sum_{j=0}^{m-k}d_jd_{j+k}\Big)^2\sigma^4+\Big\{\sum_{k=1}^{n-m}\sum_{j=0}^{m}d_j^2c_{j+k}-(n-m)\Big\}^2\sigma^4\Big]+o\Big(\frac{1}{n}\Big), \tag{2.5}$$

where $b_k=\sum_{i=1}^{n-k}c_ic_{i+k}$ *for* $k=1,\ldots,m$.

The proof of Theorem 1 is given in Section 5. When the ε_i are normally distributed, $\text{var}(\varepsilon_i^2)=2c_i^2\sigma^4$, and the optimal sequence $\{d'_j\}$ is obtained by minimizing

$$\sum_{i=1}^{n}c_i^2\Big(\sum_{j=0}^{m}{d'_j}^2\Big)^2+2\sum_{k=1}^{m}b_k\Big(\sum_{j=0}^{m-k}d'_jd'_{j+k}\Big)^2. \tag{2.6}$$

For $m=1$, the optimal sequence is $\{d'_0,d'_1\}=\{1/\sqrt{\omega_0+\omega_1},-1/\sqrt{\omega_0+\omega_1}\}$. For $m\geq 2$, the optimal sequence can be numerically computed by the Lagrange multiplier method. When the ε_i are non-normally distributed, the optimal sequence $\{d'_j\}$ is obtained by minimizing

$$\sum_{i=1}^{n}(\gamma_i-1)c_i^2\Big(\sum_{j=0}^{m}{d'_j}^2\Big)^2+4\sum_{k=1}^{m}b_k\Big(\sum_{j=0}^{m-k}d'_jd'_{j+k}\Big)^2, \tag{2.7}$$

where $\gamma_i=E(\varepsilon_i^4)/(c_i^2\sigma^4)$ for $i=1,\ldots,n$. In general, we need estimates of γ_i to substitute into (2.7) to obtain the optimal sequence $\{d'_j\}$. Accordingly, we can use (2.5) to obtain the optimal sequence of $\{d_j\}$ in Hall's estimator $\hat{\sigma}^2_{\text{H}}$.

2.2. Correlated errors

We consider correlated errors. Specifically, let $\Sigma=(c_{jk})_{n\times n}$ be a known non-diagonal matrix. Under this setting, the expectation of Hall's estimator is

$$E(\hat{\sigma}^2_{\text{H}})=\frac{1}{n-m}\Big\{\sigma^2\sum_{v=1}^{n-m}\sum_{k=0}^{m}\sum_{j=0}^{m}d_kd_jc_{j+v,k+v}+\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m}d_jf_{j+k}\Big)^2\Big\}.$$

In general, $\hat{\sigma}^2_{\rm H}$ is an asymptotically biased estimator of σ^2.

Let $\omega = \sum_{v=1}^{n-m}\sum_{k=0}^{m}\sum_{j=0}^{m} d_k d_j c_{j+v,k+v}/(n-m)$ and $\tilde{d}_j = d_j/\sqrt{\omega}$ for $j = 0,\ldots,m$. Take

$$\tilde{\sigma}^2_{\rm new} = \frac{1}{n-m}\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m}\tilde{d}_j Y_{j+k}\Big)^2, \tag{2.8}$$

where the difference sequence $\{\tilde{d}_j\}$ satisfies

$$\sum_{j=0}^{m}\tilde{d}_j = 0 \quad\text{and}\quad \sum_{j=0}^{m}\omega\tilde{d}_j^2 = 1. \tag{2.9}$$

Then $\tilde{\sigma}^2_{\rm new}$ can be written as the quadratic form $\boldsymbol{Y}^T\tilde{A}\boldsymbol{Y}/(n-m)$, where $\tilde{A} = \tilde{D}^T\tilde{D}$ and $\tilde{D}$ is a matrix that replaces d'_j in matrix D by $\tilde{d}_j$ accordingly. Under Assumptions 1 and 2, we can verify that $E(\tilde{\sigma}^2_{\rm new}) = \sigma^2 + O(1/n^2)$ for any fixed m. The asymptotic MSE of $\tilde{\sigma}^2_{\rm new}$ is stated in a theorem.

Theorem 2. *Let Σ be a known non-diagonal positive semi-definite matrix and suppose Assumptions* 1−3 *hold. Then for any fixed m,*

$$\begin{aligned}\mathrm{MSE}(\tilde{\sigma}^2_{\rm new}) &= \frac{1}{(n-m)^2}E\Big\{\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m}\tilde{d}_j\varepsilon_{j+k}\Big)^2\Big\}^2 \\ &\quad -\frac{\sigma^4}{(n-m)^2}\Big(\sum_{v=1}^{n-m}\sum_{k=0}^{m}\sum_{j=0}^{m}\tilde{d}_k\tilde{d}_j c_{j+v,k+v}\Big)^2 + O\Big(\frac{1}{n^2}\Big),\\ \mathrm{MSE}(\hat{\sigma}^2_{\rm H}) &= \frac{1}{(n-m)^2}E\Big\{\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m}d_j\varepsilon_{j+k}\Big)^2\Big\}^2 \\ &\quad -\frac{\sigma^4}{(n-m)}\Big\{2\sum_{v=1}^{n-m}\sum_{k=0}^{m}\sum_{j=0}^{m}d_k d_j c_{j+v,k+v} - (n-m)\Big\} + O\Big(\frac{1}{n^2}\Big).\end{aligned}$$

The proof is given in Section 5. Here the optimal sequence $\{\tilde{d}_j\}$ is obtained by minimizing

$$E\Big\{\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m}\tilde{d}_j\varepsilon_{j+k}\Big)^2\Big\}^2 - \sigma^4\Big(\sum_{v=1}^{n-m}\sum_{k=0}^{m}\sum_{j=0}^{m}\tilde{d}_k\tilde{d}_j c_{j+v,k+v}\Big)^2.$$

For $m = 1$, the optimal sequence is $\{\tilde{d}_0,\tilde{d}_1\} = \{1/\sqrt{2\omega}, -1/\sqrt{2\omega}\}$, where $\omega = \sum_{v=1}^{n-1}(c_{v,v}+c_{v+1,v+1}-c_{v,v+1}-c_{v+1,v})/2(n-1)$. For $m \ge 2$, the optimal sequence can be numerically computed by the Lagrange multiplier method.

When Σ has a special structure, one can say more. For instance, when Σ is the Toeplitz matrix

$$\Sigma = \begin{pmatrix} a_0 & a_1 & a_2 & \dots & \dots & a_{n-1} \\ a_1 & a_0 & a_1 & a_2 & \dots & a_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ a_{n-2} & a_{n-3} & \dots & \dots & a_0 & a_1 \\ a_{n-1} & a_{n-2} & a_{n-3} & \dots & a_1 & a_0 \end{pmatrix},$$

it can be shown that the optimal sequence $\{\tilde{d}_j\}$ is obtained by minimizing

$$E\bigg\{\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m}\tilde{d}_j\varepsilon_{j+k}\Big)^2\bigg\}^2 - (n-m)^2\Big(a_0\sum_{j=0}^{m}\tilde{d}_j^2 + 2\sum_{0\le s<t\le m}\tilde{d}_s\tilde{d}_t a_{t-s}\Big)^2\sigma^4.$$

In particular, for $m=1$ the optimal sequence is given as $\{\tilde{d}_0, \tilde{d}_1\} = \{1/\sqrt{2(a_0-a_1)}, -1/\sqrt{2(a_0-a_1)}\}$. The higher-order optimal sequence, accordingly, can be obtained numerically by the Lagrange multiplier method.

Consider the first-order autoregressive model

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \qquad t = 1, 2, \dots, \tag{2.10}$$

where ε_t are the i.i.d. random errors with mean zero and variance σ^2, with ρ known. Let $\boldsymbol{Y} = (Y_{t+1}, Y_{t+2}, \dots, Y_{t+k})^T$, $\boldsymbol{f} = (f(Y_t), \dots, f(Y_{t+k-1}))^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_{t+1}, \varepsilon_{t+2}, \dots, \varepsilon_{t+k})^T$. In matrix notation, the model can be written as $\boldsymbol{Y} = \boldsymbol{f} + \boldsymbol{\varepsilon}$, where $E(\boldsymbol{\varepsilon}) = 0$, $\mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2\Sigma$ with

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{k-1} \\ \rho & 1 & \rho & \dots & \rho^{k-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{k-1} & \rho^{k-2} & \rho^{k-3} & \dots & 1 \end{pmatrix}.$$

Let $\omega = 1 + 2\sum_{0\le s<t\le m} d_s d_t \rho^{t-s}$. We then estimate σ^2 by $\hat{\sigma}^2_{\mathrm{AR}} = \sum_{v=1}^{k-m} (\sum_{j=0}^{m}\tilde{d}_j Y_{j+v})^2/(k-m)$. The asymptotic MSE of the estimator $\hat{\sigma}^2_{\mathrm{AR}}$ is

$$\frac{1}{(k-m)^2}\bigg[E\bigg\{\sum_{v=1}^{k-m}\Big(\sum_{j=0}^{m}\tilde{d}_j\varepsilon_{j+v}\Big)^2\bigg\}^2 - (k-m)^2\Big(\sum_{j=0}^{m}\tilde{d}_j^2 + 2\sum_{0\le s<t\le m}\tilde{d}_s\tilde{d}_t\rho^{t-s}\Big)^2\sigma^4\bigg].$$

For $m = 1$, by minimizing its asymptotic MSE we have the optimal sequence as $\{\tilde{d}_0, \tilde{d}_1\} = \{1/\sqrt{2(1-\rho)}, -1/\sqrt{2(1-\rho)}\}$, consistent with what was found for a general Toeplitz matrix.

2.3. Simulation study

We conducted three simulation studies to evaluate the finite sample performance of the proposed estimators. We compared our proposed estimators with Hall's estimator $\hat{\sigma}^2_{\text{H1}}$ using the optimal sequence derived in Hall, Kay, and Titterington (1990), and Hall's estimator $\hat{\sigma}^2_{\text{H2}}$ using the optimal sequence derived in Theorem 1.

The first simulation study considered a regression model with uncorrelated errors. Let $x_i = i/n$ for $i = 1, \ldots, n$, and $m = 2$. We simulated ε_i independently from $N(0, c_i\sigma^2)$, where $c_i = nc'_i / \sum_{i=1}^n c'_i$ with c'_i being generated from the standard exponential distribution with parameter $\lambda = 1$, so $\sum_{i=1}^n c_i = n$ and $\Sigma = \text{diag}(c_1, \ldots, c_n)$. With $m = 2$, the optimal sequences of $\{d_0, d_1, d_2\}$ and $\{d'_0, d'_1, d'_2\}$ can be obtained by minimizing the corresponding quantities in Theorem 1 along with the information in Σ. We considered the mean functions $f_1(x) = 5x$, $f_2(x) = 5x(1-x)$, and $f_3(x) = 5\sin(2\pi x)$. We took $\sigma = 0.5$ and 2, corresponding to small and large variances, and $n = 20$, 30 and 40. In total, there are 18 combinations of simulation settings. For each setting, we repeated the simulation 1,000 times, and report the means, standard errors and MSEs for $\hat{\sigma}^2_{\text{H1}}$, $\hat{\sigma}^2_{\text{H2}}$ and $\hat{\sigma}^2_{\text{new}}$ in Table 1.

From Table 1, we see that $\text{MSE}(\hat{\sigma}^2_{\text{new}}) \le \text{MSE}(\hat{\sigma}^2_{\text{H2}}) \le \text{MSE}(\hat{\sigma}^2_{\text{H1}})$ for most cases, no matter whether the smoothness function is rough or the variance is large. The proposed estimator $\hat{\sigma}^2_{\text{new}}$, being asymptotically unbiased, provides a smaller standard error than the other two estimators. The biases of $\hat{\sigma}^2_{\text{H1}}$ and $\hat{\sigma}^2_{\text{H2}}$ are comparable, suggesting that $\hat{\sigma}^2_{\text{H2}}$ does not provide bias reduction compared to $\hat{\sigma}^2_{\text{H1}}$.

The second study considered unknown $c'_i s$. For the nonparametric regression $Y_i = f(x_i) + v(x_i)^{1/2}\varepsilon_i$ with ε_i i.i.d. random errors with mean 0 and variance 1, we took $v(x) = a + bx$. For estimating a and b, we let $s_i = (Y_{i+1} - Y_i)^2/2$ and $z_i = (x_i + x_{i+1})/2$, so $E(s_i) = a + bz_i + o(1)$. Then we fit s_i as a linear model of z_i, $s_i = a + bz_i + \epsilon_i$, where $i = 1, \ldots, n-1$. In matrix notation, $S = Z\beta + \epsilon$, with $S = (s_1, \ldots, s_{n-1})^T$, $\beta = (a, b)^T$, $\epsilon = (\epsilon_1, \ldots, \epsilon_{n-1})^T$, and Z an $(n-1) \times 2$ matrix with i-th row elements $(1, z_i)$. The least squares estimator of β is $\hat{\beta} = (\hat{a}, \hat{b})^T = (Z^TZ)^{-1}Z^TS$, where $\hat{a}$ and $\hat{b}$ are consistent estimators of a and b. Then $\hat{v}(x)$ is a consistent estimator of $v(x)$. With $\tilde{c}_i = \hat{v}(x_i)$, we have the estimator of c_i as $\hat{c}_i = n\tilde{c}_i / \sum_{i=1}^n \tilde{c}_i$.

To check whether $\hat{a}$ and $\hat{b}$ are consistently estimated, we conducted a simulation study with $x_i = i/n$ for $i = 1, \ldots, n$, and the mean functions $f_1(x) = 5x$, $f_2(x) = 5x(1-x)$ and $f_3(x) = 5\sin(2\pi x)$. The variance function $v(x)$ was $1+0.5x$ and the sample size was $n = 1{,}000$. We repeated the simulation 1,000 times for each setting and report the estimated values of a and b in Table 2.

Table 1. Means, standard errors (SE) and mean squared errors (MSE) of $\hat{\sigma}^2_{\mathrm{H1}}$, $\hat{\sigma}^2_{\mathrm{H2}}$, and $\hat{\sigma}^2_{\mathrm{new}}$ for the equidistant design with uncorrelated errors under various settings.

			$\hat{\sigma}^2_{\mathrm{H1}}$			$\hat{\sigma}^2_{\mathrm{H2}}$			$\hat{\sigma}^2_{\mathrm{new}}$		
n	σ	f	Mean	SE	MSE	Mean	SE	MSE	Mean	SE	MSE
20	0.5	f_1	0.3392	0.1267	0.0240	0.3293	0.1253	0.0220	0.3215	0.1218	0.0199
		f_2	0.2821	0.1264	0.0170	0.2798	0.1252	0.0165	0.2705	0.1215	0.0151
		f_3	1.6039	0.1880	1.8685	1.4667	0.1692	1.5090	1.4511	0.1749	1.4733
	2	f_1	4.2536	2.0105	4.1026	4.2307	1.9956	4.0317	4.0966	1.9344	3.7478
		f_2	4.1962	2.0098	4.0739	4.1837	1.9949	4.0095	4.0453	1.9338	3.7380
		f_3	5.5167	2.0792	6.6192	5.3592	2.0439	6.0211	5.2251	1.9930	5.4694
30	0.5	f_1	0.2915	0.1017	0.0120	0.2905	0.1016	0.0119	0.2768	0.0946	0.0096
		f_2	0.2668	0.1018	0.0106	0.2665	0.1016	0.0106	0.2481	0.0946	0.0089
		f_3	0.8867	0.1180	0.4193	0.8694	0.1172	0.3974	0.8052	0.1090	0.3201
	2	f_1	4.1435	1.6248	2.6580	4.1425	1.6226	2.6508	3.8635	1.5105	2.2981
		f_2	4.1187	1.6252	2.6530	4.1184	1.6231	2.6458	3.8346	1.5108	2.3077
		f_3	4.7389	1.6437	3.2451	4.7215	1.6407	3.2100	4.3919	1.5271	2.4834
40	0.5	f_1	0.2789	0.0892	0.0088	0.2768	0.0889	0.0086	0.2689	0.0868	0.0079
		f_2	0.2651	0.0894	0.0082	0.2643	0.0891	0.0081	0.2575	0.0869	0.0076
		f_3	0.6244	0.0953	0.1492	0.5889	0.0941	0.1237	0.5820	0.0908	0.1185
	2	f_1	4.1672	1.4286	2.0668	4.1629	1.4242	2.0531	4.0528	1.3902	1.9336
		f_2	4.1528	1.4292	2.0639	4.1499	1.4248	2.0506	4.0410	1.3903	1.9327
		f_3	4.5187	1.4355	2.3277	4.4806	1.4300	2.2739	4.3701	1.3964	2.0852

Table 2. The estimated values of a and b for various mean functions.

f	$\hat{a}$	$\hat{b}$
f_1	0.9938191	0.5083688
f_2	0.9938108	0.5083684
f_3	0.9940543	0.5083678

We looked at the proposed estimator with $v(x)$ estimated, in a simulation study with $n = 20$, 30 and 40 and the other settings those of the first study. With the estimated $\hat{c}_i$ values, for $m = 2$ we can obtain the optimal sequences of $\{d_0, d_1, d_2\}$ and $\{d'_0, d'_1, d'_2\}$ by minimizing the corresponding quantities in Theorem 1. We then repeated the simulation 1,000 times and report the means, standard errors and mean squared errors of the estimator of σ^2 in Table 3. There the proposed estimator outperforms the existing estimator with a parametric form of the variance function assumed.

A third study considered correlated errors, with Σ a $n \times n$ symmetric matrix, diagonal elements of 2.5 and off-diagonal elements of $0.9^{|i-j|}$ for $1 \le i, j \le n$ and $i \neq j$. We simulated the random errors ε from $N_n(0, \Sigma\sigma^2)$ independently. All other settings were the same as before. We chose $m = 2$ and compared $\tilde{\sigma}^2_{\mathrm{new}}$ in Section 2.2 with $\hat{\sigma}^2_{\mathrm{H1}}$ and $\hat{\sigma}^2_{\mathrm{H2}}$. The optimal sequences of $\{d_0, d_1, d_2\}$ and

Table 3. Means, standard errors (SE) and mean squared errors (MSE) of $\hat{\sigma}^2_{\mathrm{H1}}$, $\hat{\sigma}^2_{\mathrm{H2}}$ and $\hat{\sigma}^2_{\mathrm{new}}$ for the estimated variances under various settings.

			$\hat{\sigma}^2_{\mathrm{H1}}$			$\hat{\sigma}^2_{\mathrm{H2}}$			$\hat{\sigma}^2_{\mathrm{new}}$		
n	σ	f	Mean	SE	MSE	Mean	SE	MSE	Mean	SE	MSE
20	0.5	f_1	0.3868	0.1138	0.0316	0.3874	0.1138	0.0318	0.3119	0.0913	0.0121
		f_2	0.3294	0.1163	0.0198	0.3296	0.1164	0.0198	0.2647	0.0934	0.0089
		f_3	1.6480	0.1924	1.9914	1.6575	0.1931	2.0186	1.3476	0.1563	1.2292
	2	f_1	5.0178	1.8158	4.3301	5.0188	1.8161	4.3332	4.0274	1.4574	2.1226
		f_2	4.9592	1.8272	4.2554	4.9598	1.8275	4.2579	3.9793	1.4667	2.1495
		f_3	6.2667	1.9082	8.7760	6.2766	1.9097	8.8265	5.0503	1.5336	3.4528
30	0.5	f_1	0.3452	0.0976	0.0185	0.3496	0.0991	0.0197	0.2757	0.0782	0.0067
		f_2	0.3201	0.0975	0.0144	0.3244	0.0987	0.0152	0.2561	0.0779	0.0061
		f_3	0.9398	0.1175	0.4897	0.9591	0.1180	0.5168	0.7494	0.0928	0.2580
	2	f_1	4.9997	1.5595	3.4293	5.0589	1.5835	3.6265	3.9956	1.2502	1.5614
		f_2	4.9733	1.5592	3.3761	5.0337	1.5820	3.5688	3.9761	1.2490	1.5590
		f_3	5.5927	1.5796	5.0297	5.6659	1.6013	5.3371	4.4673	1.2638	1.8140
40	0.5	f_1	0.3299	0.0823	0.0131	0.3298	0.0823	0.0131	0.2642	0.0658	0.0045
		f_2	0.3165	0.0824	0.0112	0.3165	0.0824	0.0112	0.2534	0.0658	0.0043
		f_3	0.6728	0.0879	0.1865	0.6722	0.0879	0.1860	0.5381	0.0705	0.0880
	2	f_1	4.9879	1.3181	2.7117	4.9881	1.3181	2.7121	3.9961	1.0541	1.1100
		f_2	4.9753	1.3188	2.6888	4.9755	1.3188	2.6893	3.9855	1.0540	1.1102
		f_3	5.3293	1.3196	3.5069	5.3289	1.3196	3.5057	4.2701	1.0570	1.1891

$\{\tilde{d}_0, \tilde{d}_1, \tilde{d}_2\}$ were obtained by the Lagrange multiplier method. We repeated the simulation 1,000 times for each setting and report the means, standard errors and MSEs of the estimators in Table 4. In Table 4, our estimator performs better than its competitors under most settings.

3. Applications

In this section, we applied the proposed method to the nonparametric regression model with repeated measurements, and to the semiparametric partially linear model.

3.1. Nonparametric regression with repeated measurements

Consider the nonparametric regression model with repeated measurements

$$Y_{ij} = f(x_i) + \varepsilon_{ij}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, r_i, \tag{3.1}$$

where ε_{ij} are i.i.d. normal with zero mean and variance σ^2. All other settings are the same as those at (1.1).

To estimate σ^2, consider the within-design-point variation. Let $S_i^2 = \sum_{j=1}^{r_i} (Y_{ij} - \bar{Y}_i)^2/(r_i - 1)$ be the sample variance of the repeated observations in the ith design point, where $\bar{Y}_i = \sum_{j=1}^{r_i} Y_{ij}/r_i$. For the special case $r_i = 1$, we let

Table 4. Means, standard errors (SE) and mean squared errors (MSE) of $\hat{\sigma}^2_{\rm H1}$, $\hat{\sigma}^2_{\rm H2}$, and $\tilde{\sigma}^2_{\rm new}$ for the equidistant design with correlated errors, under various settings.

			$\hat{\sigma}^2_{\rm H1}$			$\hat{\sigma}^2_{\rm H2}$			$\tilde{\sigma}^2_{\rm new}$		
n	σ	f	Mean	SE	MSE	Mean	SE	MSE	Mean	SE	MSE
20	0.5	f_1	112.01	36.281	13807.1	0.3859	0.1734	0.0485	0.2346	0.1054	0.0113
		f_2	16.892	13.561	460.69	0.3861	0.1734	0.0485	0.2347	0.1054	0.0113
		f_3	3.2421	4.1069	25.803	0.4080	0.1736	0.0550	0.2480	0.1055	0.0111
	2	f_1	149.37	154.11	44863	6.1759	2.7753	12.429	3.7545	1.6871	2.9040
		f_2	59.341	80.683	9566.0	6.1760	2.7754	12.431	3.7546	1.6872	2.9041
		f_3	48.586	65.846	6319.4	6.1981	2.7757	12.528	3.7679	1.6874	2.8983
30	0.5	f_1	175.39	49.407	33113.8	0.3951	0.1404	0.0407	0.2402	0.0853	0.0073
		f_2	24.746	17.847	918.29	0.3951	0.1404	0.0407	0.2402	0.0853	0.0073
		f_3	3.6395	5.1569	38.056	0.3994	0.1405	0.0420	0.2428	0.0854	0.0073
	2	f_1	222.62	206.64	90457.1	6.3226	2.2473	10.440	3.8437	1.3662	1.8891
		f_2	76.119	102.57	15712.1	6.3226	2.2473	10.440	3.8437	1.3662	1.8891
		f_3	57.285	82.625	9659.4	6.3269	2.2475	10.460	3.8462	1.3663	1.8886
40	0.5	f_1	242.06	61.728	62281.1	0.3930	0.1235	0.0357	0.2389	0.0751	0.0057
		f_2	33.531	22.450	1611.1	0.3930	0.1235	0.0357	0.2389	0.0751	0.0057
		f_3	3.9996	5.4911	44.182	0.3944	0.1235	0.0361	0.2397	0.0751	0.0057
	2	f_1	306.80	265.23	161974	6.2894	1.9767	9.1449	3.8234	1.2017	1.4737
		f_2	94.878	126.66	24287.5	6.2894	1.9767	9.1449	3.8235	1.2016	1.4737
		f_3	63.515	87.793	11242.0	6.2907	1.9767	9.1509	3.8242	1.2017	1.4735

$S_i^2 = 0$. For independent normal errors, since $(r_i - 1)S_i^2/\sigma^2$ is chisquared with $r_i - 1$ degrees of freedom,

$$\sum_{i=1}^{n} \frac{(r_i - 1)S_i^2}{\sigma^2} \sim \chi^2_\kappa, \tag{3.2}$$

where $\kappa = \sum_{i=1}^n (r_i - 1)$. We take

$$\hat{\sigma}_1^2 = \frac{1}{\kappa}\sum_{i=1}^{n}(r_i - 1)S_i^2. \tag{3.3}$$

As only information within design points is being used, refer to $\hat{\sigma}_1^2$ as the within-design-points estimator.

For the between-design-point variation, $\bar{\varepsilon}_i = \sum_{j=1}^{r_i} \varepsilon_{ij}/r_i$. Then, $E(\bar{\varepsilon}_i) = 0$ and $\mathrm{var}(\bar{\varepsilon}_i) = \sigma^2/r_i$. Using the average information, the regression model (3.1) reduces to

$$\bar{Y}_i = f(x_i) + \bar{\varepsilon}_i, \tag{3.4}$$

a heteroscedastic nonparametric regression model with uncorrelated errors $\mathrm{var}(\bar{\varepsilon}_i) = \sigma^2/r_i$. Thus, by Section 2.1 we can also estimate σ^2 by

$$\hat{\sigma}_2^2 = \frac{n}{(n-m)R}\sum_{k=1}^{n-m}\left(\sum_{j=0}^{m} d_j' \bar{Y}_{j+k}\right)^2, \tag{3.5}$$

where $R = \sum_{i=1}^n 1/r_i$, $d'_j = d_j/\sqrt{\nu_j}$ and $\nu_j = (n/[(n-m)R]) \sum_{k=1}^{n-m} 1/r_{j+k}$. The sequence $\{d'_j\}$ satisfies $\sum_{j=0}^m d'_j = 0$, $\sum_{j=0}^m \nu_j {d'_j}^2 = 1$. We refer to $\hat\sigma_2^2$ as the between-design-points estimator as it does not take the advantage of the repeated measurements.

As $\hat\sigma_1^2$ and $\hat\sigma_2^2$ are suboptimal, we propose the estimator

$$\hat\sigma_3^2(\alpha) = (1-\alpha)\hat\sigma_1^2 + \alpha\hat\sigma_2^2, \tag{3.6}$$

where $0 \le \alpha \le 1$ is the tuning parameter. As $\bar{Y}_i$ and S_i^2 are independent of each other, $\hat\sigma_1^2$ and $\hat\sigma_2^2$ are independent, and

$$\text{MSE}(\sigma_3^2(\alpha)) = (1-\alpha)^2\text{MSE}(\hat\sigma_1^2) + \alpha^2\text{MSE}(\hat\sigma_2^2). \tag{3.7}$$

The optimal tuning parameter α is then chosen to minimize $\text{MSE}(\sigma_3^2(\alpha))$,

$$\alpha_{\text{opt}} = \frac{\text{MSE}(\hat\sigma_1^2)}{\text{MSE}(\hat\sigma_1^2) + \text{MSE}(\hat\sigma_2^2)}. \tag{3.8}$$

From Theorem 1 and the fact that $\text{MSE}(\hat\sigma_1^2) = 2\sigma^4/\kappa$, we estimate α by

$$\hat\alpha_{\text{opt}} = \left[1 + \frac{\kappa n^2}{R^2(n-m)^2}\left\{\frac{n^2}{R^2}\sum_{i=1}^n \frac{1}{r_i^2}\Big(\sum_{j=0}^m {d'_j}^2\Big)^2 + 2\sum_{k=1}^m h_k \Big(\sum_{j=0}^{m-k} d'_j d'_{j+k}\Big)^2\right\}\right]^{-1}, \tag{3.9}$$

where $\nu_j = n/\{(n-m)R\} \sum_{k=1}^{n-m} 1/r_{j+k}$, $h_k = n^2/R^2 \sum_{i=1}^{n-k} 1/(r_i r_{i+k})$, $R = \sum_{i=1}^n 1/r_i$ and $\kappa = \sum_{i=1}^n (r_i - 1)$. In the special case when r_i are all the same, the estimated optimal tuning parameter can be simplified as $\hat\alpha_{\text{opt}} = \{1 + (2m+1)\kappa/(2mn)\}^{-1}$. With the estimated tuning parameter, we estimate σ^2 by $\hat\sigma_3^2(\hat\alpha_{\text{opt}})$ and refer to it as the combined estimator.

3.1.1. Simulation study

We conducted a simulation study to evaluate the finite sample performance of $\hat\sigma_1^2$, $\hat\sigma_2^2$, and $\hat\sigma_3^2(\hat\alpha_{\text{opt}})$. The design points were $x_i = i/n$ and the ε_{ij} were i.i.d. $N(0,1)$. We took $f(x) = \sin(2\pi x)$. The measurement numbers of all samples were set to $1 + b_i$, where the b_i were generated from the binomial $(3, 0.4)$. We took $n = 25, 50$, and 200. In our simulations, the difference sequence $\{d_j\}$ was the optimal sequence. The order m of optimal sequence was $m = 1, 2, 3$ and 4. For comparison, we also considered a residual-based estimator of the form $\hat\sigma^2 = (1/N)\sum_{i=1}^n \sum_{j=1}^{r_i} (Y_{ij} - \hat f(x_i))^2$, where $N = \sum_{i=1}^n r_i$. We used cubic spline smoothing to estimate f, with smoothing parameter selected by generalized cross validation, and refer to the variance as $\hat\sigma^2_{\text{spline}}$. In total, we had 12 combinations of simulation settings.

Table 5. Mean squared errors (MSE) of $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\sigma}_3^2(\hat{\alpha}_{\text{opt}})$, and $\hat{\sigma}^2_{\text{spline}}$ under various settings.

n	m	MSE($\hat{\sigma}_1^2$)	MSE($\hat{\sigma}_2^2$)	MSE($\hat{\sigma}_3^2(\hat{\alpha}_{\text{opt}})$)	MSE($\hat{\sigma}^2_{\text{spline}}$)
25	1	0.0757	0.1286	0.0561	0.0682
	2	0.0757	0.1177	0.0555	0.0682
	3	0.0757	0.1087	0.0557	0.0682
	4	0.0757	0.1456	0.0571	0.0682
50	1	0.0357	0.0661	0.0277	0.0287
	2	0.0357	0.0563	0.0269	0.0287
	3	0.0357	0.0553	0.0268	0.0287
	4	0.0357	0.0585	0.0269	0.0287
200	1	0.0080	0.0169	0.0064	0.0050
	2	0.0080	0.0148	0.0062	0.0050
	3	0.0080	0.0143	0.0061	0.0050
	4	0.0080	0.0142	0.0061	0.0050

For each setting, we generated observations and computed the estimators $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\sigma}_3^2(\hat{\alpha}_{\text{opt}})$ and $\hat{\sigma}^2_{\text{spline}}$. 1,000 times, and report the mean squared errors in Table 5. Simulation results indicate that the proposed estimator $\hat{\sigma}_3^2(\hat{\alpha}_{\text{opt}})$ has a smaller MSE than the other estimators in all settings when the sample size is small; for large sample sizes, the performance of $\hat{\sigma}_3^2(\hat{\alpha}_{\text{opt}})$ is still comparable with that of the residual-based estimator.

3.1.2. Data examples

The first data set was from Interactive Data Analysis (McNeil (1977)) and named “cars”, it can be downloaded in the R package “datasets”. There are 50 observation measurements on the speed of a car in mph and the distance taken to stop, in ft. We took *distance* as the response variable. From the scatter plot and the fitted smoothing spline curve (the left figure in Figure 1), we observe a nonlinear relationship with heteroscedastic errors. The estimated σ^2 values with $m = 2$ are $\hat{\sigma}_1^2 = 221.8963$, $\hat{\sigma}_2^2 = 256.3992$, $\hat{\sigma}_3^2(\hat{\alpha}_{\text{opt}}) = 224.3759$, and $\hat{\sigma}^2_{\text{spline}} = 219.0512$, respectively. Here $\hat{\sigma}_3^2(\hat{\alpha}_{\text{opt}})$ and $\hat{\sigma}^2_{\text{spline}}$ are close, suggesting that the proposed estimator performs as well as the residual-based estimator.

The second data set consisted of measurements of the fetal mandible. It is named “Mandible” and can be downloaded in the R package “lmtest” (Chitty, Campbell, and Altman (1993), Royston and Altman (1994)). There are 167 observations on gestational age in weeks and mandible length in mm. The scatter plot in Figure 1 shows the smoothing spline fitted curve with evident heteroscedastic errors. With $m = 2$ the estimated σ^2 values are $\hat{\sigma}_1^2 = 4.0164$, $\hat{\sigma}_2^2 = 8.7518$, $\hat{\sigma}_3^2(\hat{\alpha}_{\text{opt}}) = 5.4903$ and $\hat{\sigma}^2_{\text{spline}} = 5.3030$. Again, we observe that $\hat{\sigma}_3^2(\hat{\alpha}_{\text{opt}})$ and $\hat{\sigma}^2_{\text{spline}}$ perform very similarly.

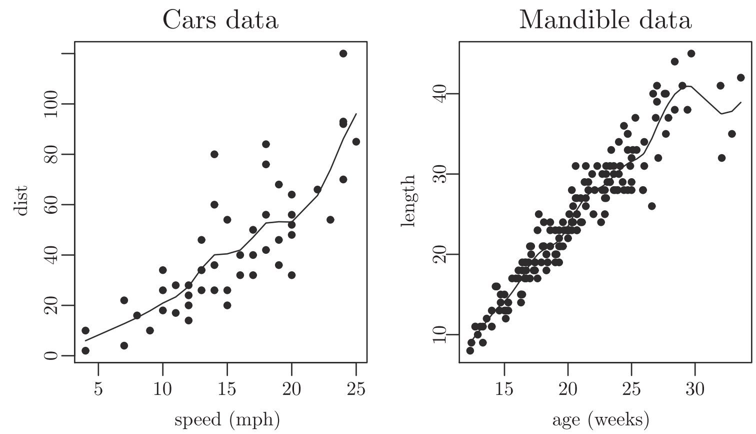

Figure 1. The Cars data and Mandible data together with the fitted curves by smoothing spline.

3.2. Semiparametric partially linear model

Consider the semiparametric partially linear model

$$Y_i = X_i\beta + f(z_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{3.10}$$

where the X_i are known p-dimensional vectors with $p < n$, β is an unknown p-dimensional parameter vector, f is an unknown smooth function, and the ε_i are i.i.d. random errors with mean zero and variance σ^2. To estimate σ^2, Wang, Brown, and Cai (2011) proposed a difference-based method. They applied Hall, Kay, and Titterington (1990)'s optimal sequence to remove the nonparametric part, and then used a linear regression to estimate β. They used the residuals to estimate σ^2.

Our proposed method can be applied when $\Sigma \neq I$, or when $\Sigma = I$ but there are repeated measurements at certain design points (X_i, z_i). We discuss another setting in which our method can be applied.

Rewrite (3.10) in matrix form as

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{f} + \boldsymbol{\varepsilon}, \tag{3.11}$$

where $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^T$, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$, $\boldsymbol{f} = (f(z_1), \ldots, f(z_n))^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$. With $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 I$, we can employ the difference sequence $\{d_0, d_1, \ldots, d_m\}$ with constraint (1.2) to remove the nonparametric part by letting

$$\bar{D} = \begin{pmatrix} d_0 \ d_1 \ d_2 \ \ldots d_m \ 0 \ \ldots\ldots \ 0 \\ 0 \ d_0 \ d_1 \ d_2 \ \ldots d_m \ 0 \ \ldots \ 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ 0 \ \ldots\ldots d_0 \ d_1 \ d_2 \ \ldots d_m \ 0 \\ 0 \ \ 0 \ \ldots\ldots d_0 \ d_1 \ d_2 \ \ldots d_m \end{pmatrix}.$$

Then $\bar{D}\boldsymbol{Y} = \bar{D}\boldsymbol{X}\beta + \bar{D}\boldsymbol{f} + \bar{D}\boldsymbol{\varepsilon} \approx \bar{D}\boldsymbol{X}\beta + \bar{D}\boldsymbol{\varepsilon}$. This leads to

$$\tilde{\boldsymbol{Y}} \approx \tilde{\boldsymbol{X}}\beta + \tilde{\varepsilon}, \tag{3.12}$$

where $\tilde{\boldsymbol{Y}} = \bar{D}\boldsymbol{Y}$, $\tilde{\boldsymbol{X}} = \bar{D}\boldsymbol{X}$, and $\tilde{\varepsilon} = \bar{D}\boldsymbol{\varepsilon}$. Accordingly, the the least squares estimator of β is $\hat{\beta} = (\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{Y}}$. Substituting $\hat{\beta}$ into (3.11), we have

$$\check{\boldsymbol{Y}} = \boldsymbol{f} + \boldsymbol{e}, \tag{3.13}$$

where $\check{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\beta}$. It is easy to verify that $E(\boldsymbol{e}) = 0$ and $\mathrm{var}(\boldsymbol{e}) = \sigma^2\Sigma$, where $\Sigma = \Big(I - \boldsymbol{X}(\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^T\bar{D}\Big)\Big(I - \boldsymbol{X}(\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^T\bar{D}\Big)^T$. Here Σ is totally specified since the X_i are known, and $\bar{D}$ can be derived too. Hence, (3.13) is a heteroscedastic nonparametric regression with a known covariance matrix, and suitable for our proposed method.

3.2.1. Simulation study

We conducted a simulation study to evaluate the performance of our proposed estimator in partially linear models, and compared it with the residual-based estimator in Wang, Brown, and Cai (2011), denoted by $\hat{\sigma}^2_{\mathrm{W}}$. For the linear component, we took $\beta = (2, 2, 4)^T$ and generated X_i from $N_3((1, 2, 4)^T, I_3)$, where I_3 is a 3×3 identity matrix. As the nonparametric component, we took $f(z) = 5\sin(\omega\pi z)$ with $\omega = 1$, 2, and 4. The design points were $z_i = i/n$ for $i = 1, \ldots, n$. The random errors ε_i were independent $N(0, \sigma^2)$ with $\sigma = 0.5$ and 2. We considered the difference order $m = 2$. For $\hat{\sigma}^2_{\mathrm{W}}$, the sequence is $d = \{\sqrt{m/(m+1)}, -1/\sqrt{m(m+1)}, -1/\sqrt{m(m+1)}\}$, recommended in Wang, Brown, and Cai (2011). For $\tilde{\sigma}^2_{\mathrm{new}}$, the sequence was obtained by the Lagrange multiplier method. For each setting with $n = 50$, 200, and 500, we repeated the simulation 500 times and here report the means, standard errors, and MSEs of $\hat{\sigma}^2_{\mathrm{W}}$ and $\tilde{\sigma}^2_{\mathrm{new}}$ in Table 6. Our proposed estimator performs as well as the residual-based estimator $\hat{\sigma}^2_{\mathrm{W}}$.

4. Discussion

We have considered the estimation of the residual variance in nonparametric regression when the random errors are heteroscedastic but known. Our proposed estimators are asymptotically unbiased and also have a smaller MSE than existing estimators. We applied the proposed methods to nonparametric regression models with repeated measurements and to semiparametric partially linear models. Data examples and simulation studies suggest that the proposed methods work well in a wide range of problems. Of course, we hope to clarify that when the

Table 6. Means, standard errors (SE) and mean squared errors (MSE) of $\hat{\sigma}^2_{\mathrm{W}}$ and $\tilde{\sigma}^2_{\mathrm{new}}$ under various settings.

			$\hat{\sigma}^2_{\mathrm{W}}$			$\tilde{\sigma}^2_{\mathrm{new}}$		
n	σ	f	Mean	SE	MSE	Mean	SE	MSE
50	0.5	f_1	0.3213	0.0613	0.0088	0.3198	0.0615	0.0086
		f_2	0.5459	0.0663	0.0919	0.5264	0.0659	0.0807
		f_3	1.4215	0.0833	1.3795	1.3332	0.0804	1.1798
	2	f_1	4.0083	0.9675	0.9344	4.0743	0.9731	0.9506
		f_2	4.2339	0.9748	1.0032	4.2817	0.9797	1.0373
		f_3	5.1062	0.9765	2.1755	5.0854	0.9789	2.1345
200	0.5	f_1	0.2530	0.0279	0.0007	0.2539	0.0280	0.0007
		f_2	0.2668	0.0280	0.0010	0.2669	0.0280	0.0010
		f_3	0.3226	0.0280	0.0060	0.3189	0.0281	0.0055
	2	f_1	3.9782	0.4478	0.2006	3.9981	0.4482	0.2005
		f_2	3.9919	0.4478	0.2002	4.0111	0.4484	0.2008
		f_3	4.0479	0.4476	0.2022	4.0636	0.4485	0.2048
500	0.5	f_1	0.2502	0.0184	0.0003	0.2504	0.0183	0.0003
		f_2	0.2525	0.0184	0.0003	0.2523	0.0183	0.0003
		f_3	0.2614	0.0184	0.0004	0.2597	0.0183	0.0004
	2	f_1	3.9933	0.2957	0.0873	3.9987	0.2933	0.0858
		f_2	3.9955	0.2957	0.0872	4.0006	0.2933	0.0858
		f_3	4.0044	0.2956	0.0872	4.0080	0.2932	0.0859

prior information on Σ is limited, the proposed method may not be practically useful. Further research is needed in this direction.

5. Proofs

Proof of Theorem 1. By (2.2), we have

$$E(\hat{\sigma}^2_{\mathrm{new}}) = \sigma^2 + \frac{1}{n-m}\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m} d'_j f_{j+k}\Big)^2.$$

By Cauchy inequality and the fact that $\min\{w_0, \ldots, w_m\} \geq c_0 > 0$, we have

$$\begin{aligned}\Big|\sum_{j=0}^{m} d'_j f_{j+k}\Big| &= \Big|\sum_{j=0}^{m} \frac{d_j}{\sqrt{w_j}} f_{j+k}\Big| = O\Big(\frac{1}{n}\Big|\sum_{j=0}^{m}(j+1)\frac{d_j}{\sqrt{w_j}}\Big|\Big)\\ &= O\Big(\frac{1}{n}\Big\{\sum_{j=0}^{m}\frac{(j+1)^2}{w_j}\sum_{j=0}^{m} d_j^2\Big\}^{1/2}\Big) = O\Big(\frac{m^{3/2}}{n}\Big).\end{aligned}$$

This leads to

$$\frac{1}{n-m}\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m} d'_j f_{j+k}\Big)^2 = O\Big(\frac{m^3}{n^2}\Big).$$

Therefore, $E(\hat\sigma^2_{\rm new}) = \sigma^2 + O(m^3/n^2)$.

For the variance of $\hat\sigma^2_{\rm new}$, we have

$$\begin{aligned}
\mathrm{var}(\hat\sigma^2_{\rm new}) &= \frac{1}{(n-m)^2}\mathrm{var}(\boldsymbol{Y}^T A\boldsymbol{Y}) \\
&= \frac{1}{(n-m)^2}\mathrm{var}(\boldsymbol{f}^T A\boldsymbol{f} + 2\boldsymbol{f}^T A\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T A\boldsymbol{\varepsilon}) \\
&= \frac{1}{(n-m)^2}\left\{4\mathrm{var}(\boldsymbol{f}^T A\boldsymbol{\varepsilon}) + 4\mathrm{Cov}(\boldsymbol{f}^T A\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}^T A\boldsymbol{\varepsilon}) + \mathrm{var}(\boldsymbol{\varepsilon}^T A\boldsymbol{\varepsilon})\right\}. \quad (5.1)
\end{aligned}$$

We calculate these terms in order. By Lemma A.1 in Yatchew (2000),

$$\begin{aligned}
4\mathrm{var}(\boldsymbol{f}^T A\boldsymbol{\varepsilon}) &= 4\boldsymbol{f}^T A\Sigma A^T\boldsymbol{f}\sigma^2 = 4(D\boldsymbol{f})^T D\Sigma D^T(D\boldsymbol{f})\sigma^2 \\
&= 4\sum_{i=1}^{n-m}\sum_{j=1}^{n-m} r_{ij}\Big(\sum_{l=0}^{m} d'_l f_{l+i}\Big)\Big(\sum_{l=0}^{m} d'_l f_{l+j}\Big)\sigma^2 \\
&= O(m^3), \quad (5.2) \\
4\mathrm{Cov}(\boldsymbol{f}^T A\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}^T A\boldsymbol{\varepsilon}) &= 4E\{(D\boldsymbol{f})^T D\boldsymbol{\varepsilon}(D\boldsymbol{\varepsilon})^T D\boldsymbol{\varepsilon}\} \\
&= 4\sum_{k=1}^{n-m}\sum_{v=1}^{n-m}\sum_{j=0}^{m} d'_j f_{j+v} E\Big\{\Big(\sum_{j=0}^{m} d'_j\varepsilon_{j+v}\Big)\Big(\sum_{j=0}^{m} d'_j\varepsilon_{j+k}\Big)^2\Big\} \\
&= O(m^{7/2}). \quad (5.3) \\
\mathrm{var}(\boldsymbol{\varepsilon}^T A\boldsymbol{\varepsilon}) &= \mathrm{tr}(\eta\odot A\odot A - 3\sigma^4\Sigma^2\odot A\odot A) + 2\sigma^4\mathrm{tr}(\Sigma A\Sigma A),
\end{aligned}$$

where $R = D\Sigma D^T = (r_{ij})_{(n-m)\times(n-m)}$ and $\eta = \mathrm{diag}(E(\varepsilon_1^4), E(\varepsilon_2^4), \ldots, E(\varepsilon_n^4))$. It is easy to verify that

$$\begin{aligned}
&\mathrm{tr}(\eta\odot A\odot A - 3\sigma^4\Sigma^2\odot A\odot A) \\
&= \Big\{\sum_{i=0}^{m-1}\Big(\sum_{j=0}^{i}{d'_j}^2\Big)^2 E(\varepsilon_{i+1}^4) + \sum_{i=m+1}^{n-m}\Big(\sum_{j=0}^{m}{d'_j}^2\Big)^2 E(\varepsilon_i^4) + \sum_{i=1}^{m}\Big(\sum_{j=i}^{m}{d'_j}^2\Big)^2 E(\varepsilon_{n-m+i}^4)\Big\} \\
&\quad -3\sigma^4\Big\{\sum_{i=0}^{m-1}\Big(\sum_{j=0}^{i}{d'_j}^2\Big)^2 c_{i+1}^2 + \sum_{i=m+1}^{n-m}\Big(\sum_{j=0}^{m}{d'_j}^2\Big)^2 c_i^2 + \sum_{i=1}^{m}\Big(\sum_{j=i}^{m}{d'_j}^2\Big)^2 c_{n-m+i}^2\Big\},
\end{aligned}$$

$$\begin{aligned}
\mathrm{tr}(\Sigma A\Sigma A) =& \sum_{i=0}^{m-1} c_{i+1}^2\Big(\sum_{j=0}^{i}{d'_j}^2\Big)^2 + \sum_{i=m+1}^{n-m} c_i^2\Big(\sum_{j=0}^{m}{d'_j}^2\Big)^2 + \sum_{i=1}^{m} c_{n-m+i}^2\Big(\sum_{j=i}^{m}{d'_j}^2\Big)^2 \\
&+2\sum_{k=1}^{m}\Big\{\sum_{i=0}^{m-1} c_{i+1}c_{i+1+k}\Big(\sum_{j=0}^{i} d'_j d'_{j+k}\Big)^2 + \sum_{i=m+1}^{n-m-k} c_i c_{i+k}\Big(\sum_{j=0}^{m-k} d'_j d'_{j+k}\Big)^2 \\
&+\sum_{i=1}^{m-k} c_{n-m+i}c_{n-m+i+k}\Big(\sum_{j=i}^{m-k} d'_j d'_{j+k}\Big)^2\Big\}.
\end{aligned}$$

This leads to

$$\begin{aligned}
&\text{var}(\boldsymbol{\varepsilon}^T A\boldsymbol{\varepsilon})\\
&= \sum_{i=0}^{m-1}\Big(\sum_{j=0}^{i} {d'_j}^2\Big)^2 \text{var}(\varepsilon_{i+1}^2) + \sum_{i=m+1}^{n-m}\Big(\sum_{j=0}^{m} {d'_j}^2\Big)^2 \text{var}(\varepsilon_i^2)\\
&\quad + \sum_{i=1}^{m}\Big(\sum_{j=i}^{m} {d'_j}^2\Big)^2 \text{var}(\varepsilon_{n-m+i}^2) + 4\sigma^4 \sum_{k=1}^{m}\bigg\{\sum_{i=0}^{m-1} c_{i+1}c_{i+1+k}\Big(\sum_{j=0}^{i} d'_j d'_{j+k}\Big)^2\\
&\quad + \sum_{i=m+1}^{n-m-k} c_i c_{i+k}\Big(\sum_{j=0}^{m-k} d'_j d'_{j+k}\Big)^2 + \sum_{i=1}^{m-k} c_{n-m+i}c_{n-m+i+k}\Big(\sum_{j=i}^{m-k} d'_j d'_{j+k}\Big)^2\bigg\}.
\end{aligned} \tag{5.4}$$

Substituting (5.2)−(5.4) into (5.1), we can verify that

$$\begin{aligned}
\text{var}(\hat{\sigma}^2_{\text{new}}) &= \frac{1}{(n-m)^2}\text{var}(\boldsymbol{\varepsilon}^T A\boldsymbol{\varepsilon}) + O\Big(\frac{m^3}{n^2}\Big) + O\Big(\frac{m^{5/2}}{n^2}\Big)\\
&= \frac{1}{(n-m)^2}\bigg\{\sum_{i=1}^{n}\Big(\sum_{j=0}^{m} {d'_j}^2\Big)^2 \text{var}(\varepsilon_i^2) + 4\sigma^4\sum_{k=1}^{m} b_k\Big(\sum_{j=0}^{m-k} d'_j d'_{j+k}\Big)^2\bigg\}\\
&\quad + O\Big(\frac{m^{7/2}}{n^2}\Big).
\end{aligned}$$

Noting that $\text{Bias}(\hat{\sigma}^2_{\text{new}}) = O(m^3/n^2)$, for any $m = n^r$ with $0 < r < 2/7$ we have

$$\begin{aligned}
\text{MSE}(\hat{\sigma}^2_{\text{new}}) &= \text{var}(\hat{\sigma}^2_{\text{new}}) + \big\{\text{Bias}(\hat{\sigma}^2_{\text{new}})\big\}^2\\
&= \frac{1}{(n-m)^2}\bigg\{\sum_{i=1}^{n}\Big(\sum_{j=0}^{m} {d'_j}^2\Big)^2 \text{var}(\varepsilon_i^2) + 4\sigma^4\sum_{k=1}^{m} b_k\Big(\sum_{j=0}^{m-k} d'_j d'_{j+k}\Big)^2\bigg\}\\
&\quad + o\Big(\frac{1}{n}\Big).
\end{aligned}$$

The derivation of $\text{MSE}(\hat{\sigma}^2_{\text{H}})$ is similar and so is omitted here.

Proof of Theorem 2. By (2.8), we have

$$E(\tilde{\sigma}^2_{\text{new}}) = \sigma^2 + \frac{1}{n-m}\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m} \tilde{d}_j f_{j+k}\Big)^2.$$

For any $m = O(1)$, under Assumptions 1 and 2 we have $|\sum_{j=0}^{m} \tilde{d}_j f_{j+k}| = O(1/n)$. This leads to

$$\text{Bias}(\tilde{\sigma}^2_{\text{new}}) = \frac{1}{n-m}\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m} \tilde{d}_j f_{j+k}\Big)^2 = O\Big(\frac{1}{n^2}\Big).$$

For the variance of $\tilde{\sigma}^2_{\rm new}$, we write

$$\mathrm{var}(\tilde{\sigma}^2_{\rm new}) = \frac{1}{(n-m)^2}\Big\{4\mathrm{var}(\boldsymbol{f}^T\tilde{A}\boldsymbol{\varepsilon}) + 4\mathrm{Cov}(\boldsymbol{f}^T\tilde{A}\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}^T\tilde{A}\boldsymbol{\varepsilon}) + \mathrm{var}(\boldsymbol{\varepsilon}^T\tilde{A}\boldsymbol{\varepsilon})\Big\},$$

where $\Sigma = (c_{i,j})$ and $\tilde{A} = \tilde{D}^T\tilde{D}$. Let $\tilde{D}\Sigma\tilde{D}^T = (\gamma_{ij})_{(n-m)\times(n-m)}$. For any $m = O(1)$, it is easy to verify that

$$\begin{aligned}4\mathrm{var}(\boldsymbol{f}^T\tilde{A}\boldsymbol{\varepsilon}) &= 4\sigma^2\boldsymbol{f}^T\tilde{A}\Sigma\tilde{A}^T\boldsymbol{f}\\ &= 4\sigma^2\sum_{i=1}^{n-m}\sum_{j=1}^{n-m}\gamma_{ij}\Big(\sum_{l=0}^{m}\tilde{d}_l f_{l+i}\Big)\Big(\sum_{l=0}^{m}\tilde{d}_l f_{l+j}\Big)\\ &= O(1),\end{aligned}$$

$$\begin{aligned}4\mathrm{Cov}(\boldsymbol{f}^T\tilde{A}\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}^T\tilde{A}\boldsymbol{\varepsilon}) &= 4E\{(\tilde{D}\boldsymbol{f})^T\tilde{D}\boldsymbol{\varepsilon}(\tilde{D}\boldsymbol{\varepsilon})^T\tilde{D}\boldsymbol{\varepsilon}\}\\ &= 4\sum_{k=1}^{n-m}\sum_{v=1}^{n-m}\sum_{j=0}^{m}\tilde{d}_j f_{j+v}E\Big\{\Big(\sum_{j=0}^{m}\tilde{d}_j\varepsilon_{j+v}\Big)\Big(\sum_{j=0}^{m}\tilde{d}_j\varepsilon_{j+k}\Big)^2\Big\}\\ &= O(1).\end{aligned}$$

For the third term, see Schott (1997),

$$\begin{aligned}\mathrm{var}(\boldsymbol{\varepsilon}^T\tilde{A}\boldsymbol{\varepsilon}) &= \mathrm{tr}\{(\tilde{A}\otimes\tilde{A})E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\otimes\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)\} - \{\mathrm{tr}(\tilde{A}\Sigma)\}^2\sigma^4\\ &= E\{\mathrm{tr}(\tilde{A}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\otimes\tilde{A}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)\} - \{\mathrm{tr}(\tilde{A}\Sigma)\}^2\sigma^4\\ &= E\{\mathrm{tr}(\tilde{A}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)\}^2 - \{\mathrm{tr}(\tilde{A}\Sigma)\}^2\sigma^4\\ &= E\Big\{\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m}\tilde{d}_j\varepsilon_{j+k}\Big)^2\Big\}^2 - \Big(\sum_{v=1}^{n-m}\sum_{k=0}^{m}\sum_{j=0}^{m}\tilde{d}_k\tilde{d}_j c_{j+v,k+v}\Big)^2\sigma^4.\end{aligned}$$

According, we have

$$\begin{aligned}\mathrm{MSE}(\tilde{\sigma}^2_{\rm new}) &= \mathrm{var}(\tilde{\sigma}^2_{\rm new}) + \mathrm{Bias}^2(\tilde{\sigma}^2_{\rm new})\\ &= \frac{1}{(n-m)^2}\Big[E\Big\{\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m}\tilde{d}_j\varepsilon_{j+k}\Big)^2\Big\}^2\\ &\quad -\Big(\sum_{v=1}^{n-m}\sum_{k=0}^{m}\sum_{j=0}^{m}\tilde{d}_k\tilde{d}_j c_{j+v,k+v}\Big)^2\sigma^4\Big] + O\Big(\frac{1}{n^2}\Big).\end{aligned}$$

Similarly, for Hall's estimator we have

$$\mathrm{MSE}(\hat{\sigma}^2_{\rm H}) = \frac{1}{(n-m)^2}\Big[E\Big\{\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m}d_j\varepsilon_{j+k}\Big)^2\Big\}^2 - \Big(\sum_{v=1}^{n-m}\sum_{k=0}^{m}\sum_{j=0}^{m}d_k d_j c_{j+v,k+v}\Big)^2\sigma^4\Big]$$

$$
\begin{aligned}
&+\frac{1}{(n-m)^2}\Big\{\sum_{v=1}^{n-m}\Big(\sum_{k=0}^{m}\sum_{j=0}^{m}d_k d_j c_{j+v,k+v}-1\Big)\Big\}^2\sigma^4+O\Big(\frac{1}{n^2}\Big)\\
=&\frac{1}{(n-m)^2}E\Big\{\sum_{k=1}^{n-m}\Big(\sum_{j=0}^{m}d_j\varepsilon_{j+k}\Big)^2\Big\}^2\\
&-\frac{1}{(n-m)}\Big\{2\sum_{v=1}^{n-m}\sum_{k=0}^{m}\sum_{j=0}^{m}d_k d_j c_{j+v,k+v}-(n-m)\Big\}\sigma^4+O\Big(\frac{1}{n^2}\Big).
\end{aligned}
$$

This completes the proof of Theorem 2.

Acknowledgements

Yebin Cheng's research was supported in part by NSFC Grant 11271241. Lie Wang's research was supported in part by NSF Grant DMS-1005539. Tiejun Tong's research was supported in part by Hong Kong RGC Grant 202711 and HKBU Grants FRG2/11-12/110, FRG1/13-14/018, and FRG2/13-14/062. The authors thank the editor, an associate editor, and two reviewers for their helpful comments and suggestions that have substantially improved the paper.

References

Brown, L. and Levine, M. (2007). Variance estimation in nonparametric regression via the difference sequence method. *Ann. Statist.* **35**, 2219-2232.

Buckley, M. J., Eagleson, G. K. and Silverman, B. W. (1988). The estimation of residual variance in non-parametric regression. *Biometrika* **75**, 189-199.

Carroll, R. J. (1987). The effects of variance function estimation on prediction and calibration: an example. *Statist. Decision Theory and Related Topics* **2**, 273-280.

Carroll, R. J. and Ruppert, D. (1988). *Transforming and Weighting in Regression.* Chapman and Hall, London.

Chitty, L. S., Campbell, S. and Altman, D. G. (1993). Measurement of the fetal mandible - feasibility and construction of a centile chart. *Prenatal Diagnosis* **13**, 749-756.

Dette, H., Munk, A. and Wagner, T. (1998). Estimating the variance in nonparametric regression - what is a reasonable choice. *J. Roy. Statist. Soc. Ser. B* **60**, 751-764.

Eubank, R. L. and Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85**, 387-392.

Gasser, T., Kneip, A. and Kohler, W. (1991). A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.* **86**, 643-652.

Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika 73*, 625-633.

Hall, P., Kay, J. W. and Titterington, D. M. (1990). Asymptotically optimal difference- based estimation of variance in nonparametric regression. *Biometrika* **77**, 521-528.

Härdle, W. and Tsybakov, A. (1997). Local polynomial estimators of the valatility func- tion in nonparametric autoregression. *J. Econom.* **81**, 223-242.

Kay, J. (1988). On the choice of the regularisation parameter in image restoration. *Lecture Notes in Computer Sci.* **301**, 587-596.

Kulasekera, K. B. and Gallagher, C. (2002). Variance estimation in nonparametric multiple regression. *Comm. Statist. Theory Methods* **31**, 1373-1383.

Levine, M. (2006). Bandwidth selection for a class of difference-based variance estimators in the nonparametric regression: a possible approach. *Comput. Statist. Data Anal.* **50**, 3405-3431.

McNeil, D. (1977). *Interactive Data Analysis*. Wiley, New York.

Müller, U., Schick, A. and Wefelmeyer, W. (2003). Estimating the error variance in nonparametric regression by a covariate-matched U-statistic. *Statistics* **37**, 179-188.

Rice, J. A. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12**, 1215-1230.

Royston, P. and Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Appl. Statist.* **43**, 429-453.

Schott, J. R. (1997). *Matrix Analysis for Statistics*. John Wiley and Sons, New York.

Tong, T., Ma, Y. and Wang, Y. (2013). Optimal variance estimation without estimating the mean function. *Bernoulli* **19**, 1839-1854.

Tong, T. and Wang, Y. (2005). Estimating residual variance in nonparametric regression using least squares. *Biometrika* **92**, 821-830.

Wang, L., Brown, L. and Cai, T. (2011). A difference based approach to the semiparametric partial linear model. *Electron. J. Statist.* **5**, 619-641.

Yatchew, A. (2000). Differencing methods in nonparametric regression: simple techniques for the applied econometrician. `http://www.economics.utoronto.ca/yatchew`.

Zhao, H. and You, J. (2011). Difference based estimation for partially linear regression models with measurement errors. *J. Multivariate Anal.* **102**, 1321-1338.

School of Science, Anhui University of Science and Technology, Huainan, Anhui, 232001, China.

E-mail: yjchow@163.com

Glorious Sun School of Business and Management, Donghua University, Shanghai, 200051, China.

E-mail: Y.Cheng@dhu.edu.cn

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

E-mail: liewang@math.mit.edu

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

E-mail: tongt@hkbu.edu.hk

(Received January 2013; accepted August 2014)