Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 585146, 14 pages
http://dx.doi.org/10.1155/2014/585146

*Research Article*

# A Least Squares Method for Variance Estimation in Heteroscedastic Nonparametric Regression

**Yuejin Zhou,[1,2] Yebin Cheng,[2] and Tiejun Tong[3]**

[1] *School of Science, Anhui University of Science and Technology, Huainan 232001, China*
[2] *School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 20043, China*
[3] *Department of Mathematics, Hong Kong Baptist University, Hong Kong*

Correspondence should be addressed to Tiejun Tong; tongt@hkbu.edu.hk

Received 28 March 2014; Revised 7 June 2014; Accepted 9 June 2014; Published 3 July 2014

Academic Editor: Li Ma



Interest in variance estimation in nonparametric regression has grown greatly in the past several decades. Among the existing methods, the least squares estimator in Tong and Wang (2005) is shown to have nice statistical properties and is also easy to implement. Nevertheless, their method only applies to regression models with homoscedastic errors. In this paper, we propose two least squares estimators for the error variance in heteroscedastic nonparametric regression: the intercept estimator and the slope estimator. Both estimators are shown to be consistent and their asymptotic properties are investigated. Finally, we demonstrate through simulation studies that the proposed estimators perform better than the existing competitor in various settings.

## 1. Introduction

Consider the nonparametric regression model

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where $y_i$ are observations, $x_i$ are design points with $0 \le x_1 \le \cdots \le x_n \le 1$, $f(\cdot)$ is an unknown mean function, and $\varepsilon_i$ are independent random errors with mean zero and variance $c_i\sigma^2$, respectively. In the special case when $c_i$ are all the same, model (1) reduces to a homoscedastic nonparametric regression. In this paper, we are interested in estimating the variance $\sigma^2$ in the situation when $c_i$ are not all the same but known constants. Note that such a setting can arise in various situations. As an illustration, we consider a regression model with $r_i$ repeated observations on design points $x_i$, respectively, where the measurement errors are normal. If in practice we only report the average values on each design point, we have the new model as $\overline{y}_i = f(x_i) + \overline{\varepsilon}_i$, where $\mathrm{var}(\overline{\varepsilon}_i) = c_i\sigma^2$ with $c_i = 1/r_i$.

Needless to say, an accurate estimate of variance is important in nonparametric regression. For instance, it is required in constructing confidence bands, in choosing the amount of smoothing, in testing the goodness of fit, and in estimating the detection limits of immunoassay [1–8]. In the past several decades, researchers have proposed many methods for estimating $\sigma^2$, especially when the regression model is homoscedastic. Among the existing methods, one popular class is referred to as difference-based estimators. The first-order difference-based estimator was proposed in Rice [9],

$$\widehat{\sigma}_R^2 = \frac{1}{2(n-1)} \sum_{i=2}^{n} (y_i - y_{i-1})^2. \tag{2}$$

Assume that $f(\cdot)$ is a Lipschitz continuous function and $\max_{2\le i\le n}\{x_i - x_{i-1}\} = O(1/n)$. Note that $y_i - y_{i-1} = f(x_i) - f(x_{i-1}) + \varepsilon_i - \varepsilon_{i-1} \approx \varepsilon_i - \varepsilon_{i-1}$ as $n \to \infty$. Therefore, $\widehat{\sigma}_R^2$ is an asymptotically unbiased estimator of $\sigma^2$. Since then, many difference-based estimators have been proposed in the literature. For instance, Gasser et al. [10] proposed a second-order difference-based estimator. Hall et al. [11] proposed an $m$th-order difference-based estimator with $m \ge 2$ a finite number. Other significant works include Dette et al. [12], Müller et al. [13], Tong et al. [14], Du and Schick [15], and Wang et al. [16], among others. Furthermore, Brown and Levine [17], Wang et al. [18], and Cai and Wang [19] considered the difference-based kernel and wavelet estimators for the variance function

in nonparametric regression. Note that the difference-based estimators do not require an estimate of the mean function and so are popular in practice.

As a variation of the difference-based estimation, Tong and wang [20] proposed a least squares estimator of $\sigma^2$. Let the lag-$k$ Rice estimator be

$$\widehat{\sigma}_R^2(k) = \frac{1}{2(n-k)} \sum_{i=k+1}^{n} (y_i - y_{i-k})^2. \tag{3}$$

For the equally spaced design with $x_i = i/n$, it can be shown that $\widehat{\sigma}_R^2(k) = \sigma^2 + Jd_k + o(d_k)$ for any $k = o(n)$, where $J = \int_0^1 (f'(x))^2 dx/2$ and $d_k = k^2/n^2$. This indicates that the lag-$k$ Rice estimators are always positively biased estimators of $\sigma^2$, especially when the sample size $n$ is small. To reduce bias, Tong and Wang regressed $\widehat{\sigma}_R^2(k)$ on $d_k$ using a simple linear regression and then estimate $\sigma^2$ as the intercept. The least squares estimator achieves the asymptotically optimal rate that is usually possessed by residual-based estimators only. In addition, Tong et al. [21] established the asymptotic normality and also demonstrated the efficiency of the least squares estimator. We also note that Park et al. [22] investigated the least squares method in small sample nonparametric regression via a local quadratic approximation to determine the regressor and weights.

The aforementioned methods have significantly advanced our understanding on the difference-based estimation of the error variance. Nevertheless, most of the above methods, including the least squares method, only applied to nonparametric regression models with homoscedastic errors. In practice, it is not uncommon that the errors may have different variances. In such situations, we note that the bias term of the least squares estimator in Tong and Wang [20] will be significantly enlarged; for more details, see Sections 2 and 3. Inspired by this, we propose two adaptive least squares estimators for the residual variance in heteroscedastic nonparametric regression.

The remainder of this paper is organized as follows. In Section 2, we propose two least squares estimators for the error variance: the intercept estimator and the slope estimator. In Section 3, we investigate the asymptotic properties of the proposed estimators and present some theoretical results including the asymptotic normalities of the estimators. In Section 4, we conduct simulation studies to evaluate the proposed estimators and compare them with the existing competitor in the literature. We then conclude the paper in Section 5 with a brief discussion and provide the technical proofs in Section 6.

## 2. Methodology

For model (1), without loss of generality, we assume that $\sum_{i=1}^{n} c_i = n$. In matrix notation, the model is written as

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\varepsilon}, \tag{4}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$, $\mathbf{f} = (f(x_1), f(x_2), \ldots, f(x_n))'$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$. The covariance matrix of $\boldsymbol{\varepsilon}$ is $\sigma^2\Sigma$, where

$$\Sigma = \begin{pmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & c_n \end{pmatrix}. \tag{5}$$

When $\Sigma = I$, namely, $c_i = 1$ for all $i$, it reduced to the homoscedastic setting in Tong and Wang [20]. In this paper, we assume that the $c_i$ values are not all the same.

For this setting, one naive approach is to apply the transformation $\Sigma^{-1/2}\mathbf{y} = \Sigma^{-1/2}\mathbf{f} + \Sigma^{-1/2}\boldsymbol{\varepsilon}$. Through this transformation the errors become homogeneous. Nevertheless, meanwhile, it makes the transformed mean function $\Sigma^{-1/2}\mathbf{f}$ no longer a Lipschitz continuous function. Specifically, if $c_i \neq c_{i-1}$ and $f(x_i) \neq 0$, the difference $f(x_i)/\sqrt{c_i} - f(x_{i-1})/\sqrt{c_{i-1}}$ will not be negligible when $x_i - x_{i-1}$ tends to be zero. As a consequence, the difference-based methods fail to apply in such situations.

To advance the research project, we reconsider the lag-$k$ Rice estimator defined in Tong and Wang [20]. Suppose that $f$ has a bounded first derivative. For model (4), the expectation of the lag-$k$ Rice estimator is

$$E\left(\widehat{\sigma}_R^2(k)\right) = a_k\sigma^2 + b_kJ + O\left(\frac{k^3}{n^2(n-k)}\right) + o\left(\frac{1}{n^2}\right), \tag{6}$$

where $a_k = \sum_{i=k+1}^{n}(c_i + c_{i-k})/2(n-k)$, $b_k = k^2/n^2$, and $J = \int_0^1 (f'(x))^2 dx/2$. Note that $a_k = (n + c_{k+1} + \cdots + c_{n-k})/(n + n - 2k) \neq 1$ when $c_{k+1} + \cdots + c_{n-k} \neq n - 2k$. Therefore, for model (4) with heteroscedastic errors, it is not guaranteed that $\widehat{\sigma}_R^2(k)$ is an asymptotically unbiased estimator of $\sigma^2$.

In what follows, we develop two new estimators for $\sigma^2$: (i) the first method estimates $\sigma^2$ as the intercept and (ii) the second method estimates $\sigma^2$ as the slope. For the first method, we let $z_k = \widehat{\sigma}_R^2(k)/a_k$ and $t_k = b_k/a_k$. Then, for any $k = o(n)$, we have

$$E(z_k) = \sigma^2 + Jt_k + o(t_k). \tag{7}$$

Now treating $z_k$ as the response variable and $t_k$ as the independent variable, we fit the following simple linear regression and estimate $\sigma^2$ as the fitted intercept,

$$z_k = \alpha + \beta t_k + e_k, \quad k = 1, 2, \ldots, m, \tag{8}$$

where $e_k$ are the random errors and $m$ is the total number of pairs used for the fit. Note that $z_k$ involves $(n-k)$ pairs of difference; we assign weights $w_k = (n-k)/N$, where $N = (n-1) + \cdots + (n-m) = mn - m(m+1)/2$, to the response variable $z_k$. We then fit the linear model (8) using the weight least squares that minimizes the weighted sum of squares $\sum_{k=1}^{m} w_k(z_k - \alpha - \beta t_k)^2$. Specifically, the estimated error variance is

$$\widehat{\sigma}_1^2 = \widehat{\alpha} = \overline{z}_w - \widehat{\beta}\,\overline{t}_w, \tag{9}$$

where $\bar{z}_w = \sum_{k=1}^m w_k z_k$, $\bar{t}_w = \sum_{k=1}^m w_k t_k$, and $\hat{\beta} = \sum_{k=1}^m w_k z_k (t_k - \bar{t}_w) / \sum_{k=1}^m w_k (t_k - \bar{t}_w)^2$. Let $d_0 = 0$ and

$$d_k = \frac{1}{a_k}\left(1 - \frac{(t_k - \bar{t}_w)\bar{t}_w}{\sum_{k=1}^m w_k (t_k - \bar{t}_w)^2}\right), \quad k = 1, \ldots, m. \tag{10}$$

The quadratic form of $\hat{\sigma}_1^2$ can be represented as $\hat{\sigma}_1^2 = \mathbf{y}' D\mathbf{y} / \operatorname{tr}(D\Sigma)$, where $D = (d_{ij})_{n\times n}$ is a symmetric matrix with $d_{ij} = \sum_{k=1}^m d_k + \sum_{k=0}^{\min\{i-1, n-i, m\}} d_k$ for $i = j$, $d_{ij} = -d_{|i-j|}$ for $0 < |i-j| \le m$, and $d_{ij} = 0$ otherwise.

For the second method, we fit the linear regression with two independent variables $a_k$ and $b_k$ and with no intercept term. Specifically, we fit

$$\hat{\sigma}_R^2(k) = \beta_1 a_k + \beta_2 b_k + \xi_k, \quad k = 1, 2, \ldots, m, \tag{11}$$

where $\xi_k$ are the random errors associated with the linear regression. We then estimate $\sigma^2$ as the fitted slope $\hat{\beta}_1$. For ease of notation, let $s_k = \hat{\sigma}_R^2(k)$. By minimizing the weighted sum of squares $\sum_{k=1}^m w_k (s_k - \beta_1 a_k - \beta_2 b_k)^2$, we have the second estimator of $\sigma^2$ as

$$\begin{aligned}\hat{\sigma}_2^2 &= \hat{\beta}_1 \\ &= \frac{\sum_{k=1}^m w_k a_k b_k \sum_{k=1}^m w_k b_k s_k - \sum_{k=1}^m w_k b_k^2 \sum_{k=1}^m w_k a_k s_k}{\left(\sum_{k=1}^m w_k a_k b_k\right)^2 - \sum_{k=1}^m w_k a_k^2 \sum_{k=1}^m w_k b_k^2}.\end{aligned} \tag{12}$$

Let $v_0 = 0$ and

$$v_k = \frac{b_k \sum_{k=1}^m w_k a_k b_k - a_k \sum_{k=1}^m w_k b_k^2}{\left(\sum_{k=1}^m w_k a_k b_k\right)^2 - \sum_{k=1}^m w_k a_k^2 \sum_{k=1}^m w_k b_k^2}, \quad k = 1, \ldots, m. \tag{13}$$

It is easy to verify that $\hat{\sigma}_2^2$ has the quadratic form $\hat{\sigma}_2^2 = \mathbf{y}' H\mathbf{y} / \operatorname{tr}(H\Sigma)$, where $H = (h_{ij})_{n\times n}$ is a symmetric matrix with $h_{ij} = \sum_{k=1}^m v_k + \sum_{k=0}^{\min\{i-1, n-i, m\}} v_k$ for $i = j$, $h_{ij} = -v_{|i-j|}$ for $0 < |i-j| \le m$, and $h_{ij} = 0$ otherwise.

## 3. Main Results

This section investigates the statistical properties of the proposed least squares estimators. Note that $\hat{\sigma}_1^2$ in (9) and $\hat{\sigma}_2^2$ in (12) are two similar estimators, except that (9) treats $e_k$ as i.i.d. random errors and (12) treats $\xi_k = a_k e_k$ as i.i.d. random errors. For simplicity, in what follows, we present the asymptotic results for $\hat{\sigma}_1^2$ only. To evaluate the achievement of the proposed estimators, we will also investigate the behavior of $\hat{\sigma}_{\mathrm{TW}}^2$ in Tong and Wang [20] under the new model (4). Recall that for $\hat{\sigma}_{\mathrm{TW}}^2$, we have

$$\hat{\sigma}_{\mathrm{TW}}^2 = \frac{\mathbf{y}' \widetilde{D}\mathbf{y}}{2N}, \tag{14}$$

where $\tilde{d}_k = 1 - \bar{b}_w (b_k - \bar{b}_w) / \sum_{k=1}^m w_k (b_k - \bar{b}_w)^2$, $\bar{b}_w = \sum_{k=1}^m w_k b_k$, and $\widetilde{D} = (\tilde{d}_{ij})_{n\times n}$ is a symmetric matrix with $\tilde{d}_{ij} = \sum_{k=1}^m \tilde{d}_k + \sum_{k=0}^{\min\{i-1, n-i, m\}} \tilde{d}_k$ for $1 \le i = j \le n$, $\tilde{d}_{ij} = -\tilde{d}_{|i-j|}$ for $0 < |i-j| \le m$, and $\tilde{d}_{ij} = 0$ otherwise.

**Theorem 1.** *For the equally spaced design, the estimator $\hat{\sigma}_1^2$ in (9) is an unbiased estimator of $\sigma^2$ when $f$ is a linear function, regardless of the choice of $m$ and $c_i$. Under the same setting, however, the estimator $\hat{\sigma}_{TW}^2$ in (14) does not preserve the unbiasedness property. More specifically, the bias term of $\hat{\sigma}_{TW}^2$ has the expression*

$$\operatorname{Bias}\left(\hat{\sigma}_{TW}^2\right) = \left\{\frac{1}{2N}\sum_{k=1}^m \tilde{d}_k \sum_{i=k+1}^n (c_i + c_{i-k}) - 1\right\}\sigma^2. \tag{15}$$

**Theorem 2.** *Assume that $f$ has a bounded second derivative and $E(\varepsilon^4) < \infty$ with $\varepsilon = \varepsilon_1/\sqrt{c_1}$. When $\max_{1\le i\le n} c_i = O(1)$, for any $m = n^r$ with $0 < r < 1$ and the equally spaced design, then*

$$\begin{aligned}\operatorname{Bias}\left(\hat{\sigma}_1^2\right) &= O\left(\frac{m^2}{n^2}\right), \\ \operatorname{Var}\left(\hat{\sigma}_1^2\right) &= \frac{C_1}{n}\operatorname{Var}\left(\varepsilon^2\right) + o\left(\frac{1}{n}\right),\end{aligned} \tag{16}$$

*where*

$$C_1 = \frac{n}{4N^2}\sum_{i=1}^n c_i^2 \left(\sum_{k=1}^m d_k + \sum_{j=0}^{\min(i-1, n-i, m)} d_j\right)^2. \tag{17}$$

*As a comparison, the bias and variance of $\hat{\sigma}_{TW}^2$ are*

$$\begin{aligned}\operatorname{Bias}\left(\hat{\sigma}_{TW}^2\right) &= \left\{\frac{1}{2N}\sum_{k=1}^m \tilde{d}_k \sum_{i=k+1}^n (c_i + c_{i-k}) - 1\right\}\sigma^2 \\ &\quad + o\left(\frac{m^2}{n^2}\right), \\ \operatorname{Var}\left(\hat{\sigma}_{TW}^2\right) &= \frac{C_2}{n}\operatorname{Var}\left(\varepsilon^2\right) + o\left(\frac{1}{n}\right),\end{aligned} \tag{18}$$

*where*

$$C_2 = \frac{n}{4N^2}\sum_{i=1}^n c_i^2 \left(\sum_{k=1}^m \tilde{d}_k + \sum_{j=0}^{\min(i-1, n-i, m)} \tilde{d}_j\right)^2. \tag{19}$$

**Theorem 3.** *Assume that $f$ has a bounded second derivative and $E(\varepsilon^4) < \infty$ with $\varepsilon = \varepsilon_1/\sqrt{c_1}$. When $\max_{1\le i\le n} c_i = O(n^s)$ with $0 < s < 2/5$, for any $m = n^r$ with $0 < r < 1$ and the equally spaced design, then*

$$\operatorname{Bias}\left(\hat{\sigma}_1^2\right) = O\left(\frac{m^2}{n^2}\right), \qquad \operatorname{Var}\left(\hat{\sigma}_1^2\right) = O\left(\frac{1}{n^{1-s}}\right). \tag{20}$$

*As a comparison,*

$$\operatorname{Bias}\left(\hat{\sigma}_{TW}^2\right) = O\left(\frac{m}{n^{1-s}}\right), \qquad \operatorname{Var}\left(\hat{\sigma}_{TW}^2\right) = O\left(\frac{1}{n^{1-s}}\right). \tag{21}$$

**Theorem 4.** *Assume that $f$ has a bounded second derivative and $E(\varepsilon^4) < \infty$ with $\varepsilon = \varepsilon_1/\sqrt{c_1}$. For $\max_{1\le i\le n} c_i = O(n^s)$ with $0 < s < 1/4$ and any $m = n^r$ with $0 < r < 1/2$, then*

$$\sqrt{\frac{n}{\delta}}\left(\widehat{\sigma}_1^2 - \sigma^2\right) \xrightarrow{d} N\left(0, \left(\gamma_4 - 1\right)\sigma^4\right), \quad \text{as } n \longrightarrow \infty, \tag{22}$$

*where $\delta = \sum_{i=1}^n c_i^2/n$, $\gamma_4 = E(\varepsilon_1^4)/(c_1^2\sigma^4)$, and $\xrightarrow{d}$ denotes convergence in distribution.*

The proofs of the theorems are given in Section 6, respectively. Theorems 1 and 2 indicate that $\widehat{\sigma}_1^2$ is an unbiased or asymptotically unbiased estimator of $\sigma^2$ whereas $\widehat{\sigma}_{\mathrm{TW}}^2$ is not. The comparison on the asymptotic variances, or equivalently on $C_1$ and $C_2$, will be presented in Section 4. Furthermore, when the heteroscedasticity level is high, Theorem 3 shows that the bias term of $\widehat{\sigma}_{\mathrm{TW}}^2$ is getting more severe so that it does not remain to be a consistent estimator. The asymptotic normality in Theorem 4 can be used to construct confidence intervals for $\sigma^2$. When $n > \delta(\gamma_4 - 1)z_{\alpha/2}^2$, an approximate $1-\alpha$ confidence interval for $\sigma^2$ is

$$\left(\frac{\widehat{\sigma}_1^2}{\left\{1 + z_{\alpha/2}\sqrt{\delta\left(\gamma_4 - 1\right)/n}\right\}}, \frac{\widehat{\sigma}_1^2}{\left\{1 - z_{\alpha/2}\sqrt{\delta\left(\gamma_4 - 1\right)/n}\right\}}\right), \tag{23}$$

where $z_\alpha$ is the upper $\alpha$-th percentile of the standard normal distribution. When $\varepsilon_i$ are from normal distribution with variance $c_i\sigma^2$, we have $\gamma_4 = 3$ so that the confidence interval is fully specified. In general, we need to give an estimate for the unknown $\gamma_4$.

## 4. Simulation Studies

In this section, we conduct simulation studies to evaluate the finite sample performance of the proposed estimators, $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$. Their performance will also be compared with the estimator $\widehat{\sigma}_{\mathrm{TW}}^2$. Let $x_i = i/n$ for $i = 1, \ldots, n$. Throughout the simulations, we choose the bandwidth $m = n^{1/3}$, as suggested in Tong and Wang [20].

Our first simulation study considers only one $c_i$ value being different from the others. Specifically, for a given location $j$, we let $c_j = nc/(c+n-1)$ and $c_i = n/(c+n-1)$ for any $i \neq j$, where $c$ is a constant. Note that $\sum_{i=1}^n c_i = n$ is satisfied. In this study, we let $c = 30$. To investigate the behavior of the estimators along with the variance pattern, we consider the mean function $f = 5x$ and $f = 5\sin(2\pi x)$ and $\sigma = 0.5$ and $\sigma = 2$, respectively. Given the $c_i$ and $\sigma$ values, we then simulate $\varepsilon_i$ independently from $N(0, c_i\sigma^2)$. With 1000 repetitions, we plot the relative mean squared errors, $\mathrm{MSE}/(2\sigma^4/n)$, along with the location $j$ for $n = 30$ in Figure 1. It is evident that our estimators $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ perform better than $\widehat{\sigma}_{\mathrm{TW}}^2$ in most locations. To check the behavior near the boundary, we also plot the values of $C_1$ and $C_2$ along with the location $j$ for $n = 30, 50, 100$ (chosen $c = 30$) and $n = 500$ (chosen $c = 100$) in Figure 2. Combining Figures 1 and 2, we recommend the use of the new estimators when no significant different variance appears in the boundaries.

Our second simulation study is to investigate the average improvement of $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ over $\widehat{\sigma}_{\mathrm{TW}}^2$ when one or more variances are different from the others. To proceed, we consider three mean functions,

$$\begin{aligned} f_1(x) &= 5\sin(\pi x), \\ f_2(x) &= 5\sin(2\pi x), \\ f_3(x) &= 5\sin(4\pi x), \end{aligned} \tag{24}$$

two standard deviations, $\sigma = 0.5$ and 2, and three sample sizes, $n = 30$, 100, and 500, respectively. In total, there are 18 combinations. The $c$ values corresponding to $n = 30, 100$ and 500 are $c = 30$, 100, and 200, respectively. We then randomly sample (i) one location or (ii) five locations from the set $\{m, \ldots, n-m\}$ without replacement. For (i), the choice of the $c_i$ values follows the previous study. For (ii) with the five locations $\mathscr{L} = \{j_1, \ldots, j_5\}$, we let $c_j = nc/(5c + n - 5)$ for $j \in \mathscr{L}$ and $c_i = n/(c + n - 1)$ for $i \notin \mathscr{L}$. This results in $\sum_{i=1}^n c_i = n$. For each combination setting, we repeat the simulation 1000 times and report the relative MSEs in Table 1 for (i) and in Table 2 for (ii). From the simulation results, we observe that $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ have smaller relative MSEs than $\widehat{\sigma}_{\mathrm{TW}}^2$ in all the settings. In addition, we note that the performances of $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ are almost identical.

## 5. Conclusion

In this paper, we have proposed two least squares estimators for the error variance in heteroscedastic nonparametric regression: the intercept estimator and the slope estimator. Both estimators are shown to be consistent and their asymptotic properties are investigated, including the consistency and the asymptotic normalities. Simulation studies indicate that the proposed estimators perform better than the existing competitor in most settings. In the boundaries, however, we note that the proposed estimators behave not as well as expected when significantly different variances appear in the boundaries of design points. As a practical rule, we have suggested adopting the boundaries as $[1, m)$ and $(n-m, n]$. Further research may be necessary in this direction.

## 6. Proofs

This section provides the technical proofs of the theorems in Section 3. To prove the theorems, we first establish two lemmas. For ease of notation, let $f_i = f(x_i)$ and $f_i' = f'(x_i)$.

**Lemma 5.** *Assume that $f$ has a bounded second derivative. When $\max_{1\le i\le n} c_i = O(1)$, for any $m = n^r$ with $0 < r < 1$ and the equally spaced design, then*

(a) $\sum_{k=1}^m k^t d_k = O(m^{t+1})$, $t = 0, 1, 2, 3$;

(b) $\mathbf{f}'D\mathbf{f} = O(m^3/n)$;

(c) $\mathbf{f}'D\Sigma D'\mathbf{f} = o(m^4/n)$;

(d) $\mathbf{f}'\widetilde{D}\mathbf{f} = o(m^3/n)$;

(e) $\mathbf{f}'\widetilde{D}\Sigma\widetilde{D}'\mathbf{f} = o(m^4/n)$.

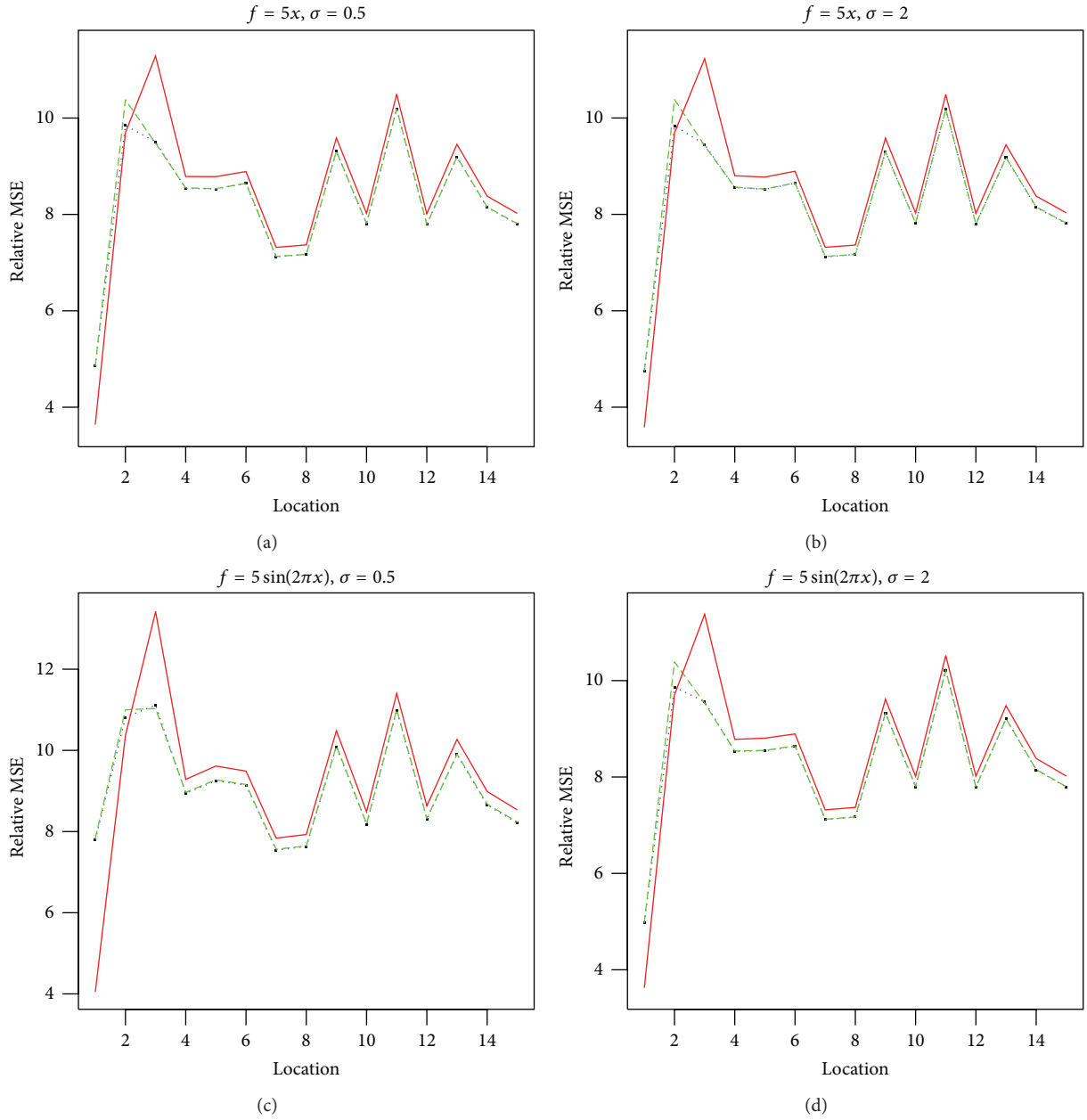


FIGURE 1: Plots of the relative MSEs of the estimators versus the location $j$ for $n = 30$. The dotted, dashed, and solid lines correspond to $\widehat{\sigma}_1^2$, $\widehat{\sigma}_2^2$, and $\widehat{\sigma}_{\mathrm{TW}}^2$, respectively.

*Proof.* (a) For simplicity, we prove only for $t = 2$. Let $u_k = 1 - ((t_k - \bar{t}_w)\bar{t}_w / \sum_{k=1}^m w_k(t_k - \bar{t}_w)^2)$ and $v = \bar{t}_w / \sum_{k=1}^m w_k(t_k - \bar{t}_w)^2$. By the definition of $a_k$ and $\sum_{i=1}^n c_i = n$, we have

$$\frac{n-k}{n} < \frac{1}{a_k} < \frac{2(n-k)}{n}, \quad k = 1, 2, \ldots, m. \tag{25}$$

First, we consider the upper bound of $\sum_{k=1}^m w_k(t_k - \bar{t}_w)^2$. We know

$$\sum_{k=1}^m w_k(t_k - \bar{t}_w)^2 = \sum_{k=1}^m w_k t_k^2 - \bar{t}_w^2 < \sum_{k=1}^m w_k t_k^2. \tag{26}$$

Thus, we have

$$\sum_{k=1}^m w_k(t_k - \bar{t}_w)^2 < \frac{4m^4}{5n^4} + o\left(\frac{m^4}{n^4}\right). \tag{27}$$

Next, we consider the lower bound of $\sum_{k=1}^m w_k(t_k - \bar{t}_w)^2$. By the definition, we can know

$$\begin{aligned} \frac{k^2(n-k)}{n^3} < t_k < \frac{2k^2(n-k)}{n^3}, \\ \frac{m^2}{3n^2} + o\left(\frac{m^2}{n^2}\right) < \bar{t}_w < \frac{2m^2}{3n^2} + o\left(\frac{m^2}{n^2}\right). \end{aligned} \tag{28}$$

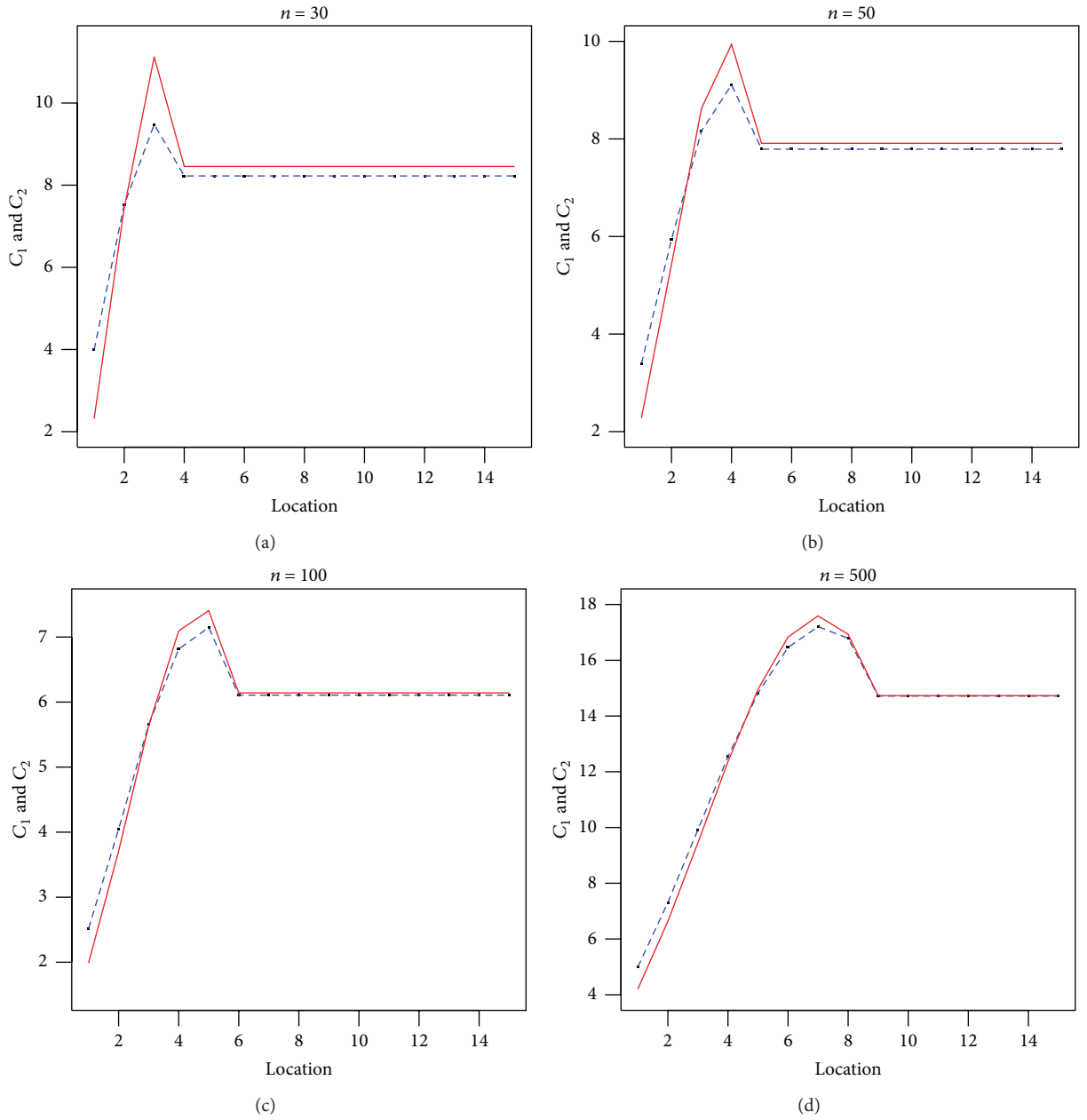


FIGURE 2: Plots of $C_1$ in $\mathrm{Var}(\widehat{\sigma}_1^2)$ and $C_2$ in $\mathrm{Var}(\widehat{\sigma}_{\mathrm{TW}}^2)$ versus the location $j$, under various $n$. The dashed and solid lines correspond to $C_1$ and $C_2$, respectively.

Let $\alpha_k = k^2(n-k)/n^3 - 2m^2/3n^2$ and $\beta_k = 2k^2(n-k)/n^3 - m^2/3n^2$. Then, $\alpha_k$ and $\beta_k$ are monotonically increasing about $k$, and

$$\alpha_k + o\left(\frac{m^2}{n^2}\right) < t_k - \bar{t}_w < \beta_k + o\left(\frac{m^2}{n^2}\right), \quad k = 1, 2, \ldots, m. \tag{29}$$

Note that $\beta_k$ is a monotonically increasing function of $k$ for $1 \le k \le m = o(n)$ with $\beta_1 < 0$ and $\beta_m > 0$. Let $m_1$ be the unique integer such that $\beta_{m_1} \le 0$ and $\beta_{m_1+1} > 0$. Therefore, we have

$$\begin{aligned}\sum_{k=1}^{m} w_k (t_k - \bar{t}_w)^2 &> \sum_{k=1}^{m_1} w_k (t_k - \bar{t}_w)^2 \\ &> \sum_{k=1}^{m_1} w_k \beta_k^2 + o\left(\frac{m^4}{n^4}\right) \\ &= \frac{4m_1^4}{5n^4} - \frac{4mm_1^3}{9n^4} + \frac{m_1 m^3}{9n^4} + o\left(\frac{m^4}{n^4}\right).\end{aligned} \tag{30}$$

Table 1: The relative MSEs of $\widehat{\sigma}_1^2$, $\widehat{\sigma}_2^2$, and $\widehat{\sigma}_{\mathrm{TW}}^2$ for scenario (i) under various simulation settings.

| $n$ | $\sigma$ | $f$ | $\widehat{\sigma}_1^2$ | $\widehat{\sigma}_2^2$ | $\widehat{\sigma}_{\mathrm{TW}}^2$ |
|---|---|---|---|---|---|
| 30 | 0.5 | $f_1$ | 8.592259 | 8.588712 | 8.971043 |
| | | $f_2$ | 9.105655 | 9.109716 | 9.59135 |
| | | $f_3$ | 27.52861 | 27.93159 | 29.50932 |
| | 2 | $f_1$ | 8.568082 | 8.564596 | 8.922776 |
| | | $f_2$ | 8.577039 | 8.57355 | 8.937509 |
| | | $f_3$ | 8.655637 | 8.653321 | 9.055993 |
| 100 | 0.5 | $f_1$ | 24.68565 | 24.67358 | 25.16409 |
| | | $f_2$ | 24.63531 | 24.63209 | 25.13371 |
| | | $f_3$ | 25.1026 | 25.10747 | 25.67267 |
| | 2 | $f_1$ | 24.64412 | 24.63515 | 25.1233 |
| | | $f_2$ | 24.62441 | 24.61774 | 25.1069 |
| | | $f_3$ | 24.62372 | 24.61822 | 25.11145 |
| 500 | 0.5 | $f_1$ | 47.85348 | 47.85206 | 48.00883 |
| | | $f_2$ | 47.86474 | 47.8618 | 48.01908 |
| | | $f_3$ | 47.88284 | 47.87881 | 48.03735 |
| | 2 | $f_1$ | 47.85314 | 47.85136 | 48.00809 |
| | | $f_2$ | 47.85561 | 47.85345 | 48.01027 |
| | | $f_3$ | 47.85829 | 47.85587 | 48.01283 |

Table 2: The relative MSEs of $\widehat{\sigma}_1^2$, $\widehat{\sigma}_2^2$, and $\widehat{\sigma}_{\mathrm{TW}}^2$ for scenario (ii) under various simulation settings.

| $n$ | $\sigma$ | $f$ | $\widehat{\sigma}_1^2$ | $\widehat{\sigma}_2^2$ | $\widehat{\sigma}_{\mathrm{TW}}^2$ |
|---|---|---|---|---|---|
| 30 | 0.5 | $f_1$ | 5.356385 | 5.358625 | 5.76456 |
| | | $f_2$ | 5.825295 | 5.844773 | 6.410941 |
| | | $f_3$ | 24.08339 | 24.74582 | 27.14087 |
| | 2 | $f_1$ | 5.291982 | 5.292094 | 5.659067 |
| | | $f_2$ | 5.284907 | 5.285037 | 5.659967 |
| | | $f_3$ | 5.367277 | 5.370764 | 5.809319 |
| 100 | 0.5 | $f_1$ | 14.70915 | 14.70799 | 15.1483 |
| | | $f_2$ | 14.6843 | 14.68265 | 15.1374 |
| | | $f_3$ | 15.01971 | 15.02443 | 15.58368 |
| | 2 | $f_1$ | 14.71244 | 14.71141 | 15.14792 |
| | | $f_2$ | 14.70282 | 14.70165 | 15.13869 |
| | | $f_3$ | 14.69388 | 14.6925 | 15.13526 |
| 500 | 0.5 | $f_1$ | 42.08678 | 42.08626 | 42.35964 |
| | | $f_2$ | 42.09082 | 42.0904 | 42.36408 |
| | | $f_3$ | 42.10065 | 42.10021 | 42.37591 |
| | 2 | $f_1$ | 42.0839 | 42.08347 | 42.3567 |
| | | $f_2$ | 42.08482 | 42.08442 | 42.35764 |
| | | $f_3$ | 42.08672 | 42.08632 | 42.35965 |

Let $m_1 = cm$ with $0 < c < 1$. It is easy to verify that $c > 1/3$. Then, $4m_1^4/5n^4 - 4mm_1^3/9n^4 + m_1m^3/9n^4 = (m^4/n^4)((4/5)c^4 - (4/9)c^3 + (1/9)c)$. Let $g(c) = (4/5)c^4 - (4/9)c^3 + (1/9)c$. Since $g'(c) = (16/5)c^3 - (4/3)c^2 + (1/9) > 0$ for $1/3 < c < 1$, then $g(c) > 0$ for $1/3 < c < 1$. Then,

$$g(c)\frac{m^4}{n^4} + o\left(\frac{m^4}{n^4}\right) < \sum_{k=1}^{m} w_k\left(t_k - \bar{t}_w\right)^2 < \frac{4m^4}{5n^4} + o\left(\frac{m^4}{n^4}\right). \tag{31}$$

Consequently, we obtain $\bar{t}_w = O(m^2/n^2)$ and $v = O(n^2/m^2)$. Note that

$$\left|\sum_{k=1}^{m} k^2 d_k\right| \le \sum_{k=1}^{m} \frac{k^2}{a_k}|u_k| \le \left|1 + v\bar{t}_w\right| \sum_{k=1}^{m} \frac{k^2}{a_k} + v\sum_{k=1}^{m} \frac{k^2 t_k}{a_k}. \tag{32}$$

So, we can get $\sum_{k=1}^{m} k^2 d_k = O(m^3)$. (b) Note that

$$\begin{aligned}
\mathbf{f}'D\mathbf{f} &= \sum_{k=1}^{m}\left\{d_k \sum_{i=k+1}^{n}\left(f_i - f_{i-k}\right)^2\right\} \\
&= \sum_{k=1}^{m}\left\{d_k \sum_{i=k+1}^{n}\left(f_i'\frac{k}{n} + O\left(\frac{k^2}{n^2}\right)\right)^2\right\} \\
&= \sum_{k=1}^{m}\left[d_k\left\{\frac{k^2}{n^2}\sum_{i=k+1}^{n} f_i'^2 + O\left(\frac{(n-k)k^3}{n^3}\right)\right\}\right] \\
&= \frac{J}{n}\sum_{k=1}^{m} k^2 d_k + O\left(\frac{1}{n^2}\right)\sum_{k=1}^{m} k^3 d_k,
\end{aligned} \tag{33}$$

where $J = \int_0^1 f'(x)^2 dx$. By (a), for any $m = o(n)$, we have

$$\mathbf{f}'D\mathbf{f} = O\left(\frac{m^3}{n}\right) + O\left(\frac{m^4}{n^2}\right) = O\left(\frac{m^3}{n}\right). \tag{34}$$

(c) Let $d_0 = 0$, $f_0 = 0$. We know

$$\begin{aligned}
&\mathbf{f}'D\Sigma D'\mathbf{f} \\
&= \sum_{i=1}^{m} c_i\left\{-\sum_{j=0}^{i-1} d_j f_{i-j} + \left(\sum_{k=1}^{m} d_k + \sum_{j=0}^{i-1} d_j\right) f_i - \sum_{j=1}^{m} d_j f_{i+j}\right\}^2 \\
&\quad + \sum_{i=m+1}^{n-m} c_i\left\{-\sum_{j=1}^{m} d_j f_{i-j} + 2\sum_{k=1}^{m} d_k f_i - \sum_{j=1}^{m} d_j f_{i+j}\right\}^2 \\
&\quad + \sum_{i=n-m+1}^{n} c_i\left\{-\sum_{j=1}^{m} d_j f_{i-j} + \left(\sum_{k=1}^{m} d_k + \sum_{j=1}^{n-i} d_j\right) f_i \right. \\
&\qquad \left. - \sum_{j=1}^{n-i} d_j f_{i+j}\right\}^2 \\
&= \sum_{i=1}^{m} c_i\left(\sum_{j=i}^{m} j d_j\right)^2\left(\frac{f_i'}{n} + o\left(\frac{1}{n}\right)\right)^2 \\
&\quad + 4\sum_{i=m+1}^{n-m} c_i\left(o\left(\frac{1}{n}\right)\sum_{j=1}^{m} j d_j\right)^2 \\
&\quad + \sum_{i=n-m+1}^{n} c_i\left(\sum_{j=n-i+1}^{m} j d_j\right)^2\left(\frac{f_i'}{n} + o\left(\frac{1}{n}\right)\right)^2 \\
&= O\left(\frac{m^5}{n^2}\right) + o\left(\frac{m^4}{n}\right).
\end{aligned} \tag{35}$$

For any $m = o(n)$, we have

$$\mathbf{f}' D\Sigma D'\mathbf{f} = o\left(\frac{m^4}{n}\right). \tag{36}$$

(d) We know

$$\begin{aligned}\mathbf{f}'\widetilde{D}\mathbf{f} &= \sum_{k=1}^{m}\left\{\widetilde{d}_k \sum_{i=k+1}^{n}\left(f_i - f_{i-k}\right)^2\right\} \\ &= \frac{J}{n}\sum_{k=1}^{m}k^2\widetilde{d}_k + O\left(\frac{1}{n^2}\right)\sum_{k=1}^{m}k^3\widetilde{d}_k.\end{aligned} \tag{37}$$

Note that

$$\sum_{k=1}^{m}k^2\widetilde{d}_k = o\left(m^3\right), \qquad \sum_{k=1}^{m}k^3\widetilde{d}_k = O\left(m^4\right). \tag{38}$$

Thus,

$$\mathbf{f}'\widetilde{D}\mathbf{f} = o\left(\frac{m^3}{n}\right). \tag{39}$$

(e) We know

$$\begin{aligned}\mathbf{f}'\widetilde{D}\Sigma\widetilde{D}'\mathbf{f} &= \frac{1}{n^2}\sum_{i=1}^{m}c_i\left(f_i' + o\left(1\right)\right)^2\left(\sum_{k=i}^{m}k\widetilde{d}_k\right)^2 \\ &\quad + 4\sum_{i=m+1}^{n-m}c_i\left(\sum_{k=1}^{m}k\widetilde{d}_k\right)^2 o\left(\frac{1}{n^2}\right) \\ &\quad + \frac{1}{n^2}\sum_{i=n-m+1}^{n}c_i\left(f_i' + o\left(1\right)\right)^2\left(\sum_{k=n-i+1}^{m}k\widetilde{d}_k\right)^2.\end{aligned} \tag{40}$$

Note that

$$\sum_{k=1}^{m}k\widetilde{d}_k = O\left(m^2\right), \quad \sum_{k=i}^{m}k\widetilde{d}_k = O\left(m^2\right), \quad 1 \le i \le m. \tag{41}$$

Therefore, we have

$$\mathbf{f}'\widetilde{D}\Sigma\widetilde{D}'\mathbf{f} = o\left(\frac{m^4}{n}\right). \tag{42}$$

□

**Lemma 6.** *Assume that $f$ has a bounded second derivative. When $\max_{1\le i\le n}c_i = O(n^s)$ with $0 < s < 1$, for any $m = n^r$ with $0 < r < 1$ and the equally spaced design, then*

(i) $\sum_{k=1}^{m}k^t d_k = O(m^{t+1})$, $t = 0, 1, 2, 3$;

(ii) $\mathbf{f}' D\mathbf{f} = O(m^3/n)$;

(iii) $\mathbf{f}' D\Sigma D'\mathbf{f} = o(m^4/n^{1-s})$;

(iv) $\mathbf{f}'\widetilde{D}\mathbf{f} = o(m^3/n)$;

(v) $\mathbf{f}'\widetilde{D}\Sigma\widetilde{D}'\mathbf{f} = o(m^4/n^{1-s})$.

*Proof.* (i) Here we only consider the proof of $t = 2$. For $\max_{1\le i\le n}c_i = O(n^s)$, it is similar with (a) in Lemma 5 to verify that $\bar{t}_w = O(m^2/n^2)$ and $\upsilon = O(n^2/m^2)$. Thus, we can get

$$\sum_{k=1}^{m}k^2 d_k = O\left(m^3\right). \tag{43}$$

(ii) According to (i), we have

$$\mathbf{f}' D\mathbf{f} = \frac{J}{n}\sum_{k=1}^{m}k^2 d_k + O\left(\frac{1}{n^2}\right)\sum_{k=1}^{m}k^3 d_k = O\left(\frac{m^3}{n}\right). \tag{44}$$

(iii) By part (c) in Lemma 5, we know

$$\begin{aligned}\mathbf{f}' D\Sigma D'\mathbf{f} &= \frac{1}{n^2}\sum_{i=1}^{m}c_i\left(f_i' + o\left(1\right)\right)^2\left(\sum_{k=i}^{m}kd_k\right)^2 \\ &\quad + 4\sum_{i=m+1}^{n-m}c_i\left(\sum_{k=1}^{m}kd_k\right)^2 o\left(\frac{1}{n^2}\right) \\ &\quad + \frac{1}{n^2}\sum_{i=n-m+1}^{n}c_i\left(f_i' + o\left(1\right)\right)^2\left(\sum_{k=n-i+1}^{m}kd_k\right)^2.\end{aligned} \tag{45}$$

For any $m = o(n)$, we have

$$\mathbf{f}' D\Sigma D'\mathbf{f} = o\left(\frac{m^4}{n^{1-s}}\right). \tag{46}$$

(iv) By (d) in Lemma 5, we have

$$\mathbf{f}'\widetilde{D}\mathbf{f} = \frac{J}{n}\sum_{k=1}^{m}k^2\widetilde{d}_k + O\left(\frac{1}{n^2}\right)\sum_{k=1}^{m}k^3\widetilde{d}_k = o\left(\frac{m^3}{n}\right). \tag{47}$$

(v) Similar to (e) in Lemma 5, it is easy to get

$$\begin{aligned}\mathbf{f}'\widetilde{D}\Sigma\widetilde{D}'\mathbf{f} &= \frac{1}{n^2}\sum_{i=1}^{m}c_i\left(f_i' + o\left(1\right)\right)^2\left(\sum_{k=i}^{m}k\widetilde{d}_k\right)^2 \\ &\quad + 4\sum_{i=m+1}^{n-m}c_i\left(\sum_{k=1}^{m}k\widetilde{d}_k\right)^2 o\left(\frac{1}{n^2}\right) \\ &\quad + \frac{1}{n^2}\sum_{i=n-m+1}^{n}c_i\left(f_i' + o\left(1\right)\right)^2\left(\sum_{k=n-i+1}^{m}k\widetilde{d}_k\right)^2 \\ &= O\left(\frac{m^5}{n^{2-s}}\right) + o\left(\frac{m^4}{n}\right) = o\left(\frac{m^4}{n^{1-s}}\right).\end{aligned} \tag{48}$$

□

**Lemma 7** (see [23]). *Let $a_{n;tk}$ be entries of a real symmetric matrix $A_n = (a_{n;tk})$, let $\{Z_t, t = 1, \ldots, n\}$ be i.i.d. random variables, and $T_n = \sum_{t,k=1}^{n}a_{n;tk}Z_tZ_k$. Assume that $\|A_n\|_{sp}/\|A_n\| \to 0$ and $EZ_t^4 < \infty$; then*

$$\left(\mathrm{Var}\left(T_n\right)\right)^{-1/2}\left(T_n - ET_n\right) \xrightarrow{d} N\left(0, 1\right), \tag{49}$$

*where $\|A_n\| = (\sum_{t,k=1}^{n}a_{n;tk}^2)^{1/2}$ and $\|A_n\|_{sp} = \max_{\|x\|=1}\|A_nx\|$ are the Euclidean norm and the spectral norm of the matrix $A_n$, respectively.*

*6.1. Proof of Theorem 1.* Let $f(x) = \mu + \delta x$. It is easy to verify that

$$Ez_k = \sigma^2 + \frac{k^2}{2n^2 a_k}\delta^2 = \sigma^2 + \frac{1}{2}t_k\delta^2, \tag{50}$$

$$E\overline{z}_w = E\left(\sum_{k=1}^{m} w_k z_k\right) = \sigma^2 + \frac{1}{2}\overline{t}_w\delta^2. \tag{51}$$

By combining (50) and (51), we have

$$\begin{aligned} E\left(\widehat{\sigma}_1^2\right) &= E\left(\overline{z}_w - \overline{t}_w\widehat{\beta}\right) \\ &= E\left(\overline{z}_w\right) - \frac{\overline{t}_w}{\sum_{k=1}^{m} w_k\left(t_k - \overline{t}_w\right)^2} \\ &\quad \times \left\{\sum_{k=1}^{m} w_k t_k Ez_k - \overline{t}_w E\left(\overline{z}_w\right)\right\} \\ &= \sigma^2 + \frac{1}{2}\delta^2\overline{t}_w - \frac{\overline{t}_w}{\sum_{k=1}^{m} w_k\left(t_k - \overline{t}_w\right)^2} \\ &\quad \times \left(\sum_{k=1}^{m} w_k t_k^2 - \overline{t}_w^2\right)\frac{1}{2}\delta^2 = \sigma^2. \end{aligned} \tag{52}$$

This shows that $\widehat{\sigma}_1^2$ is an unbiased estimator of $\sigma^2$. In what follows we consider $\widehat{\sigma}_{\mathrm{TW}}^2$. Therefore,

$$E\left(\widehat{\sigma}_{\mathrm{TW}}^2\right) = \frac{\mathrm{tr}\left(\widetilde{D}\Sigma\right)}{2N}\sigma^2 + \frac{1}{2N}\mathbf{f}'\widetilde{D}\mathbf{f}. \tag{53}$$

For $f(x) = \mu + \delta x$, we have

$$\begin{aligned} \frac{1}{2N}\mathbf{f}'\widetilde{D}\mathbf{f} &= \frac{1}{2N}\sum_{k=1}^{m}\left\{\widetilde{d}_k\sum_{i=k+1}^{n}\left(f_i - f_{i-k}\right)^2\right\} \\ &= \frac{1}{2N}\sum_{k=1}^{m}\left(\widetilde{d}_k\sum_{i=k+1}^{n}\delta^2\frac{k^2}{n^2}\right) = \frac{\delta^2}{2N}\sum_{k=1}^{m}(n-k)\,b_k\widetilde{d}_k \\ &= \frac{\delta^2}{2}\sum_{k=1}^{m} w_k b_k\left(1 - \frac{\left(b_k - \overline{b}_w\right)\overline{b}_w}{\sum_{k=1}^{m} w_k\left(b_k - \overline{b}_w\right)^2}\right) = 0, \end{aligned} \tag{54}$$

$$\begin{aligned} \mathrm{tr}\left(\widetilde{D}\Sigma\right) &= \sum_{i=1}^{m} c_i\left(\sum_{k=1}^{m}\widetilde{d}_k + \sum_{k=0}^{i-1}\widetilde{d}_k\right) \\ &\quad + 2\sum_{i=m+1}^{n-m} c_i\sum_{k=1}^{m}\widetilde{d}_k + \sum_{i=n-m+1}^{n} c_i\left(\sum_{k=1}^{m}\widetilde{d}_k + \sum_{k=1}^{n-i}\widetilde{d}_k\right) \\ &= \sum_{k=1}^{m}\widetilde{d}_k\sum_{i=k+1}^{n}\left(c_i + c_{i-k}\right). \end{aligned} \tag{55}$$

By (54) and (55), we get

$$\mathrm{Bias}\left(\widehat{\sigma}_{\mathrm{TW}}^2\right) = \left\{\frac{\sum_{k=1}^{m}\widetilde{d}_k\sum_{i=k+1}^{n}\left(c_i + c_{i-k}\right)}{2N} - 1\right\}\sigma^2. \tag{56}$$

*6.2. Proof of Theorem 2.* It is easy to verify that $\mathrm{tr}(D\Sigma) = 2N$. This leads to

$$\begin{aligned} E\left(\widehat{\sigma}_1^2\right) &= E\left(\frac{1}{2N}\mathbf{y}'D\mathbf{y}\right) \\ &= \frac{1}{2N}E\left(\mathbf{f}'D\mathbf{f} + 2\mathbf{f}'D\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'D\boldsymbol{\varepsilon}\right) \\ &= \frac{1}{2N}\left\{\mathbf{f}'D\mathbf{f} + E\left(\boldsymbol{\varepsilon}'D\boldsymbol{\varepsilon}\right)\right\} \\ &= \sigma^2 + \frac{1}{2N}\mathbf{f}'D\mathbf{f}. \end{aligned} \tag{57}$$

Thus,

$$\mathrm{Bias}\left(\widehat{\sigma}_1^2\right) = E\left(\widehat{\sigma}_1^2\right) - \sigma^2 = \frac{1}{2N}\mathbf{f}'D\mathbf{f}. \tag{58}$$

By Lemma 5, for any $m = o(n)$, we have

$$\mathrm{Bias}\left(\widehat{\sigma}_1^2\right) = O\left(\frac{m^2}{n^2}\right). \tag{59}$$

In what follows, we calculate $\mathrm{Var}(\widehat{\sigma}_1^2)$. Note that

$$\begin{aligned} \mathrm{Var}\left(\widehat{\sigma}_1^2\right) &= \mathrm{Var}\left(\frac{1}{2N}\mathbf{y}'D\mathbf{y}\right) \\ &= \frac{1}{4N^2}\mathrm{Var}\left(\mathbf{f}'D\mathbf{f} + \boldsymbol{\varepsilon}'D\boldsymbol{\varepsilon} + 2\mathbf{f}'D\boldsymbol{\varepsilon}\right) \\ &= \frac{1}{4N^2}\left\{\mathrm{Var}\left(\boldsymbol{\varepsilon}'D\boldsymbol{\varepsilon}\right) + 4\mathbf{f}'D\,\mathrm{Var}\left(\boldsymbol{\varepsilon}\right)D'\mathbf{f}\right. \\ &\qquad \left. + 2\mathrm{Cov}\left(\boldsymbol{\varepsilon}'D\boldsymbol{\varepsilon}, 2\mathbf{f}'D\boldsymbol{\varepsilon}\right)\right\} \\ &= \frac{1}{4N^2}\left(I_1 + I_2 + I_3\right). \end{aligned} \tag{60}$$

For $I_1$, we have

$$\begin{aligned} I_1 &= \sum_{i=1}^{n} d_{ii}^2\,\mathrm{Var}\left(\varepsilon_i^2\right) + 4\sum_{1\le i<j\le n} d_{ij}^2 E\left(\varepsilon_i^2\varepsilon_j^2\right) \\ &= \sum_{i=1}^{n} c_i^2\left(\sum_{k=1}^{m} d_k + \sum_{j=0}^{\min(i-1,n-i,m)} d_j\right)^2\mathrm{Var}\left(\varepsilon^2\right) \\ &\quad + 4\sigma^4\sum_{k=1}^{m} d_k^2\sum_{i=1}^{n-k} c_i c_{i+k}, \end{aligned} \tag{61}$$

where $d_0 = 0, \varepsilon = \varepsilon_i/\sqrt{c_i}$. For $I_2$, by Lemma 5, for any $m = o(n)$, we have

$$I_2 = 4\mathbf{f}'D\,\mathrm{Var}\left(\boldsymbol{\varepsilon}\right)D'\mathbf{f} = 4\mathbf{f}'D\Sigma D'\mathbf{f}\sigma^2 = o\left(\frac{m^4}{n}\right). \tag{62}$$

Finally, we consider $I_3$,

$$\begin{aligned} I_3 &= 4E\left\{\mathbf{f}'D\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'D\boldsymbol{\varepsilon}\right\} \\ &= 4\left[\sum_{s=1}^{m}\left\{\frac{f_s'}{n}\sum_{k=s}^{m}kd_k + o\left(\frac{1}{n}\right)\sum_{k=s}^{m}kd_k\right\}\right. \\ &\qquad \times E\left\{\varepsilon_s\sum_{k=1}^{m}d_k\sum_{i=k+1}^{n}\left(\varepsilon_i-\varepsilon_{i-k}\right)^2\right\} \\ &\qquad + \sum_{s=m+1}^{n-m}E\left\{\varepsilon_s\sum_{k=1}^{m}d_k\sum_{i=k+1}^{n}\left(\varepsilon_i-\varepsilon_{i-k}\right)^2\right\}o\left(\frac{m^2}{n}\right) \\ &\qquad + \sum_{s=n-m+1}^{n}\left\{\frac{f_s'}{n}\sum_{k=n-s+1}^{m}kd_k + o\left(\frac{1}{n}\right)\sum_{k=n-s+1}^{m}kd_k\right\} \\ &\qquad \left.\times E\left\{\varepsilon_s\sum_{k=1}^{m}d_k\sum_{i=k+1}^{n}\left(\varepsilon_i-\varepsilon_{i-k}\right)^2\right\}\right] \\ &= O\left(\frac{m^4}{n}\right)+o\left(m^3\right). \end{aligned} \tag{63}$$

Combining (61), (62), and (63), we know

$$\begin{aligned} \operatorname{Var}\left(\widehat{\sigma}_1^2\right) &= \frac{\operatorname{Var}\left(\varepsilon^2\right)}{4N^2}\sum_{i=1}^{n}c_i^2\left(\sum_{k=1}^{m}d_k+\sum_{j=0}^{\min(i-1,n-i,m)}d_j\right)^2 \\ &\quad + \frac{\sigma^4}{N^2}\sum_{k=1}^{m}d_k^2\left(\sum_{i=1}^{n-k}c_ic_{i+k}\right) \\ &\quad + O\left(\frac{m^2}{n^3}\right)+o\left(\frac{m}{n^2}\right). \end{aligned} \tag{64}$$

Note that, for $\max_{1\le i\le n}c_i = O(1)$ and any $m = n^r$ with $0 < r < 1$, we have

$$\frac{\sigma^4}{N^2}\sum_{k=1}^{m}d_k^2\left(\sum_{i=1}^{n-k}c_ic_{i+k}\right) = o\left(\frac{1}{n}\right). \tag{65}$$

Therefore, we get

$$\operatorname{Var}\left(\widehat{\sigma}_1^2\right) = \frac{C_1}{n}\operatorname{Var}\left(\varepsilon^2\right)+o\left(\frac{1}{n}\right), \tag{66}$$

where

$$C_1 = \frac{n}{4N^2}\sum_{i=1}^{n}c_i^2\left(\sum_{k=1}^{m}d_k+\sum_{j=0}^{\min(i-1,n-i,m)}d_j\right)^2. \tag{67}$$

Let $U_1 = \operatorname{Var}(\boldsymbol{\varepsilon}'\widetilde{D}\boldsymbol{\varepsilon})$, $U_2 = \operatorname{Cov}(\boldsymbol{\varepsilon}'\widetilde{D}\boldsymbol{\varepsilon}, \mathbf{f}'\widetilde{D}\boldsymbol{\varepsilon})$. Then, we have

$$\begin{aligned} U_1 &= \sum_{i=1}^{n}c_i^2\left(\sum_{k=1}^{m}\widetilde{d}_k+\sum_{j=0}^{\min(i-1,n-i,m)}\widetilde{d}_j\right)^2\operatorname{Var}\left(\varepsilon^2\right) \\ &\quad + 4\sigma^4\sum_{k=1}^{m}\widetilde{d}_k^2\sum_{i=1}^{n-k}c_ic_{i+k}, \\ U_2 &= \sum_{s=1}^{m}\sum_{j=s}^{m}j\widetilde{d}_j\left\{\frac{f_s'}{n}+o\left(\frac{1}{n}\right)\right\}E\left\{\varepsilon_s\sum_{k=1}^{m}\widetilde{d}_k\sum_{i=k+1}^{n}\left(\varepsilon_i-\varepsilon_{i-k}\right)^2\right\} \\ &\quad + \sum_{s=m+1}^{n-m}E\left\{\varepsilon_s\sum_{k=1}^{m}\widetilde{d}_k\sum_{i=k+1}^{n}\left(\varepsilon_i-\varepsilon_{i-k}\right)^2\right\}o\left(\frac{m^2}{n}\right) \\ &\quad + \sum_{s=n-m+1}^{n}\sum_{j=n-s}^{m}j\widetilde{d}_j\left\{\frac{f_s'}{n}+o\left(\frac{1}{n}\right)\right\} \\ &\qquad \times E\left\{\varepsilon_s\sum_{k=1}^{m}\widetilde{d}_k\sum_{i=k+1}^{n}\left(\varepsilon_i-\varepsilon_{i-k}\right)^2\right\}. \end{aligned} \tag{68}$$

By (68) and (d) and (e) in Lemma 5, we can get

$$\begin{aligned} &\operatorname{Bias}\left(\widehat{\sigma}_{\mathrm{TW}}^2\right) \\ &= E\left(\widehat{\sigma}_{\mathrm{TW}}^2\right)-\sigma^2 = \frac{\operatorname{tr}\left(\widetilde{D}\Sigma\right)}{2N}\sigma^2+\frac{1}{2N}\mathbf{f}'\widetilde{D}\mathbf{f}-\sigma^2 \\ &= \left\{\frac{\sum_{k=1}^{m}\widetilde{d}_k\sum_{i=k+1}^{n}\left(c_i+c_{i-k}\right)}{2N}-1\right\}\sigma^2+o\left(\frac{m^2}{n^2}\right), \\ &\operatorname{Var}\left(\widehat{\sigma}_{\mathrm{TW}}^2\right) = \frac{1}{4N^2}\left\{U_1+4U_2+4\mathbf{f}'\widetilde{D}\Sigma\widetilde{D}'\mathbf{f}\sigma^2\right\} \\ &\qquad = \frac{C_2}{n}\operatorname{Var}\left(\varepsilon^2\right)+o\left(\frac{1}{n}\right), \end{aligned} \tag{69}$$

where $C_2 = (n/4N^2)\sum_{i=1}^{n}c_i^2(\sum_{k=1}^{m}\widetilde{d}_k+\sum_{j=0}^{\min(i-1,n-i,m)}\widetilde{d}_j)^2$ and $\widetilde{d}_0 = 0$. This completes the proof of the theorem.

*6.3. Proof of Theorem 3.* By Lemma 6, we know

$$\operatorname{Bias}\left(\widehat{\sigma}_1^2\right) = E\left(\widehat{\sigma}_1^2\right)-\sigma^2 = \frac{1}{2N}\mathbf{f}'D\mathbf{f} = O\left(\frac{m^2}{n^2}\right). \tag{70}$$

According to (63), for $\max_{1\le i\le n}c_i = O(n^s)$, we have

$$\operatorname{Cov}\left(\boldsymbol{\varepsilon}'D\boldsymbol{\varepsilon}, \mathbf{f}'D\boldsymbol{\varepsilon}\right) = O\left(\frac{m^4}{n^{1-(3/2)s}}\right)+o\left(m^3n^{(3/2)s}\right). \tag{71}$$

Note that, under the condition $\sum_{i=1}^{n}c_i = n$ and $\max_{1\le i\le n}c_i = O(n^s)$, it can be shown that

$$\sum_{i=1}^{n}c_i^2 = O\left(n^{1+s}\right). \tag{72}$$

In addition, by Cauchy-Schwarz inequality, we know

$$\sum_{k=1}^{m} d_k^2 \sum_{i=1}^{n-k} c_i c_{i+k} = O\left(mn^{1+s}\right). \tag{73}$$

Thus,

$$\begin{aligned}\operatorname{Var}\left(\widehat{\sigma}_1^2\right) &= \frac{1}{4N^2}\operatorname{Var}\left(\boldsymbol{\varepsilon}' D\boldsymbol{\varepsilon}\right) + \frac{1}{N^2}\mathbf{f}' D \operatorname{Var}(\boldsymbol{\varepsilon}) D'\mathbf{f} \\ &\quad + \frac{1}{N^2}\operatorname{Cov}\left(\boldsymbol{\varepsilon}' D\boldsymbol{\varepsilon}, \mathbf{f}' D\boldsymbol{\varepsilon}\right) \\ &= O\left(\frac{1}{n^{1-s}}\right) + O\left(\frac{1}{mn^{1-s}}\right) + o\left(\frac{m^2}{n^{3-s}}\right) \\ &\quad + o\left(\frac{m}{n^{2-(3/2)s}}\right).\end{aligned} \tag{74}$$

When $r$ and $s$ satisfy $r - s > 0$ and $r + (3/2)s \le 1$, namely, $0 < s < 2/5$ and $0 < r < 1$, then

$$\operatorname{Var}\left(\widehat{\sigma}_1^2\right) = O\left(\frac{1}{n^{1-s}}\right). \tag{75}$$

Next, we consider the order of the bias and variance of $\widehat{\sigma}_{\mathrm{TW}}^2$. By (55), we have

$$\operatorname{tr}\left(\widetilde{D}\Sigma\right) = \sum_{k=1}^{m} \widetilde{d}_k \sum_{i=k+1}^{n} \left(c_i + c_{i-k}\right). \tag{76}$$

Note that

$$\sum_{k=1}^{m} \widetilde{d}_k = m - \frac{5m^2}{16n} + o\left(\frac{m^2}{n}\right), \qquad \sum_{i=1}^{n} c_i = n. \tag{77}$$

By (iv) in Lemma 6, then

$$\begin{aligned}&\operatorname{Bias}\left(\widehat{\sigma}_{\mathrm{TW}}^2\right) \\ &= \left\{\left(\left(2n\sum_{k=1}^{m}\widetilde{d}_k - \sum_{k=1}^{m}\widetilde{d}_k\sum_{i=1}^{k}c_i \right.\right.\right. \\ &\qquad \left.\left.\left. - \sum_{k=1}^{m}\widetilde{d}_k \sum_{i=n-k+1}^{n} c_i\right)(2N)^{-1}\right) - 1\right\}\sigma^2 \\ &\quad + o\left(\frac{m^2}{n^2}\right) \\ &= \left\{-\frac{5m}{16n} - \left(\sum_{k=1}^{m}\widetilde{d}_k\sum_{i=1}^{k}c_i \right.\right. \\ &\qquad \left.\left. + \sum_{k=1}^{m}\widetilde{d}_k\sum_{i=n-k+1}^{n}c_i\right)(2N)^{-1}\right\}\sigma^2 \\ &\quad + o\left(\frac{m}{n}\right).\end{aligned} \tag{78}$$

Consequently, it shows that

$$\left|\operatorname{Bias}\left(\widehat{\sigma}_{\mathrm{TW}}^2\right)\right| \le \left\{\frac{5m}{16n} + \sum_{k=1}^{m} k\widetilde{d}_k \frac{O\left(n^s\right)}{N}\right\}\sigma^2 + o\left(\frac{m}{n}\right). \tag{79}$$

Thus, we get

$$\operatorname{Bias}\left(\widehat{\sigma}_{\mathrm{TW}}^2\right) = O\left(\frac{m}{n^{1-s}}\right). \tag{80}$$

By (68) and (v) in Lemma 6, for $0 < s < 2/5$ and $0 < r < 1$, it is similar with (74) to get

$$\begin{aligned}\operatorname{Var}\left(\widehat{\sigma}_{\mathrm{TW}}^2\right) &= \frac{1}{4N^2}\operatorname{Var}\left(\boldsymbol{\varepsilon}'\widetilde{D}\boldsymbol{\varepsilon}\right) + \frac{1}{N^2}\mathbf{f}'\widetilde{D}\operatorname{Var}(\boldsymbol{\varepsilon})\widetilde{D}'\mathbf{f} \\ &\quad + \frac{1}{N^2}\operatorname{Cov}\left(\boldsymbol{\varepsilon}'\widetilde{D}\boldsymbol{\varepsilon}, \mathbf{f}'\widetilde{D}\boldsymbol{\varepsilon}\right) \\ &= O\left(\frac{1}{n^{1-s}}\right).\end{aligned} \tag{81}$$

This completes the proof of the theorem.

*6.4. Proof of Theorem 4.* By Theorem 1, we know

$$\widehat{\sigma}_1^2 = \frac{1}{2N}\mathbf{y}' D\mathbf{y} = \frac{1}{2N}\mathbf{f}' D\mathbf{f} + \frac{1}{N}\mathbf{f}' D\boldsymbol{\varepsilon} + \frac{1}{2N}\boldsymbol{\varepsilon}' D\boldsymbol{\varepsilon}. \tag{82}$$

Note that the first term corresponds to the bias term. By Lemma 6, we know $(1/2N)\mathbf{f}' D\mathbf{f} = O(m^2/n^2)$. Thus, for any $m = n^r$ with $0 < r < 3/4$,

$$\frac{1}{2N}\mathbf{f}' D\mathbf{f} = o\left(n^{-1/2}\right). \tag{83}$$

For the second term, by Lemma 6, we have

$$E\left(\frac{\mathbf{f}' D\boldsymbol{\varepsilon}}{N}\right)^2 = \frac{\mathbf{f}' D\Sigma D'\mathbf{f}\sigma^2}{N^2} = o\left(\frac{m^2}{n^{3-s}}\right) = o\left(\frac{m^2}{n^2}\right). \tag{84}$$

Thus, for any $m = n^r$ with $0 < r < 1/2$, we have

$$\frac{\mathbf{f}' D\boldsymbol{\varepsilon}}{N} = o_p\left(n^{-1/2}\right). \tag{85}$$

Now we consider the third term. Let $\widetilde{\varepsilon}_i = \varepsilon_i/\sqrt{c_i}$, $\widetilde{\varepsilon} = (\widetilde{\varepsilon}_1, \widetilde{\varepsilon}_2, \ldots, \widetilde{\varepsilon}_n)^T$, and $C = \operatorname{diag}(\sqrt{c_1}, \sqrt{c_2}, \ldots, \sqrt{c_n})$, then $E(\widetilde{\varepsilon}_i) = 0$, $\operatorname{Var}(\widetilde{\varepsilon}_i) = \sigma^2$, and $\widetilde{\varepsilon}_i$ are i.i.d. random variables, $\boldsymbol{\varepsilon} = C\widetilde{\varepsilon}$. So we have

$$\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon} = \widetilde{\varepsilon}^T C^T DC\widetilde{\varepsilon} = \widetilde{\varepsilon}^T T\widetilde{\varepsilon} = \sum_{i,j=1}^{n} t_{ij}\widetilde{\varepsilon}_i\widetilde{\varepsilon}_j, \tag{86}$$

where $T = C^T DC = (t_{ij})_{n\times n}$ is a real symmetric matrix with $t_{ij} = c_i\left(\sum_{k=1}^{m} d_k + \sum_{k=0}^{\min\{i-1, n-i, m\}} d_k\right)$ for $i = j$, $t_{ij} = -\sqrt{c_i c_{i+|i-j|}}\, d_{|i-j|}$ for $0 < |i-j| \le m$, and $t_{ij} = 0$ otherwise.

We know that the Euclidean norm of the matrix $T$ can be denoted as

$$\begin{aligned}\|T\| &= \left(\sum_{i,j=1}^{n} t_{ij}^2\right)^{1/2} \\ &= \left\{\sum_{i=1}^{m} c_i^2\left(\sum_{k=1}^{m} d_k + \sum_{j=0}^{i-1} d_j\right)^2 \right. \\ &\quad + \sum_{i=m+1}^{n-m} c_i^2\left(2\sum_{k=1}^{m} d_k\right)^2 + \sum_{i=n-m+1}^{n} c_i^2\left(\sum_{k=1}^{m} d_k + \sum_{j=0}^{n-i} d_j\right)^2 \\ &\quad \left. +2\sum_{j=1}^{m}\left(\sum_{i=1}^{n-j} c_i c_{i+j} d_j^2\right)\right\}^{1/2},\end{aligned} \tag{87}$$

where $d_0 = 0$. By the definition of spectral norm, we have

$$\|T\|_{\mathrm{sp}} = \max_{\|x\|=1}\|Tx\| = \left(\text{maximum eigenvalue of } T^C T\right)^{1/2}, \tag{88}$$

where $T^C$ is the conjugate transpose of the matrix $T$. Since $T$ is a real symmetric matrix, $T^C = T$. Then, $T^C T = T^2$.

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be eigenvalues of matrix $T$ and ordered to be nonincreasing in absolute value. Then, $\lambda_i^2, i = 1, \ldots, n$ are eigenvalues of the matrix $T^2$. Thus, $|\lambda_1|$ is the spectral norm of matrix $T$. Namely, $\|T\|_{\mathrm{sp}} = |\lambda_1|$. Let

$$L_n^2 = \max_{1\le i\le n}\sum_{j=1}^{n} t_{ij}^2, \qquad \Gamma_n = \max_{1\le i\le n}\sum_{j=1}^{n}\left|t_{ij}\right|. \tag{89}$$

It is well known that

$$L_n \le \left|\lambda_1\right| \le \Gamma_n. \tag{90}$$

Let $\max_{1\le i\le n} c_i = O(n^s)$ and $\omega = \min_{1\le i\le m, n-m+1\le i\le n}\{|\sum_{k=1}^{m} d_k + \sum_{j=0}^{i-1} d_j|, 2|\sum_{k=1}^{m} d_k|, |\sum_{k=1}^{m} d_k + \sum_{j=0}^{n-i} d_j|\}$. Then, combining (87) and (90), we have

$$\begin{aligned}\frac{\|T\|_{\mathrm{sp}}}{\|T\|} &\le \frac{\max_{1\le i\le n}\sum_{j=1}^{n}\left|t_{ij}\right|}{\|T\|} \\ &\le \frac{\max_{1\le i\le n}\left(2c_i\left|\sum_{k=1}^{m} d_k\right| + 2\sum_{k=1}^{m}\sqrt{c_i c_{i+k}}\left|d_k\right|\right)}{\|T\|} \\ &\le \left(2\left(\left|\sum_{k=1}^{m} d_k\right| + \sum_{k=1}^{m}\left|d_k\right|\right)O\left(n^s\right)\right) \\ &\quad \times \left(\left\{\sum_{i=1}^{m} c_i^2\left(\sum_{k=1}^{m} d_k + \sum_{j=0}^{i-1} d_j\right)^2 \right.\right. \\ &\quad + \sum_{i=m+1}^{n-m} c_i^2\left(2\sum_{k=1}^{m} d_k\right)^2 \\ &\quad \left.\left. + \sum_{i=n-m+1}^{n} c_i^2\left(\sum_{k=1}^{m} d_k + \sum_{j=0}^{n-i} d_j\right)^2\right\}^{1/2}\right) \\ &\le \frac{2\left(\left|\sum_{k=1}^{m} d_k\right| + \sum_{k=1}^{m}\left|d_k\right|\right)}{\omega}\frac{O\left(n^s\right)}{\sqrt{\sum_{i=1}^{n} c_i^2}}.\end{aligned} \tag{91}$$

Note that $\sum_{i=1}^{n} c_i = n$. By Cauchy-Schwarz inequality, we have

$$\sum_{i=1}^{n} c_i^2 \ge n. \tag{92}$$

Note also that

$$\frac{2\left(\left|\sum_{k=1}^{m} d_k\right| + \sum_{k=1}^{m}\left|d_k\right|\right)}{\omega} = O(1). \tag{93}$$

Consequently, for $0 < s < 1/2$, we have

$$\frac{\|T\|_{\mathrm{sp}}}{\|T\|} \le \frac{2\left(\left|\sum_{k=1}^{m} d_k\right| + \sum_{k=1}^{m}\left|d_k\right|\right)}{\omega}\frac{O\left(n^s\right)}{n^{1/2}} \longrightarrow 0, \quad \text{as } n \longrightarrow \infty. \tag{94}$$

Note that $E\tilde{\varepsilon}^4 < \infty$. Then, by Lemma 7, we obtain

$$\mathrm{Var}\left(\tilde{\varepsilon}^T T\tilde{\varepsilon}\right)^{-1/2}\left(\tilde{\varepsilon}^T T\tilde{\varepsilon} - E\left(\tilde{\varepsilon}^T T\tilde{\varepsilon}\right)\right) \xrightarrow{d} N(0,1). \tag{95}$$

That is,

$$\mathrm{Var}\left(\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}\right)^{-1/2}\left(\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon} - E\left(\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}\right)\right) \xrightarrow{d} N(0,1), \tag{96}$$

where

$$\begin{aligned}E\left(\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}\right) &= \mathrm{tr}\left(D\Sigma\right)\sigma^2 = \sum_{k=1}^{m} d_k \sum_{i=k+1}^{n}\left(c_i + c_{i-k}\right)\sigma^2, \\ \mathrm{Var}\left(\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}\right) &= \left(\gamma_4 - 1\right)\sigma^4\sum_{i=1}^{n} d_{ii}^2 c_i^2 + 4\sum_{1\le i<j\le n} d_{ij}^2 c_i c_j \sigma^4 \\ &= \left(\gamma_4 - 1\right)\sigma^4\sum_{i=1}^{n}\left(\sum_{k=1}^{m} d_k + \sum_{k=0}^{\min(i-1, n-i, m)} d_k\right)^2 c_i^2 \\ &\quad + 4\sum_{k=1}^{m} d_k^2\left(\sum_{i=1}^{n-k} c_i c_{i+k}\right)\sigma^4.\end{aligned} \tag{97}$$

Hence, we get

$$\frac{\sqrt{n}\left((1/2N)\,\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}-\mu_0\sigma^2\right)}{\sigma_0}\xrightarrow{d} N(0,1), \tag{98}$$

where $\mu_0 = \sum_{k=1}^{m} d_k \sum_{i=k+1}^{n}(c_i + c_{i-k})/(2N)$, $\sigma_0^2 = n(\gamma_4 - 1)\sigma^4/4N^2 \sum_{i=1}^{n}(\sum_{k=1}^{m} d_k + \sum_{k=0}^{\min(i-1,n-i,m)} d_k)^2 c_i^2 + (n\sigma^4/N^2)\sum_{k=1}^{m} d_k^2(\sum_{i=1}^{n-k} c_i c_{i+k})$. Combining (83), (85), and (98), by Slutsky's theorem, we have

$$\frac{\sqrt{n}\left(\widehat{\sigma}_1^2-\mu_0\sigma^2\right)}{\sigma_0^2}\xrightarrow{d} N(0,1). \tag{99}$$

Note that $\mu_0 = 1 + O(m/n^{1-s})$ and $(n\sigma^4/N^2)\sum_{k=1}^{m} d_k^2 (\sum_{i=1}^{n-k} c_i c_{i+k}) = O(n^s/m)$. Then, when $(1/2) - r - s > 0$ and $r > s$, namely, $0 < s < 1/4$ and $0 < r < 1/2$, we have $\sqrt{n}(\mu_0 - 1) = o(1)$ and

$$\begin{aligned}\sigma_0^2 &= \frac{n(\gamma_4-1)\sigma^4}{4N^2}\sum_{i=1}^{n}\left(\sum_{k=1}^{m} d_k + \sum_{k=0}^{\min(i-1,n-i,m)} d_k\right)^2 c_i^2 + o(1)\\ &= \delta(\gamma_4-1)\sigma^4 + o(1),\end{aligned} \tag{100}$$

where $\delta = \sum_{i=1}^{n} c_i^2/n$. By Slutsky's theorem, we obtain

$$\begin{aligned}\frac{\sqrt{n}\left(\widehat{\sigma}_1^2-\sigma^2\right)}{\sqrt{\delta(\gamma_4-1)\sigma^4}} &= \frac{\sigma_0}{\sqrt{\delta(\gamma_4-1)\sigma^4}}\\ &\quad\times\left\{\frac{\sqrt{n}\left(\widehat{\sigma}_1^2-\mu_0\sigma^2\right)}{\sigma_0}+\frac{\sqrt{n}(\mu_0-1)\sigma^2}{\sigma_0}\right\}\\ &\xrightarrow{d} N(0,1), \quad \text{as } n\longrightarrow\infty.\end{aligned} \tag{101}$$

This proves the theorem.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

Yebin Cheng's research was supported in part by National Natural Science Foundation of China Grant no. 11271241 and Shanghai Leading Academic Discipline Project no. 863. Tiejun Tong's research was supported in part by Hong Kong Research Grant HKBU202711 and Hong Kong Baptist University FRG Grants FRG2/10-11/020 and FRG2/11-12/110. Yuejin Zhou's research was supported in part by Doctoral Innovation Foundation of SHUFE CXJJ-2011-442. The authors thank the editor, the associate editor, and the three reviewers for their constructive comments that substantially improved an earlier draft.

## References

[1] R. J. Carroll, "The effects of variance function estimation on prediction and calibration: an example," in *Statistical Decision Theory and Related Topics*, vol. 2, pp. 273–280, 1986.

[2] R. J. Carroll and D. Ruppert, *Transforming and Weighting in Regression*, Chapman & Hall, London, UK, 1988.

[3] J. Kay, "On the choice of the regularisation parameter in image restoration," in *Pattern Recognition*, vol. 301 of *Lecture Notes in Computer Science*, pp. 587–596, 1988.

[4] M. J. Buckley, G. K. Eagleson, and B. W. Silverman, "The estimation of residual variance in nonparametric regression," *Biometrika*, vol. 75, no. 2, pp. 189–199, 1988.

[5] R. L. Eubank and C. H. Spiegelman, "Testing the goodness of fit of a linear model via nonparametric regression techniques," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 387–392, 1990.

[6] T. Gasser, A. Kneip, and W. Köhler, "A flexible and fast method for automatic smoothing," *Journal of the American Statistical Association*, vol. 86, no. 415, pp. 643–652, 1991.

[7] W. Härdle and A. Tsybakov, "Local polynomial estimators of the volatility function in nonparametric autoregression," *Journal of Econometrics*, vol. 81, no. 1, pp. 223–242, 1997.

[8] K. B. Kulasekera and C. Gallagher, "Variance estimation in nonparametric multiple regression," *Communications in Statistics A: Theory and Methods*, vol. 31, no. 8, pp. 1373–1383, 2002.

[9] J. A. Rice, "Bandwidth choice for nonparametric regression," *The Annals of Statistics*, vol. 12, no. 4, pp. 1215–1230, 1984.

[10] T. Gasser, L. Sroka, and C. Jennen-Steinmetz, "Residual variance and residual pattern in nonlinear regression," *Biometrika*, vol. 73, no. 3, pp. 625–633, 1986.

[11] P. Hall, J. W. Kay, and D. M. Titterington, "Asymptotically optimal difference-based estimation of variance in nonparametric regression," *Biometrika*, vol. 77, no. 3, pp. 521–528, 1990.

[12] H. Dette, A. Munk, and T. Wagner, "Estimating the variance in nonparametric regression—what is a reasonable choice?" *Journal of the Royal Statistical Society B. Statistical Methodology*, vol. 60, no. 4, pp. 751–764, 1998.

[13] U. U. Müller, A. Schick, and W. Wefelmeyer, "Estimating the error variance in nonparametric regression by a covariate-matched *u*-statistic," *Statistics*, vol. 37, no. 3, pp. 179–188, 2003.

[14] T. Tong, A. Liu, and Y. Wang, "Relative errors of difference-based variance estimators in nonparametric regression," *Communications in Statistics A: Theory and Methods*, vol. 37, no. 18–20, pp. 2890–2902, 2008.

[15] J. Du and A. Schick, "A covariate-matched estimator of the error variance in nonparametric regression," *Journal of Nonparametric Statistics*, vol. 21, no. 3, pp. 263–285, 2009.

[16] L. Wang, L. D. Brown, and T. T. Cai, "A difference based approach to the semiparametric partial linear model," *Electronic Journal of Statistics*, vol. 5, pp. 619–641, 2011.

[17] L. D. Brown and M. Levine, "Variance estimation in nonparametric regression via the difference sequence method," *The Annals of Statistics*, vol. 35, no. 5, pp. 2219–2232, 2007.

[18] L. Wang, L. D. Brown, T. T. Cai, and M. Levine, "Effect of mean on variance function estimation in nonparametric regression," *The Annals of Statistics*, vol. 36, no. 2, pp. 646–664, 2008.

[19] T. T. Cai and L. Wang, "Adaptive variance function estimation in heteroscedastic nonparametric regression," *The Annals of Statistics*, vol. 36, no. 5, pp. 2025–2054, 2008.

[20] T. Tong and Y. Wang, "Estimating residual variance in nonparametric regression using least squares," *Biometrika*, vol. 92, no. 4, pp. 821–830, 2005.

[21] T. Tong, Y. Ma, and Y. Wang, "Optimal variance estimation without estimating the mean function," *Bernoulli*, vol. 19, no. 5, pp. 1839–1854, 2013.

[22] C. G. Park, I. Kim, and Y.-S. Lee, "Error variance estimation via least squares for small sample nonparametric regression," *Journal of Statistical Planning and Inference*, vol. 142, no. 8, pp. 2369–2385, 2012.

[23] R. J. Bhansali, L. Giraitis, and P. S. Kokoszka, "Approximations and limit theory for quadratic forms of linear processes," *Stochastic Processes and Their Applications*, vol. 117, no. 1, pp. 71–95, 2007.