

Supplemental materials for “On the choice of difference sequence in a unified framework for variance estimation in nonparametric regression”

Wenlin Dai¹, Tiejun Tong² and Lixing Zhu²

¹King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

²Department of Mathematics, Hong Kong Baptist University, Hong Kong

This document serves as a supplement to the main manuscript and is organized as follows. In Section S1, we describe in detail the proposed unified framework for estimating the residual variance. In Section S2, we derive the asymptotic results of the unified variance estimators. In Section S3, we conduct a numerical comparison on the bias terms of the unified estimators with the “state-of-the-art” estimator in the literature. In Section S4, we provide an alternative procedure for the unified estimation for unequally spaced design. Proofs of theoretical results are provided in Section S5.

S1 Unified framework

We consider the nonparametric regression model

$$Y_i = g(x_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

where $\{Y_i\}$ are the observations, g is an unknown mean function, $\{x_i\}$ are the design points, and $\{\varepsilon_i\}$ are the independent and identically distributed (i.i.d.) random errors with mean zero and variance σ^2 .

For the difference-based variance estimation, two popular methods in the literature are the higher-order difference estimator in Hall, Kay and Titterington (1990) and the linear regression estimator in Tong and Wang (2005). To combine the two methods

systematically, for any order- r difference sequence $d = (d_0, \dots, d_r)$ satisfying

$$\sum_{j=0}^r d_j = 0 \quad \text{and} \quad \sum_{j=0}^r d_j^2 = 1, \quad (\text{S1})$$

we define

$$s_k(r) = \frac{1}{n - rk} \sum_{i=1}^{n-rk} \left(\sum_{j=0}^r d_j Y_{i+jk} \right)^2. \quad (\text{S2})$$

Let $J(r) = (\sum_{j=0}^r j d_j)^2 \int_0^1 [g'(x)]^2 dx$, $\alpha = \sigma^2$, $\beta = J(r)$ and $h_k = k^2/n^2$, we have the approximately linear regression model $s_k(r) \approx \alpha + h_k \beta$. Then we fit the linear regression model by minimizing the following weighted sum of squares

$$\sum_{k=1}^m w_k (s_k(r) - \alpha - h_k \beta)^2, \quad \beta > 0,$$

where $w_k = (n - rk)/N$ are the corresponding weights with $N = \sum_{k=1}^m (n - rk) = nm - rm(m + 1)/2$. Finally, we estimate σ^2 by the fitted intercept in the unified framework.

This leads to the unified estimator as

$$\hat{\sigma}^2(r, m) = \hat{\alpha} = \sum_{k=1}^m b_k w_k s_k(r), \quad (\text{S3})$$

where $b_k = 1 - \bar{h}_w (h_k - \bar{h}_w) / (\sum_{k=1}^m w_k h_k^2 - \bar{h}_w^2)$ and $\bar{h}_w = \sum_{k=1}^m w_k h_k$. Further, if the optimal difference sequence is used in (S2), we refer to the unified estimator as the unified optimal estimator, denoted by $\hat{\sigma}_{\text{opt}}^2(r, m)$. Otherwise if the ordinary difference sequence is used, we refer to it as the unified ordinary estimator, denoted by $\hat{\sigma}_{\text{ord}}^2(r, m)$.

Theorem S1. *For the equally spaced design, we have (i) the proposed estimator $\hat{\sigma}^2(r, m)$ is an unbiased estimator for any difference sequence satisfying (S1) when $g(x)$ is a linear function; and (ii) for the ordinary difference sequence, the estimator is an unbiased estimator for any polynomial $g(x)$ with order up to $p \leq r - 1$.*

The proof of Theorem S1 is given in Section S5. Theorem S1 shows that the linear regression does provide merit in correcting the estimation bias, especially when the mean function is not very rough.

S2 Asymptotic results

We first derive the asymptotic bias of the unified estimator (S3). To cater for the demand of difference sequence selection, we also specify the relevant results for the optimal difference sequence and ordinary difference sequence, respectively. By the following two theorems, we observe that the unified estimator $\hat{\sigma}^2(r, m)$ is capable to control the bias to order $O(m^3/n^3)$ for the optimal difference sequence, and to order $O(m^{2r}/n^{2r})$ for the ordinary difference sequence. This demonstrates that the linear regression with the ordinary difference sequence provides a smaller asymptotic bias than the linear regression with the optimal difference sequence for any $r \geq 2$.

Theorem S2. *Consider the equally spaced design with $m \rightarrow \infty$ and $m/n \rightarrow 0$, and assume that $g(x)$ has a bounded second derivative. Then for any difference sequence (d_0, \dots, d_r) under the constraint (S1) with $r \geq 1$,*

$$E(\hat{\sigma}^2(r, m)) = \sigma^2 - \frac{C_1}{16} [rC_1 \int_0^1 [g'(t)]^2 dt - (rC_1 - C_2) [g'(1)]^2 - C_2 [g'(0)]^2] \frac{m^3}{n^3} + o\left(\frac{m^3}{n^3}\right),$$

where $C_i = \sum_{j=0}^r j^i d_j / i!$ for $i = 1$ and 2 .

Theorem S3. *Consider the equally spaced design with $m \rightarrow \infty$ and $m/n \rightarrow 0$, and assume that $g(x)$ has a bounded r th derivative. Then for the ordinary difference sequence with $r \geq 2$,*

$$E(\hat{\sigma}_{\text{ord}}^2(r, m)) = \sigma^2 + \left[\frac{3(1-r)}{(2r+1)(2r+3)\binom{2r}{r}} \int_0^1 [g^{(r)}(x)]^2 dx \right] \frac{m^{2r}}{n^{2r}} + o\left(\frac{m^{2r}}{n^{2r}}\right).$$

Now we investigate the asymptotic variance of the unified estimator. Recall that an order of $r \geq 3$ is rarely recommended for practical use even for $m = 1$ (Dette, Munk and Wagner 1998). Hence, for the sake of brevity, we present the asymptotic variance of the unified estimator only for $r = 2$, i.e., the minimum order that distinguishes the optimal difference sequence from the ordinary difference sequence. Using a higher order with $r \geq 3$ will yield more tedious derivation and more complex solutions, yet the comparison results

remain the same. With $r = 2$, the unified estimator can be represented as

$$\hat{\sigma}^2(2, m) = \sum_{k=1}^m b_k w_k s_k(2) = Y^T D Y / \text{tr}(D), \quad (\text{S4})$$

where $Y = (Y_1, \dots, Y_n)^T$, $\text{tr}(\cdot)$ is the trace, and $D = (d_{ij})_{n \times n}$ is an $n \times n$ matrix with the diagonal elements

$$d_{ii} = d_0^2 \sum_{k=1}^{\min(m, \lfloor \frac{n-i}{2} \rfloor)} b_k + d_1^2 \sum_{k=1}^{\min(m, n-i, i-1)} b_k + d_2^2 \sum_{k=1}^{\min(m, \lfloor \frac{i-1}{2} \rfloor)} b_k, \quad i = 1, \dots, n,$$

and the off-diagonal elements

$$d_{ij} = \begin{cases} d_0 d_1 b_k I_{(1 \leq k \leq m)} + d_0 d_2 b_{k/2} I_{(k \in E)} & 1 \leq |i - j| = k \leq 2m, 1 \leq \min(i, j) \leq k, \\ -d_1^2 b_k I_{(1 \leq k \leq m)} + d_0 d_2 b_{k/2} I_{(k \in E)} & 1 \leq |i - j| = k \leq 2m, k + 1 \leq i, j \leq n - k, \\ d_1 d_2 b_k I_{(1 \leq k \leq m)} + d_0 d_2 b_{k/2} I_{(k \in E)} & 1 \leq |i - j| = k \leq 2m, n - k + 1 \leq \max(i, j) \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

Here, E is the set of positive even integers, I is the indicator function, and $[x]$ refers to the largest integer smaller than or equal to x . We also let $\sum_{k=1}^0 b_k = 0$ for ease of notation.

To derive the asymptotic variance of $\hat{\sigma}^2(2, m)$ with the quadratic form (S4), we apply the following formula:

$$\begin{aligned} \text{var}(\hat{\sigma}^2(2, m)) &= \frac{1}{\text{tr}(D)^2} [4\sigma^2 g^T D^2 g + 4g^T (D \text{diag}(D) u) \sigma^3 \gamma_3 \\ &\quad + \sigma^4 \text{tr}\{\text{diag}(D)^2\} (\gamma_4 - 3) + 2\sigma^4 \text{tr}(D^2)], \end{aligned} \quad (\text{S5})$$

where $g = (g(x_1), \dots, g(x_n))^T$, $u = (1, \dots, 1)^T$, $\text{diag}(D)$ denotes the diagonal matrix of D , and $\gamma_i = E(\varepsilon^i / \sigma^i)$ for $i = 3$ and 4 . Consequently, we show that

Theorem S4. *Consider the equally spaced design with $m \rightarrow \infty$ and $m/n \rightarrow 0$, and assume that $g(x)$ has a bounded second derivative. Then for any difference sequence (d_0, d_1, d_2) under the constraint (S1), we have*

$$\text{var}(\hat{\sigma}^2(2, m)) = \frac{1}{n} \text{var}(\varepsilon^2) + \frac{A_1}{mn} \sigma^4 + \frac{A_2 m}{n^2} \text{var}(\varepsilon^2) + o\left(\frac{1}{nm}\right) + o\left(\frac{m}{n^2}\right),$$

where $A_1 = 9/4 + 9d_1^2(d_1^2 - 1/2)$ and $A_2 = 9/56 + 165d_1^2(1 - d_1^2)/448$.

By Theorems S2 and S4, the optimal bandwidth is $m_{\text{opt}} = \sqrt{A_1 \sigma^4 / (A_2 \text{var}(\varepsilon^2))} n^{1/2}$, and the optimal MSE is $\text{MSE}(\hat{\sigma}^2(2, m_{\text{opt}})) = n^{-1} \text{var}(\varepsilon^2) + 2n^{-3/2} \sqrt{A_1 A_2 \sigma^4 \text{var}(\varepsilon^2)} +$

$o(n^{-3/2}) + O(m^6/n^6)$. In addition, we have $\text{MSE}(\hat{\sigma}^2(2, m)) = n^{-1} [\text{var}(\varepsilon^2) + o(1)]$ for any $m = n^\tau$ with $0 < \tau < 5/6$. This shows that the unified estimator can achieve the asymptotically optimal rate of MSE and hence is a consistent estimator of σ^2 , no matter which difference sequence is used. From this point of view, the unified estimator has improved the classical difference-based estimators in Hall et al. (1990). Note also that, by a similar proof as that for Theorem S4, we can verify that $\text{var}(\hat{\sigma}^2(r, m)) = n^{-1} [\text{var}(\varepsilon^2) + o(1)]$ for any difference sequence of order $r \geq 3$ with $m \rightarrow \infty$ and $m/n \rightarrow 0$.

S3 Bias comparison

In this section, we conduct a numerical comparison on the bias terms of the “state-of-the-art” estimator and the proposed unified estimators, as a supplement to Section 3 in the main manuscript. Specifically, we consider the TW estimator in Tong and Wang (2005), the unified ordinary estimator, and the unified optimal estimator. For simplicity of exposition, we fix $r = 2$ (the comparison results for $r \geq 3$ remain similar) and assume the errors are normally distributed. According to Theorems S1 and S2, we have

$$\begin{aligned} \text{Bias}(\hat{\sigma}_{\text{TW}}^2) &\simeq -\frac{1}{32} \left[\int_0^1 [g'(t)]^2 dt - \frac{1}{2}[g'(0)]^2 - \frac{1}{2}[g'(1)]^2 \right] \frac{m^3}{n^3}, \\ \text{Bias}(\hat{\sigma}_{\text{opt}}^2(2, m)) &\simeq -\frac{5}{32} \left[\int_0^1 [g'(t)]^2 dt - \frac{10 - \sqrt{5}}{20}[g'(0)]^2 - \frac{10 + \sqrt{5}}{20}[g'(1)]^2 \right] \frac{m^3}{n^3}, \\ \text{Bias}(\hat{\sigma}_{\text{ord}}^2(2, m)) &\simeq -\left[\frac{1}{70} \int_0^1 [g''(x)]^2 dx \right] \frac{m^4}{n^4}. \end{aligned}$$

As mentioned, the asymptotic bias of the unified ordinary estimator has a lower order than those of the TW estimator and the unified optimal estimator.

To numerically compare the coefficients in the bias terms, we consider the regression function $g(x) = A \sin(w\pi x)$, where A represents the amplitude and w represents the frequency. Note that the sine function has been frequently used in the literature for conducting simulation studies, see for example, in Seifert, Gasser and Wolf (1993), Dette et al. (1998), Tong and Wang (2005), and Tong, Ma and Wang (2013). Given the sine function, we have $\text{Bias}(\hat{\sigma}_{\text{TW}}^2) \simeq B_1 m^3/n^3$, $\text{Bias}(\hat{\sigma}_{\text{opt}}^2(2, m)) \simeq B_2 m^3/n^3$, and

Bias ($\hat{\sigma}_{\text{ord}}^2(2, m)$) $\simeq B_3 m^4/n^4$, where the coefficients of the leading terms are

$$\begin{aligned} B_1 &= -\frac{1}{64}w^2\pi^2 A^2 \left[\frac{\sin(w\pi)\cos(w\pi)}{w\pi} - \frac{1}{2} - \frac{1}{2}\cos(2w\pi) \right], \\ B_2 &= -\frac{5}{64}w^2\pi^2 A^2 \left[\frac{\sin(w\pi)\cos(w\pi)}{w\pi} - \frac{10-\sqrt{5}}{20} - \frac{10+\sqrt{5}}{20}\cos(2w\pi) \right], \\ B_3 &= -\frac{1}{140}w^4\pi^4 A^2 \left[1 - \frac{\sin(w\pi)\cos(w\pi)}{w\pi} \right]. \end{aligned}$$

First of all, as the three coefficients are all proportional to A^2 , we conclude that the amplitude A takes no effect on the comparison of the estimators. The effect of w on the corresponding bias terms is investigated in the next paragraph.

To compare the bias terms of the TW estimator and the unified optimal estimator, it suffices to compare the coefficients B_1 and B_2 . With w ranging from 0 to 4.1, we plot in Figure S1 the quantities of $|B_1|$ and $|B_2|$ for $g(x) = 5\sin(w\pi x)$. It is evident that both $|B_1|$ and $|B_2|$ have a periodic pattern along with w , and $|B_1| < |B_2|$ for most w values. In particular, we have $|B_1| = 3.86$ and $|B_2| = 19.28$ for $w = 1$, $|B_1| = 15.42$ and $|B_2| = 77.11$ for $w = 2$, and $|B_1| = 61.69$ and $|B_2| = 308.43$ for $w = 4$. This shows that the bias term of the unified optimal estimator is even more severe than that of the TW estimator. In view of this, we conclude that a combination between the linear regression method and the optimal difference sequence can be less satisfactory. For the unified ordinary estimator, instead, it is always asymptotically better than the TW estimator as it is a higher order bias term, regardless of the comparison of coefficients.

S4 Unequally spaced design

For a general domain including the unequally spaced design, we provide an alternative procedure for the unified framework of estimating the residual variance. Assume that $\{x_1, \dots, x_n\}$ are ordered design points with $x_1 < \dots < x_n$, and let $\mathcal{A}_r = \{(i, k) : x_{i+kr} - x_i \leq M, i + kr \leq n\}$ for $M = o(1)$. We define $p_{ik} = (\sum_{j=0}^r a_{ikj} Y_{i+kj})^2$, where $\{a_{ikj}\}$ is a difference sequence with order r . Then, $E(p_{ik}) = \sigma^2 + q_{ik}^2 [g'(x_i)]^2 + o(q_{ik}^2)$ where $q_{ik} = \sum_{j=1}^r a_{ikj} (x_{i+kj} - x_i)$. With this approximation, we can construct the linear regression

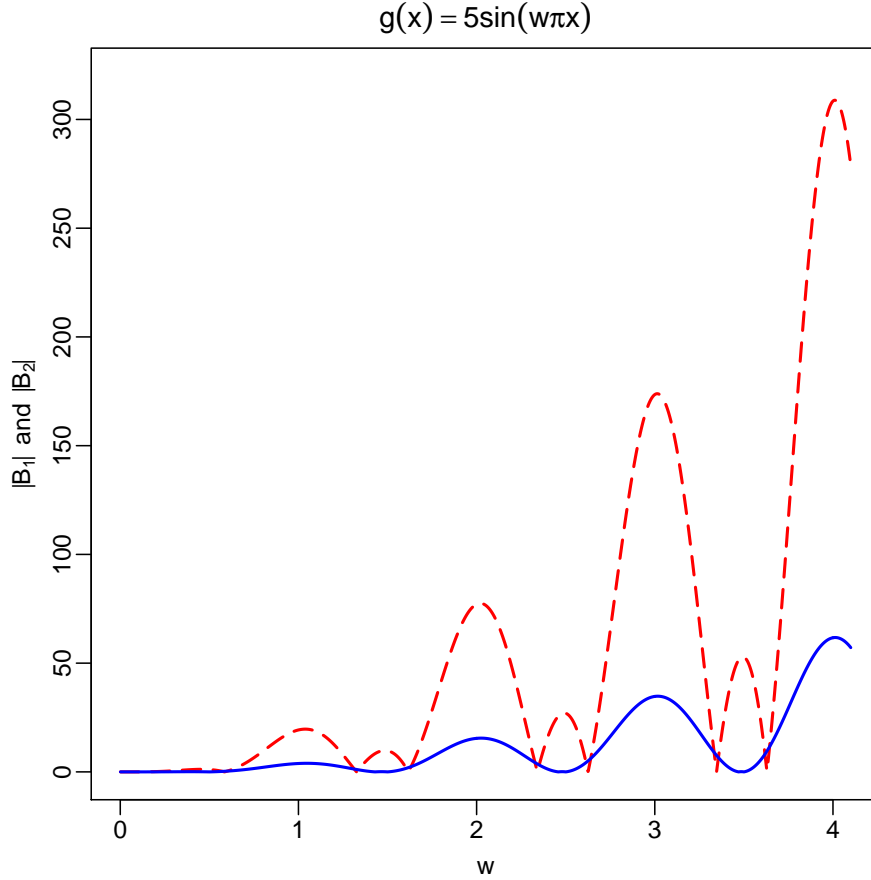


Figure S1: The absolute value of coefficients, $|B_1|$ and $|B_2|$, along with w for $g(x) = 5\sin(w\pi x)$. The blue solid line represents $|B_1|$, and the red dashed line represents $|B_2|$.

model as

$$p_{ik} = \sigma^2 + q_{ik}^2 [g'(x_i)]^2 + \xi_{ik}, \quad (i, k) \in \mathcal{A}_r.$$

Consequently, by regressing p_{ik} on q_{ik}^2 using least squares, we have the final estimator as $\hat{\sigma}^2(r, M)$. It can be verified that, with $M = mr/n$, $\hat{\sigma}^2(r, M)$ is equivalent to $\hat{\sigma}^2(r, m)$ for the equidistant design. Further, if $\{a_{ikj}\}$ satisfies $\sum_{j=0}^r a_{ikj}^2 = 1$, $\sum_{j=0}^r a_{ikj} = 0$ and $\sum_{j=1}^r a_{ikj}(x_{i+kj} - x_i)^K = 0$ for $K = 1, \dots, r-1$, we can show that $\hat{\sigma}_{\text{ord}}^2(r, M)$ is also equivalent to $\hat{\sigma}_{\text{ord}}^2(r, m)$. Further research might be needed to investigate the asymptotic properties under the unequally spaced design.

S5 Proofs of theoretical results

In this section, we provide the technical proofs for the four theorems in Sections S1 and S2, respectively.

Proof of Theorem S1

First, we have $\sum_{k=1}^m b_k w_k = 1$ and $\sum_{k=1}^m b_k w_k h_k = 0$. (i) Let $g(x) = ax + b$, then

$$\begin{aligned} E(s_k(r)) &= \sigma^2 + \frac{1}{n - kr} \sum_{i=1}^{n-kr} \left(\sum_{j=0}^r d_j \left(a \frac{(i + jk)}{n} + b \right) \right)^2 \\ &= \sigma^2 + \frac{1}{n - kr} \sum_{i=1}^{n-kr} \left(\frac{ak}{n} \sum_{j=0}^r d_j j \right)^2 \\ &= \sigma^2 + a^2 \left(\sum_{j=0}^r d_j j \right)^2 h_k. \end{aligned}$$

This leads to $E(\hat{\sigma}^2(r, m)) = \sigma^2 + a^2 (\sum_{j=0}^r d_j j)^2 \sum_{k=1}^m b_k w_k h_k = \sigma^2$.

(ii) For the polynomial $g(x)$ with order up to $p \leq r - 1$, let $g_i^{(p)}$ denote the p th order derivative at design point x_i . Then

$$\begin{aligned} E(s_k(r)) &= \sigma^2 + \frac{1}{n - kr} \sum_{i=1}^{n-kr} \left(\sum_{j=0}^r d_j g_{i+jk} \right)^2 \\ &= \sigma^2 + \frac{1}{n - kr} \sum_{i=1}^{n-kr} \left(d_0 g_i + d_1 \sum_{j=0}^p g_i^{(j)} \frac{(k/n)^j}{j!} + \dots + d_r \sum_{j=0}^p g_i^{(j)} \frac{(jk/n)^j}{j!} \right)^2 \\ &= \sigma^2 + \frac{1}{n - kr} \sum_{i=1}^{n-kr} \left(C_0 g_i + C_1 g_i' \frac{k}{n} + \dots + C_p g_i^{(p)} \frac{k^p}{n^p} \right)^2, \end{aligned}$$

where $C_0 = \sum_{j=0}^r d_j$ and $C_i = \sum_{j=0}^r j^i d_j / i!$ for $i = 1, \dots, r$. When (d_0, \dots, d_r) is the order- r ordinary difference sequence, we have $C_i = 0$ for $0 \leq i \leq p \leq r - 1$. This results in $E(s_k(r)) = \sigma^2$ and, consequently, $E(\hat{\sigma}_{\text{ord}}^2(r, m)) = \sigma^2$. \square

Proof of Theorem S2

Note that

$$E(s_k(r)) = \sigma^2 + \frac{1}{n - kr} \sum_{i=1}^{n-kr} \left(\frac{1}{n} C_1 g_i' k + O\left(\frac{k^2}{n^2}\right) \right)^2 = \sigma^2 + C_1^2 h_k \int_0^1 [g'(x)]^2 dx + O\left(\frac{k^3}{n^3}\right).$$

We have

$$\begin{aligned}
E(\hat{\sigma}^2(r, m)) &= \sum_{k=1}^m b_k w_k \left[\sigma^2 + C_1^2 h_k \int_0^1 [g'(x)]^2 dx + O\left(\frac{k^3}{n^3}\right) \right] \\
&= \sigma^2 + C_1^2 \int_0^1 [g'(x)]^2 dx \sum_{k=1}^m b_k w_k h_k + O\left(\frac{m^3}{n^3}\right) \\
&= \sigma^2 + O\left(\frac{m^3}{n^3}\right).
\end{aligned}$$

For $s_k(r)$ with any difference sequence, with $k = o(n)$ and $k \rightarrow \infty$, we have

$$\begin{aligned}
E[s_k(r)] &= \sigma^2 + \frac{1}{n - rk} \sum_{i=1}^{n-rk} \left[C_1 g'(x_i) \frac{k}{n} + C_2 g''(x_i) \frac{k^2}{n^2} + o\left(\frac{k^2}{n^2}\right) \right]^2 \\
&= \sigma^2 + C_1^2 P_A \frac{k^2}{n^2} + 2C_1 C_2 P_B \frac{k^3}{n^3} + o\left(\frac{k^3}{n^3}\right), \tag{S6}
\end{aligned}$$

where $P_A = \sum_{i=1}^{n-rk} [g'(x_i)]^2 / (n - rk)$ and $P_B = \sum_{i=1}^{n-rk} g'(x_i) g''(x_i) / (n - rk)$. For ease of notation, let $f(t) = [g'(t)]^2$. Then,

$$\begin{aligned}
P_A &= \frac{1}{n - rk} \sum_{i=1}^n f(x_i) - \frac{1}{n - rk} \sum_{i=n-rk+1}^n f(x_i) \\
&= \frac{n}{n - rk} \int_0^1 f(t) dt - \frac{n}{n - rk} \sum_{i=1}^n \int_{x_{i-1}}^{x_i} (f(t) - f(x_i)) dt \\
&\quad - \frac{n}{n - rk} \int_{x_{n-rk}}^1 f(t) dt + \frac{n}{n - rk} \sum_{i=n-rk+1}^n \int_{x_{i-1}}^{x_i} (f(t) - f(x_i)) dt \\
&\equiv P_1 - P_2 - P_3 + P_4.
\end{aligned}$$

For P_1 , we have $P_1 = \int_0^1 f(t) dt + (rk/n) \int_0^1 f(t) dt + o(k/n)$. For P_2 , we have

$$\begin{aligned}
P_2 &= \frac{n}{n - rk} \sum_{i=1}^n \int_{x_{i-1}}^{x_i} \left[f'(x_i)(t - x_i) + O\left(\frac{1}{n^2}\right) \right] dt \\
&= \frac{n}{n - rk} \sum_{i=1}^n \left[-f'(x_i) \frac{1}{2n^2} + O\left(\frac{1}{n^3}\right) \right] dt \\
&= O\left(\frac{1}{n}\right).
\end{aligned}$$

For P_3 , we let $F(s) = \int_s^1 f(t) dt$. When $0 < s < 1$ is close to 1, by Taylor expansion $F(s) = F(1) + F'(1)(s - 1) + O((1 - s)^2)$ we have

$$P_3 = \frac{n}{n - rk} F(x_{n-rk}) = \frac{rk}{n} f(1) + O\left(\frac{k^2}{n^2}\right).$$

Finally, noting that $P_4 = o(1/n)$, we have

$$P_A = \int_0^1 f(t)dt + \frac{rk}{n} \left[\int_0^1 f(t)dt - f(1) \right] + o\left(\frac{k}{n}\right). \quad (\text{S7})$$

Similarly as those for P_A , we can derive that

$$P_B = \int_0^1 g'(t)g''(t)dt + o(1) = ([g'(1)]^2 - [g'(0)]^2) / 2 + o(1). \quad (\text{S8})$$

Substituting (S7) and (S8) into (S6), we have

$$E(s_k(r)) = \sigma^2 + \frac{k^2}{n^2}T_1 + \frac{k^3}{n^3}T_2 + o\left(\frac{k^3}{n^3}\right),$$

where $T_1 = C_1^2 \int_0^1 [g'(t)]^2 dt / 2$ and $T_2 = C_1[rC_1 \int_0^1 [g'(t)]^2 dt - (rC_1 - C_2)[g'(1)]^2 - C_2[g'(0)]^2]$.

For the linear regression estimator $s_k(r)$, $N = nm - rm(m+1)/2$ and $w_k = (n - rk)/N$, we have

$$\hat{\sigma}^2(r, m) = \sum_{k=1}^m w_k s_k(r) - \frac{\bar{h}_w}{\sum_{k=1}^m w_k h_k^2 - \bar{h}_w^2} \sum_{k=1}^m w_k (h_k - \bar{h}_w) s_k(r) = Q_1 - Q_2.$$

Let $I_t = \sum_{k=1}^m k^t$ for $t = 1, \dots, 5$. For the first term, we have

$$E(Q_1) = \sigma^2 + \left(\frac{I_2}{Nn} - \frac{rI_3}{Nn^2} \right) T_1 + \left(\frac{I_3}{Nn^2} - \frac{rI_4}{Nn^3} \right) T_2 + o\left(\frac{m^3}{n^3}\right). \quad (\text{S9})$$

For the second term, we have

$$\begin{aligned} \bar{h}_w &= \frac{I_2}{Nn} - \frac{rI_3}{Nn^2} = \frac{m^2}{3n^2} + o\left(\frac{m^2}{n^2}\right), \\ \sum_{k=1}^m w_k h_k^2 - \bar{h}_w^2 &= \frac{I_4}{Nn^3} - \frac{rI_5}{Nn^4} - \left(\frac{I_2}{Nn} - \frac{rI_3}{Nn^2} \right)^2 = \frac{4m^4}{45n^4} + o\left(\frac{m^4}{n^4}\right), \\ \sum_{k=1}^m w_k (h_k - \bar{h}_w) E[s_k(r)] &= \left\{ \frac{I_4}{Nn^3} - \frac{rI_5}{Nn^4} - \left(\frac{I_2}{Nn} - \frac{rI_3}{Nn^2} \right)^2 \right\} T_1 \\ &\quad + \left\{ \frac{I_5}{Nn^4} - \frac{rI_6}{Nn^5} - \left(\frac{I_2}{Nn} - \frac{rI_3}{Nn^2} \right) \left(\frac{I_3}{Nn^2} - \frac{rI_4}{Nn^3} \right) \right\} T_2 + o\left(\frac{m^5}{n^5}\right). \end{aligned}$$

This leads to the expectation of the second term as

$$\begin{aligned} E(Q_2) &= \frac{\frac{I_2}{Nn} - \frac{rI_3}{Nn^2}}{\frac{I_4}{Nn^3} - \frac{rI_5}{Nn^4} - \left(\frac{I_2}{Nn} - \frac{rI_3}{Nn^2} \right)^2} \left[\left\{ \frac{I_4}{Nn^3} - \frac{rI_5}{Nn^4} - \left(\frac{I_2}{Nn} - \frac{rI_3}{Nn^2} \right)^2 \right\} T_1 \right. \\ &\quad \left. + \left\{ \frac{I_5}{Nn^4} - \frac{rI_6}{Nn^5} - \left(\frac{I_2}{Nn} - \frac{rI_3}{Nn^2} \right) \left(\frac{I_3}{Nn^2} - \frac{rI_4}{Nn^3} \right) \right\} T_2 + o\left(\frac{m^5}{n^5}\right) \right] \\ &= \left(\frac{I_2}{Nn} - \frac{rI_3}{Nn^2} \right) T_1 + CT_2 + o\left(\frac{m^3}{n^3}\right), \end{aligned} \quad (\text{S10})$$

where

$$\begin{aligned}
C &= \frac{\frac{I_5}{Nn^4} - \frac{rI_6}{Nn^5} - \left(\frac{I_2}{Nn} - \frac{rI_3}{Nn^2}\right) \left(\frac{I_3}{Nn^2} - \frac{rI_4}{Nn^3}\right)}{\frac{I_4}{Nn^3} - \frac{rI_5}{Nn^4} - \left(\frac{I_2}{Nn} - \frac{rI_3}{Nn^2}\right)^2} \left(\frac{I_2}{Nn} - \frac{rI_3}{Nn^2}\right) \\
&= \frac{\frac{m^5}{6n^5} + o\left(\frac{m^5}{n^5}\right) - \left(\frac{m^2}{3n^2} + o\left(\frac{m^2}{n^2}\right)\right) \left(\frac{m^3}{4n^3} + o\left(\frac{m^3}{n^3}\right)\right)}{\frac{4m^4}{45n^4} + o\left(\frac{m^4}{n^4}\right)} \left(\frac{m^2}{3n^2} + o\left(\frac{m^2}{n^2}\right)\right) \\
&= \frac{5m^3}{16n^3} + o\left(\frac{m^3}{n^3}\right).
\end{aligned}$$

Finally, by (S9) and (S10), we have $\text{Bias}(\hat{\sigma}^2(r, m)) = -(m^3/16n^3)T_2 + o(m^3/n^3)$ and this completes the derivation. \square

Proof of Theorem S3

For the ordinary difference sequence, we have $C_1 = C_2 = \dots = C_{r-1} = 0$ and $C_r = \binom{2r}{r}^{-1/2}$. Then,

$$\begin{aligned}
E(s_k(r)) &= \sigma^2 + \frac{1}{n - kr} \sum_{i=1}^{n-kr} \left(C_1 g_i' k/n + \dots + C_r g_i^{(r)} k^r/n^r + o(k^r/n^r) \right)^2 \\
&= \sigma^2 + C_r^2 \frac{k^{2r}}{n^{2r}} \int_0^1 [g^{(r)}(x)]^2 dx + o\left(\frac{k^{2r}}{n^{2r}}\right).
\end{aligned}$$

This leads to

$$\begin{aligned}
E(\hat{\sigma}^2(r, m)) &= \sum_{k=1}^m b_k w_k \left[\sigma^2 + C_r^2 \frac{k^{2r}}{n^{2r}} \int_0^1 [g^{(r)}(x)]^2 dx + o\left(\frac{k^{2r}}{n^{2r}}\right) \right] \\
&= \sigma^2 + C_r^2 \int_0^1 [g^{(r)}(x)]^2 dx \sum_{k=1}^m b_k w_k \frac{k^{2r}}{n^{2r}} + o\left(\frac{m^{2r}}{n^{2r}}\right) \\
&= \sigma^2 + \frac{3C_r^2(1-r)m^{2r}}{(2r+1)(2r+3)n^{2r}} \int_0^1 [g^{(r)}(x)]^2 dx + o\left(\frac{m^{2r}}{n^{2r}}\right),
\end{aligned}$$

where

$$\begin{aligned}
\sum_{k=1}^m b_k w_k \frac{k^{2r}}{n^{2r}} &= \sum_{k=1}^m w_k \frac{k^{2r}}{n^{2r}} - \frac{\bar{h}_w}{\sum_{k=1}^m w_k h_k^2 - \bar{h}_w^2} \sum_{k=1}^m (h_k - \bar{h}_w) \frac{k^{2r}}{n^{2r}} \\
&= \frac{1}{(2r+1)} \frac{m^{2r}}{n^{2r}} (1 + o(1)) - \frac{15}{4} \left[\frac{1}{(2r+3)} - \frac{1}{3(2r+1)} \right] \frac{m^{2r}}{n^{2r}} (1 + o(1)) \\
&= \frac{3(1-r)}{(2r+1)(2r+3)} \frac{m^{2r}}{n^{2r}} (1 + o(1)).
\end{aligned}$$

This completes the derivation. \square

Proof of Theorem S4

We first introduce three lemmas. Lemma S1 is used in the proof of Lemma S2, and Lemma S2 is used in the proof of Lemma S3. To save space, we only sketch the proof of Lemma S3.

Lemma S1. *Assume that $m \rightarrow \infty$ and $m/n \rightarrow 0$. Then for $r = 2$, we have*

- (a) $\sum_{k=1}^m h_k = \frac{m^3}{3n^2} + \frac{m^2}{2n^2} + o(\frac{m^2}{n^2})$.
- (b) $\bar{h}_w = \frac{1}{Nn^2}[\frac{1}{3}nm^3 + \frac{1}{2}nm^2 - \frac{1}{2}m^4 + o(nm^2) + o(m^4)]$.
- (c) $\sum_{k=1}^m w_k h_k^2 = \frac{1}{Nn^4}[\frac{1}{5}nm^5 + \frac{1}{2}nm^4 - \frac{1}{3}m^6 + o(nm^4) + o(m^6)]$.
- (d) $\sum_{k=1}^m h_k^2 = \frac{m^5}{5n^4} + \frac{m^4}{2n^4} + o(\frac{m^4}{n^4})$.
- (e) $\sum_{k=1}^m w_k h_k^2 - \bar{h}_w^2 = \frac{1}{N^2n^4}[\frac{4}{45}n^2m^6 + \frac{1}{6}n^2m^5 - \frac{1}{5}nm^7 + o(nm^7) + o(n^2m^5)]$.

Lemma S2. *Assume that $m \rightarrow \infty$ and $m/n \rightarrow 0$. Then for $r = 2$, we have*

- (a) $\sum_{k=1}^m b_k = m - \frac{5m^2}{8n} + o(\frac{m^2}{n})$.
- (b) $\sum_{k=1}^{l-1} b_k = \frac{9}{4}l - \frac{5l^3}{4m^2} + o(l) + O(1), \quad 1 \leq l \leq m$.
- (c) $\sum_{k=1}^{[(l-1)/2]} b_k = \frac{9}{8}l - \frac{5l^3}{32m^2} + o(l) + O(1), \quad 1 \leq l \leq 2m$.
- (d) $\sum_{k=1}^m b_k^2 = \frac{9}{4}m + o(m)$.
- (e) $\sum_{k=1}^{m/2} b_k b_{2k} = \frac{9}{8}m + o(m)$.
- (f) $\sum_{k=1}^m kb_k = O(m^2), \quad 1 \leq l \leq m$.
- (g) $\sum_{k=1}^{l-1} kb_k = O(l^2), \quad 1 \leq l \leq m$.
- (h) $\sum_{k=1}^{[(l-1)/2]} kb_k = O(l^2), \quad 1 \leq l \leq 2m$.
- (i) $\sum_{k=1}^m k^2 b_k = o(m^3)$.

Lemma S3. *Under the same conditions as in Theorem S4, for $r = 2$ we have*

- (a) $g^T D^2 g = O(m^5/n^2)$.

$$(b) \quad g^T(D\text{diag}(D)u) = O(m^4/n).$$

$$(c) \quad \text{tr}(D^2) = nm^2 - \left[\frac{103}{56} - \frac{165}{448}d_1^2(1 - d_1^2)\right]m^3 + \left[\frac{9}{8} + \frac{9}{2}(d_1^2 - \frac{1}{2})d_1^2\right]mn + o(m^3 + nm).$$

$$(d) \quad \text{tr}\{\text{diag}(D)^2\} = nm^2 - \left[\frac{103}{56} - \frac{165}{448}d_1^2(1 - d_1^2)\right]m^3 + o(m^3).$$

Proof of Lemma S3. (a) Let $g_i = g(x_i)$, $g'_i = g'(x_i)$ and $g''_i = g''(x_i)$ for $i = 1, \dots, n$. Noting that D is symmetric, we have $g^T D^2 g = g^T D^T D g = (Dg)^T Dg \triangleq p^T p$, where $p = Dg = (p_1, p_2, \dots, p_n)^T$. For $i \in [2m + 1, n - 2m]$, by Lemma S2 we have

$$\begin{aligned} p_i &= (d_0 d_1 + d_1 d_2) \sum_{k=1}^m b_k (g_{i-k} + g_{i+k} - 2g_i) + d_0 d_2 \sum_{k=1}^m b_k (g_{i-2k} + g_{i+2k} - 2g_i) \\ &= (d_0 d_1 + d_1 d_2) \sum_{k=1}^m b_k \left(\frac{k^2}{n^2} g''_i + o\left(\frac{k^2}{n^2}\right) \right) + 4d_0 d_2 \sum_{k=1}^m b_k \left(\frac{k^2}{n^2} g''_i + o\left(\frac{k^2}{n^2}\right) \right) \\ &= o\left(\frac{m^3}{n^2}\right). \end{aligned}$$

Similarly, by Lemma S2 we can derive that Similarly, we can derive that $p_i = O(m^2/n)$ for $i \in [1, m]$, $i \in [m + 1, 2m]$, $i \in [n - 2m + 1, n - m]$, and $i \in [n - m + 1, n]$, respectively. This leads to $g^T D^2 g = \sum_{i=1}^{2m} p_i^2 + \sum_{i=2m+1}^{n-2m} p_i^2 + \sum_{i=n-2m+1}^n p_i^2 = O(m^5/n^2)$.

(b) By (a), (b) and (c) in Lemma S2, we have $g^T(D\text{diag}(D)u) = p^T \text{diag}(D)u = (\sum_{i=1}^m + \sum_{i=m+1}^{2m} + \sum_{i=1+2m}^{n-2m} + \sum_{i=n-2m+1}^{n-m} + \sum_{i=n-m+1}^n) p_i d_{ii} = O(m^4/n)$.

(c) By some simple calculation, we have

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^n d_{ij}^2 &= (d_0^4 + \frac{11}{14}d_1^4 + \frac{2545}{7168}d_2^4 + \frac{13}{8}d_0^2 d_1^2 + \frac{233}{224}d_1^2 d_2^2 + \frac{67}{64}d_0^2 d_2^2)m^3 + o(m^3), \\ \sum_{i=m+1}^{2m} \sum_{j=1}^n d_{ij}^2 &= (d_0^4 + d_1^4 + \frac{8719}{7168}d_2^4 + 2d_0^2 d_1^2 + \frac{141}{64}d_1^2 d_2^2 + \frac{141}{64}d_0^2 d_2^2)m^3 + o(m^3), \\ \sum_{i=2m+1}^{n-2m} \sum_{j=1}^n d_{ij}^2 &= nm^2 - \frac{21}{4}m^3 + \left(\frac{9}{8} + \frac{9}{2}d_0 d_1^2 d_2\right)nm + o(nm) + o(m^3), \\ \sum_{i=n-2m+1}^{n-m} \sum_{j=1}^n d_{ij}^2 &= \left(\frac{8719}{7168}d_0^4 + d_1^4 + d_2^4 + \frac{141}{64}d_0^2 d_1^2 + 2d_1^2 d_2^2 + \frac{141}{64}d_0^2 d_2^2\right)m^3 + o(m^3), \\ \sum_{i=n-m+1}^n \sum_{j=1}^n d_{ij}^2 &= \left(\frac{2545}{7168}d_0^4 + \frac{11}{14}d_1^4 + d_2^4 + \frac{233}{224}d_0^2 d_1^2 + \frac{13}{8}d_1^2 d_2^2 + \frac{67}{64}d_0^2 d_2^2\right)m^3 + o(m^3). \end{aligned}$$

This leads to

$$\begin{aligned}\mathrm{tr}(D^2) &= \left(\sum_{i=1}^m \sum_{j=1}^n + \sum_{i=m+1}^{2m} \sum_{j=1}^n + \sum_{i=2m+1}^{n-2m} \sum_{j=1}^n + \sum_{i=n-2m+1}^{n-m} \sum_{j=1}^n + \sum_{i=n-m+1}^n \sum_{j=1}^n \right) d_{ij}^2 \\ &= nm^2 - \left[\frac{103}{56} - \frac{165}{448} d_1^2 (1 - d_1^2) \right] m^3 + \left[\frac{9}{8} + \frac{9}{2} (d_1^2 - \frac{1}{2}) d_1^2 \right] mn + o(m^3 + nm).\end{aligned}$$

(d) Similarly as in the proof of part (c), we have

$$\begin{aligned}\mathrm{tr}\{\mathrm{diag}(D)^2\} &= \left(\sum_{i=1}^m + \sum_{i=m+1}^{2m} + \sum_{i=1+2m}^{n-2m} + \sum_{i=n-2m+1}^{n-m} + \sum_{i=n-m+1}^n \right) d_{ii}^2 \\ &= nm^2 - \left[\frac{103}{56} - \frac{165}{448} d_1^2 (1 - d_1^2) \right] m^3 + o(m^3).\end{aligned}$$

Proof of Theorem S4. By Lemma S3, formula (S5) and the fact that $\mathrm{tr}(D) = N$, we have the variance of $\hat{\sigma}^2(2, m)$ as

$$\begin{aligned}\mathrm{var}(\hat{\sigma}^2(2, m)) &= \frac{1}{N^2} \left\{ O\left(\frac{m^5}{n^2}\right) + O\left(\frac{m^4}{n}\right) \right. \\ &\quad + (\mathrm{var}(\varepsilon^2) - 2\sigma^4) \left[nm^2 - \left[\frac{103}{56} - \frac{165}{448} d_1^2 (1 - d_1^2) \right] m^3 + o(m^3) \right] \\ &\quad + 2\sigma^4 \left[nm^2 - \left[\frac{103}{56} - \frac{165}{448} d_1^2 (1 - d_1^2) \right] m^3 \right. \\ &\quad \left. \left. + \left[\frac{9}{8} + \frac{9}{2} (d_1^2 - \frac{1}{2}) d_1^2 \right] mn + o(m^3 + nm) \right] \right\} \\ &= \frac{1}{n} \mathrm{var}(\varepsilon^2) + A_1 \frac{\sigma^4}{mn} + A_2 \frac{m}{n^2} \mathrm{var}(\varepsilon^2) + o\left(\frac{1}{nm}\right) + o\left(\frac{m}{n^2}\right),\end{aligned}$$

where A_1 and A_2 are defined in Theorem S4. This proves the theorem. \square

References

- Detle, H., Munk, A. and Wagner, T. (1998). Estimating the variance in nonparametric regression - what is a reasonable choice?, *Journal of the Royal Statistical Society, Series B* **60**: 751–764.
- Hall, P., Kay, J. W. and Titterton, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression, *Biometrika* **77**: 521–528.
- Seifert, B., Gasser, T. and Wolf, A. (1993). Nonparametric estimation of residual variance revisited, *Biometrika* **80**: 373–383.

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.