

# An estimator of the mean of the squared functions for a nonparametric regression

Chun Gun Park<sup>1</sup>

Department of Mathematics, Kyonggi University

Received 24 January 2009, revised 12 May 2009, accepted 21 May 2009

## Abstract

So far in a nonparametric regression model one of the interesting problems is estimating the error variance. In this paper we propose an estimator of the mean of the squared functions which is the numerator of SNR (Signal to Noise Ratio). To estimate SNR, the mean of the squared function should be firstly estimated. Our focus is on estimating the amplitude, that is the mean of the squared functions, in a nonparametric regression using a simple linear regression model with the quadratic form of observations as the dependent variable and the function of a lag as the regressor. Our method can be extended to nonparametric regression models with multivariate functions on unequally spaced design points or clustered designed points.

*Keywords:* Difference-based estimator, error variance, mean of the squared functions, nonparametric regression, quadratic form, SNR.

## 1. Introduction

Consider a nonparametric regression of the form

$$Y_i = f(x_i) + W_i = f_i + W_i, \quad i = 1, \dots, n, \quad (1.1)$$

where  $f$  is an unknown mean function and the error  $W_i$ 's are independent and identically distributed random variables with zero mean and variance  $\sigma^2$ . We assume that the design points  $x_i$ 's lie in  $[0, 1]$  and have been ordered.

Usually the focus of nonparametric regression models is on estimating the mean function or the error variance. The estimator of the error variance is obtained from fitting the mean function  $f$  (Park, 2004, 2008; Wahba, 1990; Hall & Carroll, 1989; Carter & Eagleson 1992; Neumann, 1994). The other approach to estimate the error variance is using difference-based variance estimation which does not require an estimator of the mean function and is to remove trend in the mean function, an idea originating in the time series analysis (Rice, 1984).

---

<sup>1</sup> Part time lecturer, Department of Mathematics, Kyonggi University, Iui-dong Yeongtog-gu, Suwon-si Gyeonggi-do 443-760, Korea. E-mail: chungunpark@daum.net

In nonparametric regression SNR has been affecting the performance of the estimators of the mean function. Our purpose is to suggest an estimator of the mean of the squared functions which is the numerator of SNR, using the  $k$ -order difference-based estimators of the error variance (see Tong and Wang, 2005).

In Sections 2 and 3 we consider equally spaced designs in  $[0, 1]$ . We present our estimator in Section 2 and asymptotic results in Section 3. A small simulation study examining the finite sample behavior of our method is given in Section 4. Section 5 provides a brief discussion. Some proofs of the technical results are deferred to Appendix.

## 2. The definition of the mean of the squared functions

Throughout this paper we assume for simplicity that all  $x_i$ 's are equal spaced. The SNR from (1.1) is given by

$$SNR = \frac{\int_0^1 f^2(x) dx}{\sigma^2}.$$

The mean of the squared functions is denoted by

$$Q = \sigma^2 SNR = \frac{1}{n} \sum_{i=1}^n f_i^2. \quad (2.1)$$

Our concern is on presenting an estimator of the mean of the squared functions  $Q$  from (2.1) and exploring the statistical properties for the estimator of  $Q$ .

### 2.1. The $k$ -order difference estimator of the error variance

Rice (1984) proposed the first-order difference-based estimator

$$\hat{\sigma}_R^2 = \frac{1}{2(n-1)} \sum_{i=2}^n (Y_i - Y_{i-1})^2.$$

Rice's estimator uses differences of all consecutive observations. We define a lag- $k$  Rice estimator  $\hat{\sigma}_R^2(k)$  as

$$\hat{\sigma}_R^2(k) = \frac{1}{2(n-k)} \sum_{i=k+1}^n (Y_i - Y_{i-1})^2, \quad k = 1, \dots, n-1. \quad (2.2)$$

From (2.2), taking expectation of the Rice estimator, we have

$$E(\hat{\sigma}_R^2(k)) = \sigma^2 + \frac{1}{2(n-k)} \sum_{i=k+1}^n (f_i - f_{i-k})^2, \quad (2.3)$$

indicating that the expectation of the estimator of the error variance is always positively biased (Rice, 1984).

### 2.2. The estimator of the mean of the squared functions

From (2.3) we propose an estimator of the mean of squared functions

$$Q_k = \frac{1}{n} \sum_{i=1}^n Y_i^2 - \frac{1}{2(n-k)} \sum_{i=k+1}^n (Y_i - Y_{i-k})^2. \tag{2.4}$$

From (1.1), (2.3) and  $E(Y_1^2) = f_1 + \sigma^2$ , taking expectation of (2.4), we have

$$E(Q_k) = \frac{1}{n} \sum_{i=1}^n f_i^2 - \frac{1}{2(n-k)} \sum_{i=k+1}^n (f_i - f_{i-k})^2, \tag{2.5}$$

indicating that the estimator of the mean of the squared functions is always negatively biased.

Suppose that  $f$  has a bounded first derivative. Then, from (2.5), we have

$$E(Q_k) = \frac{1}{n} \sum_{i=1}^n f_i^2 - \frac{k^2}{n^2} J + O\left\{ \frac{k^3}{n^2(n-k)} \right\} + o\left( \frac{1}{n^2} \right),$$

where  $J = \int_0^1 f'(x)^2 dx / 2$ .

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

$$E(Q_k) \approx \frac{1}{n} \sum_{i=1}^n f_i^2 - \frac{k^2}{n^2} J, \quad (1 \leq k \leq m). \tag{2.6}$$

We can estimate  $Q$  from (2.1), treating (2.6) as a simple linear regression model with  $k^2/n^2$  as the independent variable. Let  $d_k = k^2/n^2$  as the independent variable and  $Q_k$  as the dependent variable. We can regress  $Q_k$  on  $d_k$  to estimate  $\alpha$  as the intercept. To be specific, we fit the linear model

$$Q_k = q_n - s_k = \alpha + \beta d_k + e_k, \quad k = 1, \dots, m,$$

where  $q_n = \sum_{i=1}^n Y_i^2/n$  and  $s_k = \sum_{i=k+1}^n (Y_i - Y_{i-k})^2 / (2(n-k))$ . To estimate the parameters, we use the weighted sum of squares  $\sum_{k=1}^m w_k (Q_k - \alpha - \beta d_k)^2$ , where  $w_k = (n-k)/N$  and  $N = \sum_{k=1}^m (n-k)$ , in details see Tong and Wang (2005).

Let  $s_w = \sum_{k=1}^m w_k Q_k$  and  $\bar{d}_w = \sum_{k=1}^m w_k d_k$ . Then

$$\hat{Q} = \hat{\alpha} = s_w - \hat{\beta} \bar{d}_w,$$

where

$$\hat{\beta} = \frac{\sum_{k=1}^m w_k Q_k (d_k - \bar{d}_w)}{\sum_{k=1}^m w_k (d_k - \bar{d}_w)^2}.$$

When necessary, the dependence of  $\hat{Q}$  on  $m$  will be expressed explicitly.

**Theorem 2.1** For the equally spaced design, we have the following;

1.  $\hat{Q}$  is unbiased when  $f$  is constant and a linear function regardless of the choice of  $m$ .
2.  $\hat{Q}$  can be written as a quadratic form;  $Y^T (I_n/tr(I_n) - D/tr(D)) Y$ , where  $I_n$  is  $n \times n$  identity matrixes and  $D$  is  $n \times n$  matrixes with elements.

$$d_{ij} = \begin{cases} \sum_{k=1}^m b_k + \sum_{k=1}^{\min(i-1, n-i, m)} b_k, & 1 \leq i = j \leq n \\ -b_{|i-j|}, & 0 < |i-j| \leq m \\ 0, & \text{otherwise,} \end{cases}$$

where

$$b_0 = 0, b_k = 1 - \frac{\bar{d}_w(d_k - \bar{d}_w)}{\sum_{k=1}^m w_k(d_k - \bar{d}_w)^2} (k = 1, \dots, m).$$

The proof of Theorem 2.1 is omitted as it is straightforward and partially shown in Tong and Wang (2005).

### 3. Asymptotic results

Using the fact that  $\hat{Q}$  has a quadratic-form representation, we have the following formula for the mean squared error:

$$\begin{aligned} MSE(\hat{Q}) &= (f^T B f)^2 + 4\sigma^2 f^T B^2 f + 4g^T [B \text{diag}(B)u] \sigma^3 \gamma_3 \\ &\quad + \sigma^4 \text{tr}[\text{diag}(B)^2] (\gamma_4 - 3) + 2\sigma^4 \text{tr}(B^2), \end{aligned}$$

where  $u = (1, \dots, 1)^T$ ,  $\gamma_i = E[(\epsilon/\sigma)^i]$ , for  $i = 3, 4$ , and  $B = I_n/tr(I_n) - D/tr(D)$  from Theorem 2.1. When the random errors are normally distributed, we obtain the following result.

$$MSE(\hat{Q}) = \text{Bias}(\hat{Q}) + \text{Var}(\hat{Q}) = (f^T B f)^2 + 4\sigma^2 f^T B^2 f + 2\sigma^4 \text{tr}(B^2).$$

**Theorem 3.1** Assume that  $f$  has a bounded second derivative,  $a_1 = \int_0^1 f(x) dx < \infty$  and  $a_2 = \int_0^1 f(x)^2 dx < \infty$ . For the equally spaced design with  $m \rightarrow \infty$  and  $m/n \rightarrow 0$ , we obtain

$$\begin{aligned} MSE(\hat{Q}) &= \frac{4a_1}{n} \sigma^3 \gamma_3 + \frac{4a_2}{n} \sigma^2 + \frac{9}{4nm} \sigma^4 \\ &\quad + \frac{9m}{112n^2} \text{var}(\epsilon^2) + o\left(\frac{1}{nm}\right) + o\left(\frac{m}{n^2}\right) + O\left(\frac{m^6}{n^6}\right). \end{aligned} \quad (3.1)$$

The last term in (3.1) comes from the bias and the remaining terms comes from the variance. Theorem 3.1 indicates that  $\hat{Q}$  a consistent estimator of  $Q$ . The asymptotically optimal bandwidth is  $m_{opt} = (28n\sigma^4/\text{var}(\epsilon^2))^{1/2}$ .

### 4. A numerical study

To evaluate our estimator of the mean of the squared functions, we assume that random errors are normally distributed with mean zero and variance  $\sigma^2$ . Then  $var(\epsilon^2) = 2\sigma^4$  and  $m_{opt} = (14n)^{1/2}$  which does not dependent on  $f$  under the conditions that  $f$  has a bounded second derivative,  $m \rightarrow \infty$  and  $m/n \rightarrow 0$ . On the discussion for selecting  $m$ , see Tong and Wang (2005). Let  $m = n^{1/2}$  and  $m = n^{1/3}$ . Also we use the same simulation setting as in Seifert, Gasser and Wolf (1993) and Dette et al. (1998):  $f(x) = 5sin(w\pi x)$ , where  $w = 1, 2$  and  $4$  corresponding to low, moderate, and high oscillations respectively, three standard deviations,  $\sigma = 0.5, 1.5, 4$ , and three sample sizes,  $n = 15, 100, 500$ . We repeat this simulation 1000 times and compute MSE which consists of the squared bias and the variance of each estimator.

Table 4.1 lists MSE (mean squared errors), Bias2 (squared biases), and Var (variance) for some estimators of the mean of the squared functions under various conditions such as the sample sizes, error variances, and the oscillations.

**Table 4.1** MSE (mean squared errors), Bias 2 (squared biases), and Var (variance)

$w$	1			2			4		
$\sigma^2$	$0.5^2$	$1.5^2$	$4^2$	$0.5^2$	$1.5^2$	$4^2$	$0.5^2$	$1.5^2$	$4^2$
$n = 15$ and $\sum_{i=1}^n f(x_i)^2/n = 12.5$									
$m = n^{1/2}$	0.8572 <sup>1</sup>	7.6575	63.2308	1.2746	8.0388	58.0211	29.5193	32.0196	65.5701
	0.0139 <sup>2</sup>	0.0055	0.0073	0.4962	0.4528	0.4509	29.1491	28.2760	28.3172
	0.8433 <sup>3</sup>	7.6520	63.2235	0.7785	7.5860	57.5702	0.3701	3.7436	37.2529
$m = n^{1/3}$	0.8667 <sup>1</sup>	8.0896	86.2894	0.8492	8.6922	82.6068	1.6648	8.2781	83.0101
	0.0007 <sup>2</sup>	0.0066	0.0309	0.0217	0.0076	0.0222	0.9346	0.6571	0.7167
	0.8661 <sup>3</sup>	8.0831	86.2584	0.8275	8.6846	82.5846	0.7302	7.6210	82.2935
$n = 100$ and $\sum_{i=1}^n f(x_i)^2/n = 12.5$									
$m = n^{1/2}$	0.1262	1.1671	8.5777	0.1237	1.1077	8.2643	0.1633	1.1443	8.2808
	0.0005	0.0032	0.0048	0.0004	0.0010	0.0003	0.0342	0.0186	0.0840
	0.1257	1.1639	8.5729	0.1233	1.1067	8.2640	0.1291	1.1257	8.1968
$m = n^{1/3}$	0.1262	1.1835	9.2539	0.1237	1.1359	9.6774	0.1330	1.1722	9.2445
	0.0003	0.0034	0.0011	0.0000	0.0028	0.0056	0.0009	0.0003	0.0098
	0.1259	1.1801	9.2528	0.1237	1.1331	9.6718	0.1322	1.1719	9.2347
HKT	0.1265	1.1622	9.3374	0.1242	1.0707	9.2281	0.2769	1.2634	8.6806
	0.0013	0.0073	0.1382	0.0030	0.0003	0.1304	0.0709	0.0403	0.0020
	0.1252	1.1549	9.1992	0.1212	1.0704	9.0977	0.2060	1.2231	8.6786
$n = 500$ and $\sum_{i=1}^n f(x_i)^2/n = 12.5$									
$m = n^{1/2}$	0.0244	0.2368	1.6029	0.0247	0.2282	1.6691	0.0246	0.2460	1.5116
	0.0001	0.0004	0.0011	0.0000	0.0000	0.0000	0.0002	0.0000	0.0005
	0.0243	0.2364	1.6019	0.0247	0.2282	1.6691	0.0244	0.2459	1.5111
$m = n^{1/3}$	0.0245	0.2390	1.7326	0.0247	0.2313	1.7703	0.0245	0.2493	1.6251
	0.0002	0.0004	0.0009	0.0000	0.0000	0.0001	0.0000	0.0000	0.0012
	0.0244	0.2386	1.7316	0.0247	0.2313	1.7702	0.0245	0.2493	1.6239

1. MSE 2. Bias2 3. Variance

The comparative performance of the estimators for the mean of the squared functions depends on the smoothness of  $f$ , the sample size and error variance. The results for all conditions is illustrated in Table 4.1.  $Q(m = n^{1/2})$  has slightly smaller mean squared errors than that of  $Q(m = n^{1/3})$  except for the cases  $(n, w, \sigma) = (15, 2, 0.5^2)$ ,  $(n, w, \sigma) = (15, 4, 0.5^2)$ , and  $(n, w, \sigma) = (15, 4, 1.5^2)$  in which  $f$  is rough and sample size is small. Therefore we

recommend  $Q(m = n^{1/3})$  when the sample size is small and  $f$  is rough.

The estimator,  $\hat{\sigma}_{HKT}^2(2)$ , introduced by Hall et al. (1990) is defined as

$$\hat{\sigma}_{HKT}^2(2) = \frac{1}{n-2} \sum_{i=1}^{n-2} (0.8090Y_i - 0.5Y_{i+1} - 0.3090Y_{i+2})^2.$$

## 5. Discussion

Most subjects in nonparametric regression models have been focusing on estimating the mean function or error variance on univariate  $x_{d=1}$  or multivariate  $x_{d>1}$ . The performance for the estimators of the mean function have been affected by the smoothness of the mean functions and error variance which SNR has been represented. This paper proposes an estimator of the power of the mean functions, using the difference based estimators for error variance which Tong and Wang (2005) provided. Our estimator is obtained from the weighted least squares with weights that decrease as distance increases.

This performance of our estimators might depend on sample sizes and oscillations. In the finite sample simulations, for the large sample the MSEs corresponding to the choice of the bandwidth  $m$  would be similar. For the small sample, however, the MSEs are varying with error variances and oscillations.

Further research on the estimators of the mean of the squared functions is necessary to explore weights for weighted least squares and the choice of the bandwidth  $m$ 's under univariate and multivariate cases.

## APPENDIX

### *Proof*

We provide a sketch of the proof only. Details of the proofs can be found in a technical report available in APPENDIX at <http://www.pstat.ucsb.edu/facult/yuedong/research> (Tong and Wang, 2005).

*Quadratic form*. The proof of the quadratic form in Theorem 2.1 can be shown as the following; rewrite the proposed estimator as

$$\hat{Q} = \bar{s}_w - \hat{\beta} \bar{d}_w = \bar{s}_w(q_n) - \hat{\beta}(q_n) \bar{d}_w - \left[ \bar{s}_w(s_k) - \hat{\beta}(s_k) \bar{d}_w \right] = q_n - \hat{\sigma}^2,$$

where  $\bar{s}_w(q_n) = \sum_{k=1}^m w_k q_n$ ,  $\bar{s}_w(s_k) = \sum_{k=1}^m w_k s_k$ ,  $\hat{\beta}(q_n) = \sum_{k=1}^m w_k q_n (d_k - \bar{d}_w) / \sum_{k=1}^m w_k (d_k - \bar{d}_w)$ ,  $\hat{\beta}(s_k) = \sum_{k=1}^m w_k s_k (d_k - \bar{d}_w) / \sum_{k=1}^m w_k (d_k - \bar{d}_w)$ ,  $q_n = \bar{s}_w(q_n) - \hat{\beta}(q_n) \bar{d}_w$  and

$\hat{\sigma}^2 = \bar{s}_w(s_k) - \hat{\beta}(s_k)\bar{d}_w$ . It is not difficult to check that

$$\begin{aligned} (a) \sum_{k=1}^m w_k &= \sum_{k=1}^m \frac{n-k}{N} = 1, \\ (b) \sum_{k=1}^m w_k(d_k - \bar{d}_w) &= \sum_{k=1}^m w_k(d_k - \bar{d}_w) \\ &= \sum_{k=1}^m (w_k d_k - w_k \bar{d}_w) \\ &= \bar{d}_w - \bar{d}_w \sum_{k=1}^m w_k \\ &= \bar{d}_w(1 - \sum_{k=1}^m w_k) \\ &= 0. \end{aligned}$$

Therefore,

$$\bar{s}_w(q_n) - \hat{\beta}(q_n)\bar{d}_w = \sum_{k=1}^m w_k q_n - \frac{\sum_{k=1}^m w_k q_n (d_k - \bar{d}_w)\bar{d}_w}{\sum_{k=1}^m w_k (d_k - \bar{d}_w)^2} = q_n,$$

and the quadratic form of  $\hat{\sigma}^2$  was proved by Tong and Wang (2005).

Assume that  $n \rightarrow \infty$ ,  $m \rightarrow \infty$  and  $m/n \rightarrow 0$ .

*Asymptotic bias.* The bias is

$$\begin{aligned} E(\hat{Q}) &= E(q_n - \hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^n f_i^2 - O\left(\frac{m^3}{n^3}\right), \\ E(\hat{Q} - Q)^2 &= O\left(\frac{m^6}{n^6}\right). \end{aligned}$$

□

*Asymptotic variance.* From Appendix in Tong and Wang (2005) we prove the followings;

$$(a) f^T \left( \frac{I_n}{tr(I_n)} - \frac{D}{tr(D)} \right)^2 f = A_1 - 2A_2 + A_3,$$

$$A_1 = \frac{f^T f}{tr(I_n)^2} = \frac{1}{n^2} \sum_{i=1}^n f_i^2 = \frac{1}{tr(I_n)} \int_0^1 f(x)^2 dx,$$

$$A_2 = \frac{f^T D f}{tr(I_n)tr(D)} = \frac{1}{tr(I_n)} O\left(\frac{m^3}{n^3}\right),$$

$$A_3 = \frac{f^T D^2 f}{tr(D)^2} = \frac{1}{tr(D)^2} O\left(\frac{m^5}{n^2}\right),$$

where  $\text{tr}(D) = 2N = 2(nm - m(m+1)/2)$ ,

$$(b) f^T \left\{ \left( \frac{I_n}{\text{tr}(I_n)} - \frac{D}{\text{tr}(D)} \right) \text{diag} \left( \frac{I_n}{\text{tr}(I_n)} - \frac{D}{\text{tr}(D)} \right) u \right\} = A_4 - A_5 - A_6 + A_7,$$

$$A_4 = \frac{f^T u}{\text{tr}(I_n)^2} = \frac{1}{n^2} \sum_{i=1}^n f_i = \frac{1}{\text{tr}(I_n)} \int_0^1 f(x) dx,$$

$$\begin{aligned} A_5 &= \frac{f^T \text{diag}(D)u}{\text{tr}(I_n)\text{tr}(D)} = \frac{1}{\text{tr}(I_n)\text{tr}(D)} \left\{ \sum_{i=1}^n f_i \left( \sum_{k=1}^m b_k + \sum_{k=0}^{\min(i-1, n-i, m)} b_k \right) \right\} \\ &= \frac{1}{\text{tr}(I_n)\text{tr}(D)} \left\{ \sum_{i=1}^n f_i \left( \left( m - \frac{5m^2}{16n} + o\left(\frac{m^2}{n}\right) \right) + \left( \frac{4}{9}l - \frac{5l^3}{4m^2} + o(l) + O(1) \right) \right) \right\} \\ &= \frac{1}{\text{tr}(I_n)\text{tr}(D)} \left\{ \int_0^1 f(x) dx \left( \left( nm - \frac{5m^2}{16} + o(m^2) \right) + \left( \frac{4nl}{9} - \frac{5nl^3}{4m^2} + o(nl) + O(n) \right) \right) \right\}, \end{aligned}$$

where  $l = \min(i-1, n-i, m)$ ,

$$\begin{aligned} A_6 &= \frac{f^T D u}{\text{tr}(I_n)\text{tr}(D)} = \frac{1}{\text{tr}(I_n)\text{tr}(D)} \sum_{i=1}^n \left( \sum_{k=1}^m b_k (f_i - f_{i-k}) - \sum_{k=1}^m b_k (f_{i+k} - f_i) \right) \\ &= \frac{1}{\text{tr}(I_n)\text{tr}(D)} o\left(\frac{m^3}{n}\right), \\ A_7 &= \frac{f^T D \text{diag}(D)u}{\text{tr}(D)^2} = \frac{1}{\text{tr}(D)^2} O\left(\frac{m^4}{n}\right), \end{aligned}$$

$$(c) \text{tr} \left\{ \text{diag} \left( \frac{I_n}{\text{tr}(I_n)} - \frac{D}{\text{tr}(D)} \right)^2 \right\} = A_8 - 2A_9 + A_{10},$$

$$A_8 = \text{tr} \left\{ \text{diag} \left( \frac{I_n}{\text{tr}(I_n)} \right)^2 \right\} = \frac{1}{\text{tr}(I_n)},$$

$$\begin{aligned} A_9 &= \text{tr} \left\{ \text{diag} \left( \frac{I_n}{\text{tr}(I_n)} \right) \text{diag} \left( \frac{D}{\text{tr}(D)} \right) \right\} \\ &= \frac{1}{\text{tr}(I_n)\text{tr}(D)} \text{tr} \{ \text{diag}(D) \} \\ &= \frac{1}{\text{tr}(I_n)\text{tr}(D)} \left\{ 2 \sum_{l=1}^m \left( \sum_{k=1}^m b_k + \sum_{k=1}^{l-1} b_k \right) + \sum_{l=m+1}^{n-m} \left( 2 \sum_{k=1}^m b_k \right) \right\} \\ &= \frac{1}{\text{tr}(I_n)\text{tr}(D)} \left( 2nm + \frac{5m^3}{4n} - m^2 + m - \frac{5}{8} + o(m^2) \right), \end{aligned}$$



$$A_{10} = \text{tr} \left\{ \text{diag} \left( \frac{D}{\text{tr}(D)} \right)^2 \right\} = \frac{1}{\text{tr}(D)^2} \left( 4nm^2 - \frac{103}{28}m^3 + o(m^3) \right),$$

$$(d) \text{tr} \left\{ \left( \frac{I_n}{\text{tr}(I_n)} - \frac{D}{\text{tr}(D)} \right)^2 \right\} = A_{11} - A_8,$$

$$A_{11} = \text{tr} \left\{ \left( \frac{D}{\text{tr}(D)} \right)^2 \right\} = \frac{1}{\text{tr}(D)^2} \left( 4nm^2 - \frac{103}{28}m^3 + \frac{9}{2}nm + o(m^3) + o(nm) \right).$$

Together with the fact that  $\sigma^4(\gamma_4 - 3) = \text{var}(\epsilon^2) - 2\sigma^4$ , we have

$$\begin{aligned} \text{var}(\hat{Q}) &= 4\sigma^2 f^T B^2 f + 4f^T [B \text{diag}(B)u] \sigma^3 \gamma_3 + \sigma^4 \text{tr}[\text{diag}(B)^2](\gamma_4 - 3) + 2\sigma^4 \text{tr}(B^2) \\ &= \frac{4a_1}{n} \sigma^3 \gamma_3 + \frac{4a_2}{n} \sigma^2 + \frac{9}{4nm} \sigma^4 + \frac{9m}{112n^2} \text{var}(\epsilon^2) + o\left(\frac{1}{nm}\right) + o\left(\frac{m}{n^2}\right) + O\left(\frac{m^6}{n^6}\right). \end{aligned}$$

where  $B = I_n/\text{tr}(I_n) - D/\text{tr}(D)$ ,  $a_1 = \int_0^1 f(x)dx < \infty$  and  $a_2 = \int_0^1 f(x)^2 dx < \infty$ .  $\square$

*Asymptotic mean squared error.* The proof of (3.1) can be completed from the asymptotic bias and variance.

## References

- Carter, C. K. and Eagleson, G. K. (1992). A comparison of variance estimators in nonparametric regression. *Journal of the Royal Statistical Society: Series B*, **54**, 773-780.
- Dette, 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. and Carroll, R. J. (1989). Variance function estimation in regression: The effect of estimating the mean. *Journal of the Royal Statistical Society: Series B*, **51**, 3-14.
- Hall, P. and Marron, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika*, **77**, 415-419.
- Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73**, 625-633.
- Neumann, M. H. (1994). Fully data-driven nonparametric variance estimators. *Statistics*, **25**, 189-212.
- Park, Chun Gun (2008). A note on a Bayesian approach to the choice of wavelet basis functions at each resolution level. *Journal of Korean Data & Information Science Society*, **19**, 1465-1476.
- Park, Chun Gun, Yeong-Hwa Kim and Wan-Youn Yang (2004). Determinacy on a maximum resolution in wavelet series. *Journal of Korean Data & Information Science Society*, **15**, 467-476.
- Rice, J. A. (1984). Bandwidth choice for nonparametric regression. *Annals of statistics*, **12**, 1215-1220.
- Seifert, B., Gasser, T. and Wolf, A. (1993). Nonparametric estimation of residual variance revisited. *Biometrika*, **80**, 373-83.
- Tong, T. and Wang, Y. (2005). Estimating residual variance in nonparametric regression using least squares. *Biometrika*, **92**, 821-830.
- Wahba, G. (1990). *Spline models for observational data*, CBMS-NSF Regional Conference Series in Applied Mathematics, 59, Philadelphia, PA: SIAM.