Journal of the Korean Data & Information Science Society
2011, **22**(3), 575–587

한국데이터정보과학회지

On statistical properties of some difference-based error variance estimators in nonparametric regression with a finite sample

Chun Gun Park[1]

[1]Kyonggi University, Department of Mathematics

Received 9 February 2011, revised 6 April 2011, accepted 28 April 2011

Abstract

We investigate some statistical properties of several difference-based error variance estimators in nonparametric regression model. Most of existing difference-based methods are developed under asymptotical properties. Our focus is on the exact form of mean and variance for the lag-k difference-based estimator and the second-order difference-based estimator in a finite sample size. Our approach can be extended to Tong's estimator (2005) and be helpful to obtain optimal k.

Keywords: Difference-based estimator, error variance, Lipschitz condition, nonparametric regression, Taylor formula.

1. Introduction

In this paper we consider a particular aspect of the statistical properties for some difference-based error variance estimators in nonparametric regression model with a finite sample size.

The most basic form of our model is

$$Y_i = g(x_i) + \epsilon_i \tag{1.1}$$

where g is an unknown mean function and the error ϵ_i's are independent and identically distributed random variables with zero mean and variance σ^2. We assume that the design points x_i's are equally spaced.

There are two types of estimating the error variance. One is that the residual sum of squares method first estimates the mean function $g(\cdot)$ (Park, 2004, 2008, 2009; Wahba, 1990; Hall and Carroll, 1989; Carter and Eagleson 1992; Neumann, 1994). The other is that the estimation of the error variance σ^2 which uses differences to remove trend in the mean function has attracted a great deal of attention; see for example Rice (1984), Gasser *et al.* (1986), Hall *et al.* (1990), Dette *et al.* (1998), and Tong and Wang (2005), among

[1] Assistant Professor, Department of Mathematics, Kyonggi University, Suwon-si 443-760, Korea.
E-mail: cgpark@kgu.ac.kr

others. So far most of existing difference-based variance estimators have been developed under asymptotical properties. That is, the mean function g is smooth and a sample size is so large. Then the effect of g on the asymptotic mean squared error of the estimator $\widehat{\sigma^2}$ of σ^2 could be ignored. In finite sample cases, however, the bias does depend on the mean function.

Therefore we focus on some statistical properties of several difference-based estimators to estimate an error variance in a nonparametric regression which both satisfies Lipschitz condition and has a small sample size.

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

$$\widehat{\sigma_R^2} = \frac{1}{2(n-1)} \sum_{i=2}^{n} (Y_i - Y_{i-1})^2 .$$

Also Rice's estimator uses differences of all consecutive observations. A lag-k Rice estimator $\widehat{\sigma_R^2}(k)$ is defined as

$$\widehat{\sigma_R^2}(k) = \frac{1}{2(n-k)} \sum_{i=1+k}^{n} (Y_i - Y_{i-k})^2 . \tag{1.2}$$

Gasser *et al.* (1986) proposed the second-order difference-based estimator

$$\widehat{\sigma_{GSJ}^2} = \frac{1}{n-2} \sum_{i=2}^{n-1} c_i^2 \widehat{\epsilon}_i^2$$

where $\widehat{\epsilon_i}$ is the difference between y_i and the value at x_i of the line joining (x_{i-1}, y_{i-1}) and (x_{i+1}, y_{i+1}). The coefficients c_i are chosen such that $E\left(c_i^2 \widehat{\epsilon_i^2}\right) = \sigma^2$ for all i when g is linear. For equidistant design points, $\widehat{\sigma_{GSJ}^2}$ reduces to

$$\widehat{\sigma_{GSJ}^2} = \frac{2}{3(n-2)} \sum_{i=2}^{n-1} (0.5Y_{i-1} - Y_i + 0.5Y_{i+1})^2 \tag{1.3}$$

Hall *et al.* (1990) introduced difference based estimators via a difference sequence $\{d_i\}_{i=-m_1,\ldots,m_2}$ with $\sum_{j=-m_1}^{m_2} d_j = 0$, $\sum_{j=-m_1}^{m_2} d_j^2 = 1$ and d_{-m_1}, $d_{m_2} \neq 0$,

$$\widehat{\sigma_{HKT}^2}(m) = \frac{1}{n-r} \sum_{i=m_1+1}^{n-m_2} \left(\sum_{j=-m_1}^{m_2} d_k Y_{k+1} \right)^2 . \tag{1.4}$$

Here m_1 and m_2 are non-negative integers and $m = m_1 + m_2$ denotes the order of the variance estimator.

Recently, Müller *et al.* (2003) and Tong and Wang (2005) proposed two new types of difference-based estimators for the error variance σ^2. Such a difference-based error variance estimator had been studied under asymptotical properties. In this article, we study the statistical properties of the lag-k Rice estimator $\widehat{\sigma_R^2}(k)$ and $\widehat{\sigma_{GSJ}^2}$ which are the basic form of difference based error variance with a finite sample.

In Section 2 we consider $\widehat{\sigma_R^2}(k)$ and $\widehat{\sigma_{GSJ}^2}$ on Lipschitz condition which the mean function g satisfies. We explore some statistical properties of the estimators with a finite sample size in Section 3. A small simulation study is conducted for examining the finite sample behavior of the proposed variance estimators in Section 4. Section 5 provides a brief discussion and further work. Some proofs of the technical results are deferred to Appendix.

2. Difference-based estimator under Lipschitz condition

In nonparametric regression model (1.1), let us consider that $g(x)$ is a mean function which satisfies Lipschitz condition,

$$|g(x) - g(y)| \leq c|x - y| \tag{2.1}$$

for some constant value $c >0$.

When $g(\cdot) \in \text{Lip}[0,1]$ and $x_i = i/n$, which is denoted by the mean function g satisfying a Lipschitz condition on the support $[0,1]$, a lag-k Rice estimator (1984) is

$$\begin{aligned}
\widehat{\sigma_R^2}(k) &= \frac{1}{2(n-k)} \sum_{i=1+k}^{n} (Y_i - Y_{i-k})^2 \\
&= \frac{1}{2(n-k)} \sum_{i=1+k}^{n} (g_i - g_{i-k})^2 + \frac{1}{(n-k)} \sum_{i=1+k}^{n} (g_i - g_{i-k})(\epsilon_i - \epsilon_{i-k}) \\
&\quad + \frac{1}{2(n-k)} \sum_{i=1+k}^{n} (\epsilon_i - \epsilon_{i-k})^2 \\
&= \frac{k^2}{2n^2(n-k)} \sum_{i=1+k}^{n} c_{i(k)}^2 + \frac{k}{n(n-k)} \sum_{i=1+k}^{n} c_{i(k)} (\epsilon_i - \epsilon_{i-k}) \\
&\quad + \frac{1}{2(n-k)} \sum_{i=1+k}^{n} (\epsilon_i - \epsilon_{i-k})^2
\end{aligned}$$

where $k = 1, \cdots, n-1$, $g_i = g(x_i)$ and $(g_i - g_{i-k}) = c_{i(k)}k/n$.

Under the same condition, Gasser *et al.* (1986) proposed the estimator,

$$\begin{aligned}
\widehat{\sigma_{GSJ}^2} &= \frac{2}{3(n-2)} \sum_{i=2}^{n-1} \left(\frac{1}{2} Y_{i-1} - Y_i + \frac{1}{2} Y_{i+1} \right)^2 \\
&= \frac{1}{6(n-2)} \sum_{i=2}^{n-1} \left[(g_{i+1} - g_i) - (g_i - g_{i-1}) + (\epsilon_{i+1} - \epsilon_i) - (\epsilon_i - \epsilon_{i-1}) \right]^2 \\
&= \frac{1}{6(n-2)} \sum_{i=2}^{n-1} \left\{ \frac{1}{n^2} \left[c_{i+1(1)} - c_{i(1)} \right]^2 + \frac{2}{n} \left[c_{i+1(1)} - c_{i(1)} \right] \left[(\epsilon_{i+1} - \epsilon_i) - (\epsilon_i - \epsilon_{i-1}) \right] \right. \\
&\quad \left. + \left[(\epsilon_{i+1} - \epsilon_i) - (\epsilon_i - \epsilon_{i-1}) \right]^2 \right\}.
\end{aligned}$$

3. Statistical properties

Now we obtain mean, variance, and mean squared error (MSE) of the above estimators from Theorem 3.1.

Theorem 3.1 When $g(\cdot) \in \text{Lip}[0,1]$ and $x_i = i/n$, the statistical properties of Rice's estimator and GSJ's are the following;

(i) Rice's estimator

$$E\left(\widehat{\sigma^2_R}(k)\right) = \sigma^2 + \frac{k^2}{2n^2(n-k)} \sum_{i=1+k}^{n} c^2_{i(k)},$$
$$Var\left(\widehat{\sigma^2_R}(k)\right) = \frac{2\sigma^2 k^2}{n^2(n-k)^2}\left(\sum_{i=1+k}^{n} c^2_{i(k)} - \sum_{i=1+k}^{n-k} c_{i(k)}c_{i+k(k)}\right) + \frac{(3n-4k)\sigma^4}{(n-k)^2},$$
$$mse\left(\widehat{\sigma^2_R}(k)\right) = \text{Bias}^2\left(\widehat{\sigma^2_R}(k)\right) + Var\left(\widehat{\sigma^2_R}(k)\right),$$

(ii) GSJ's estimator

$$E\left(\widehat{\sigma^2_{GSJ}}\right) = \sigma^2 + \frac{1}{6n^2(n-2)} \sum_{i=2}^{n-1} \left(c_{i+1(1)} - 2c_{i(1)} + c_{i-1(1)}\right)^2,$$
$$Var(\widehat{\sigma^2_{GSJ}}) = \frac{\sigma^2}{9n^2(n-2)^2}(A+B) + \frac{\sigma^4}{3(n-2)},$$
$$mse\left(\widehat{\sigma^2_{GSJ}}\right) = \text{Bias}^2\left(\widehat{\sigma^2_{GSJ}}\right) + Var\left(\widehat{\sigma^2_{GSJ}}\right),$$

where $A = \sum_{i=3}^{n-3}(c_{i+1} - 2c_i + c_{i-1})^2$, $B = c_2^2 + (c_3 - 2c_2)^2 + (c_{n-2} - 2c_{n-1})^2 + c_{n-1}^2$ and $c_i = c_{i+1(1)} - c_{i(1)}$.

Therefore, the bias of a lag-k Rice estimator is always positive and may get bigger as k increasing and positive is that of the second-order difference based estimator which depends on the shape of the function g.

To compare the statistical properties of both of the estimators, two simple functions which satisfy Lipschitz condition are suggested as linear and quadratic functions. These results are summarized in Corollaries 3.1 and 3.2, respectively.

Corollary 3.1 If $g(x) = ax$ (linear), then the statistical properties are

(i) Rice's estimator

$$\begin{aligned}
E\left(\widehat{\sigma^2_R}(k)\right) &= \sigma^2 + \frac{a^2k^2}{2n^2} \\
Var\left(\widehat{\sigma^2_R}(k)\right) &= \frac{2a^2\sigma^2k^3}{n^2(n-k)^2} + \frac{(3n-4k)\sigma^4}{(n-k)^2} \\
mse\left(\widehat{\sigma^2_R}(k)\right) &= \text{Bias}^2\left(\widehat{\sigma^2_R}(k)\right) + Var\left(\widehat{\sigma^2_R}(k)\right) \\
&= \frac{a^4k^4}{4n^4} + \frac{3\sigma^4}{n-k} + \frac{2a^2\sigma^2k}{n^2(n-k)^2}\left(k + \frac{\sigma n}{\sqrt{2}a}\right)\left(k - \frac{\sigma n}{\sqrt{2}a}\right).
\end{aligned}$$

(ii) GSJ's estimator

$$E\left(\widehat{\sigma^2_{GSJ}}\right) = \sigma^2$$
$$Var\left(\widehat{\sigma^2_{GSJ}}\right) = \frac{\sigma^4}{3(n-2)}$$
$$mse\left(\widehat{\sigma^2_{GSJ}}\right) = Var\left(\widehat{\sigma^2_{GSJ}}\right).$$

We note that for n and a fixed, the optimal k which minimizes $mse\left(\widehat{\sigma^2_R}(k)\right)$ is always one. The Rice' estimator is biased but GSJ's estimator unbiased and the MSE of the lag-k is larger than that of GSJ's.

Corollary 3.2 If $g(x) = ax^2$ (quadratic), then the statistical properties are

(i) Rice's estimator

$$E\left(\widehat{\sigma^2_R}(k)\right) = \frac{a^2k^2}{2n^4(n-k)}\sum_{i=1+k}^{n}(2i-k)^2 + \sigma^2$$
$$Var\left(\widehat{\sigma^2_R}(k)\right) = \frac{2k^2\sigma^2}{n^2(n-k)^2}\left(\frac{a^2}{n^2}\sum_{i=1+k}^{n}(2i-k)^2 - \frac{a^2}{n^2}\sum_{i=1+k}^{n-k}(2i-k)(2i+k)\right) + \frac{(3n-4k)\sigma^4}{(n-k)^2}$$
$$mse\left(\widehat{\sigma^2_R}(k)\right) = \text{Bias}^2\left(\widehat{\sigma^2_R}(k)\right) + Var\left(\widehat{\sigma^2_R}(k)\right)$$

(ii) GSJ's estimator

$$E\left(\widehat{\sigma^2_{GSJ}}\right) = \sigma^2$$
$$Var(\widehat{\sigma^2_{GSJ}}) = \frac{16a^2\sigma^2}{9n^4(n-2)^2} + \frac{\sigma^4}{3(n-2)}$$
$$mse\left(\widehat{\sigma^2_{GSJ}}\right) = Var\left(\widehat{\sigma^2_{GSJ}}\right).$$

where $c_i = c_{i+1(1)} - c_{i(1)} = 2a/n$.

The proofs of Corollary 3.1 and Corollary 3.2 are summarized by Appendix. In the quadratic form of the mean function, the results of the bias and MSE are similar to those of the linear function.

Suppose that $g(x)$ has a bounded second derivative. Then by Taylor's formula, $g(x)$ can be locally approximated by the form

$$\begin{aligned} g(x) &\approx g(x_0) + (x-x_0)g\prime(x_0) + \frac{1}{2}(x-x_0)^2 g\prime\prime(x_0) \\ &= ax^2 + bx + c. \end{aligned}$$

By Lipschitz condition, the following corollary is obtained.

Corollary 3.3 If $g(x)$ has a second bounded derivative, then the results are

(i) Rice's estimator

$$E\left(\widehat{\sigma_R^2}(k)\right) \approx \sigma^2 + \frac{k^2}{2n^2(n-k)}\left[\frac{a^2}{n^2}\sum_{i=1+k}^{n}(2i-k)^2 + \frac{2ab}{n}\sum_{i=1+k}^{n}(2i-k) + (n-k)b^2\right],$$

$$Var(\widehat{\sigma_R^2}(k)) \approx \frac{2\sigma^2 k^2}{n^2(n-k)^2}\left(\frac{a^2}{n^2}\sum_{i=1+k}^{n}(2i-k)^2 + \frac{2ab}{n}\sum_{i=1+k}^{n}(2i-k)\right.$$
$$\left. - \frac{a^2}{n^2}\sum_{i=1+k}^{n-k}(2i-k)(2i+k) - \frac{4ab}{n^2}\sum_{i=1+k}^{n-k} i + kb^2\right) + \frac{(3n-4k)\sigma^4}{(n-k)^2},$$

for a and b are any value,

(ii) GSJ's estimator

$$E\left(\widehat{\sigma_{GSJ}^2}\right) \approx \sigma^2,$$

for a and b are any value, and

$$Var\left(\widehat{\sigma_{GSJ}^2}\right) \approx \frac{16a^2\sigma^2}{9n^4(n-2)^2} + \frac{\sigma^4}{3(n-2)},$$

for $a, b > 0$.

The proofs of Corollary 3.3 are summarized by Appendix. It is not easy that the comparison of both of the estimators are analytically achieved in a finite sample size. However, the numerical approach to compare them is possible for a, b fixed.

4. Simulation study

We perform a small simulation to compare the analytical and numerical results of some Rice's estimators and GSJ's. To do this, we consider two functions, the quadratic function and the cosine, for two sample size, $x_i = i/n$ and $\epsilon_i \sim N(0, \sigma^2)$. For each simulation setting, we generate observations and calculate some statistical properties. We repeat this process 100 times and the results are summarized in Table 4.1 and Table 4.2.

From Table 4.1 and Table 4.2, the biases are larger as increasing a lag-k for Rice's. In these simulation settings, GSJ's estimator has the smaller MSEs than Rice's.

5. Discussion and further work

In this paper we obtain the exact form of some statistical properties for two difference-based error variance estimators in nonparametric regression model. This is meaningful, because most of difference-based methods have been developed under asymptotical methods. And even if there is the limitation of Rice's estimator and GSJ's estimator, most of existing difference-based estimators are traced by them.

Table 4.1 Summary for $n = 30$

		lag-1		lag-3		GSJ	
		analytic	numeric	analytic	numeric	analytic	numeric
		$g(x) = (x-0.5)^2$					
	Biased	1.74e-004	3.89e-008	1.35e-003	1.82e-006	8.23e-007	4.59e-008
$\sigma = 0.02$	Variance	4.50e-008	1.27e-008	7.32e-007	4.82e-008	1.91e-009	3.58e-009
	mse	7.51e-008	1.27e-008	2.56e-006	4.82e-008	1.91e-009	3.58e-009
	Biased	1.74e-004	6.87e-006	1.35e-003	9.54e-006	8.23e-007	3.04e-005
$\sigma = 0.1$	Variance	1.09e-005	6.79e-006	2.86e-005	8.15e-006	1.19e-006	2.26e-006
	mse	1.10e-005	6.79e-006	3.04e-005	8.15e-006	1.19e-006	2.26e-006
		$g(x) = \cos(4\pi x)$					
	Biased	4.46e-002	1.98e-003	3.73e-001	1.39e-001	2.34e-003	9.10e-007
$\sigma = 0.02$	Variance	6.37e-006	2.30e-007	4.18e-004	1.38e-005	1.91e-009	5.34e-009
	mse	2.00e-003	4.15e-006	1.39e-001	1.93e-002	5.49e-006	5.34e-009
	Biased	4.46e-002	1.99e-003	3.73e-001	1.41e-001	2.34e-003	1.96e-005
$\sigma = 0.1$	Variance	1.69e-004	1.76e-005	1.05e-002	4.47e-004	1.19e-006	3.81e-006
	mse	2.16e-003	2.16e-005	1.50e-001	2.03e-002	6.68e-006	3.81e-006

Table 4.2 Summary for $n = 15$

		lag-1		lag-3		GSJ	
		analytic	numeric	analytic	numeric	analytic	numeric
		$g(x) = (x-0.5)^2$					
	Biased	6.52e-004	4.33e-007	4.33e-003	1.88e-005	1.32e-005	4.94e-008
$\sigma = 0.02$	Variance	2.55e-007	4.20e-008	4.97e-006	4.75e-007	4.10e-009	7.88e-009
	mse	6.80e-007	4.20e-008	2.37e-005	4.75e-007	4.28e-009	7.88e-009
	Biased	6.52e-004	1.36e-005	4.33e-003	4.73e-005	1.32e-005	3.26e-005
$\sigma = 0.1$	Variance	2.65e-005	1.37e-005	1.46e-004	2.73e-005	2.56e-006	4.74e-006
	mse	2.69e-005	1.37e-005	1.65e-004	2.73e-005	2.56e-006	4.74e-006
		$g(x) = \cos(4\pi x)$					
	Biased	1.73e-001	3.01e-002	9.44e-001	8.95e-001	3.40e-002	2.82e-004
$\sigma = 0.02$	Variance	9.60e-005	7.22e-006	2.49e-003	2.48e-004	4.10e-009	1.77e-007
	mse	3.01e-002	9.13e-004	8.94e-001	8.01e-001	1.16e-003	2.57e-007
	Biased	1.73e-001	3.01e-002	9.44e-001	8.94e-001	3.40e-002	1.50e-004
$\sigma = 0.1$	Variance	2.42e-003	1.77e-004	6.23e-002	4.62e-003	2.56e-006	1.17e-005
	mse	3.25e-002	1.09e-003	9.54e-001	8.05e-001	1.16e-003	1.18e-005

Tong and Wang (2005) proposed a least square estimator to estimate the error variance as the intercept in a simple linear regression which motivated from the expectation of Rice's lag-k estimator. By taking expectation of the Rice estimator, the form is as

$$E\left[\widehat{\sigma_R^2}(k)\right] \approx \sigma^2 + \frac{k^2}{n^2}J,\ 1 \le k \le m = o(n),$$

where $J = \dfrac{1}{2}\int_0^1 g'(x)^2dx$. By treating the above expectation as a simple linear regression model with a regressor $d_k = k^2/n^2$, they considered the following model,

$$s_k = \sigma^2 + \beta d_k + e_k, k = 1, 2, ..., m,$$

and estimated σ^2 as the intercept, where $s_k = \sum(y_i - y_{i-k})^2/2(n-k)$ and e_k 's are dependent random variables. Under asymptotical properties, the proposed simple model is working well. In a finite sample, however, the simple regression is not the exact form. That is, the coefficient β is not constant but depends on a lag-k or the regressor should be changed.

Further research is necessary to analysis Tong's estimator (2005) which used a least square method using the assumption of Lipschitz condition for the mean function and Taylor expansion. And the estimator of the error variance and the optimal lag-k of the Tong's estimator are investigated.

Appendix

Proof of Theorem 3.1

The variance of Rice's estimator is the form as

$$\begin{aligned}
&Var\left[\widehat{\sigma^2_R}(k)\right] \\
&= Var\left[\frac{k^2}{2n^2(n-k)}\sum_{i=1+k}^{n} c^2_{i(k)} + \frac{k}{n(n-k)}\sum_{i=1+k}^{n} c_{i(k)}\left(\epsilon_i - \epsilon_{i-k}\right)\right. \\
&\qquad \left. + \frac{1}{2(n-k)}\sum_{i=1+k}^{n}\left(\epsilon_i - \epsilon_{i-k}\right)^2\right] \\
&= Var\left[\frac{k}{n(n-k)}\sum_{i=1+k}^{n} c_{i(k)}\left(\epsilon_i - \epsilon_{i-k}\right) + \frac{1}{2(n-k)}\sum_{i=1+k}^{n}\left(\epsilon_i - \epsilon_{i-k}\right)^2\right] \\
&= Var\left[\frac{k}{n(n-k)}\sum_{i=1+k}^{n} c_{i(k)}\left(\epsilon_i - \epsilon_{i-k}\right)\right] + Var\left[\frac{1}{2(n-k)}\sum_{i=1+k}^{n}\left(\epsilon_i - \epsilon_{i-k}\right)^2\right] \\
&\quad + Cov\left[\frac{k}{n(n-k)}\sum_{i=1+k}^{n} c_{i(k)}\left(\epsilon_i - \epsilon_{i-k}\right), \frac{1}{2(n-k)}\sum_{i=1+k}^{n}\left(\epsilon_i - \epsilon_{i-k}\right)^2\right].
\end{aligned}$$

The term of covariance is

$$Cov\left[\frac{k}{n(n-k)}\sum_{i=1+k}^{n} c_{i(k)}\left(\epsilon_i - \epsilon_{i-k}\right), \frac{1}{2(n-k)}\sum_{i=1+k}^{n}\left(\epsilon_i - \epsilon_{i-k}\right)^2\right] = 0$$

by $E(\epsilon_i) = 0$, $E(\epsilon_i^3) = 0$, $\forall i$ and $E(\epsilon_i\epsilon_j) = 0$, $E(\epsilon_i^2\epsilon_j) = 0$, $\forall i \neq j$.
Therefore,

$$Var\left[\widehat{\sigma^2_R}(k)\right] = 2\sigma^2\frac{k^2}{n^2(n-k)^2}\left(\sum_{i=1+k}^{n} c^2_{i(k)} - \sum_{i=1+k}^{n-k} c_{i(k)}c_{i+k(k)}\right) + \sigma^4\frac{(3n-4k)}{(n-k)^2}.$$

That of GSJ's estimator is

$$\begin{aligned}
Var\left(\widehat{\sigma^2_{GSJ}}\right) &= Var\left\{\frac{1}{3n(n-2)}\sum_{i=2}^{n-1}\left[c_{i+1(1)} - c_{i(1)}\right]\left[(\epsilon_{i+1} - \epsilon_i) - (\epsilon_i - \epsilon_{i-1})\right]\right. \\
&\qquad \left. + \frac{1}{6(n-2)}\sum_{i=2}^{n-1}\left[(\epsilon_{i+1} - \epsilon_i) - (\epsilon_i - \epsilon_{i-1})\right]^2\right\} \\
&= Var\left\{\frac{1}{3n(n-2)}\sum_{i=2}^{n-1}\left[c_{i+1(1)} - c_{i(1)}\right]\left[(\epsilon_{i+1} - \epsilon_i) - (\epsilon_i - \epsilon_{i-1})\right]\right\}
\end{aligned}$$

$$+ Var\left\{\frac{1}{6(n-2)}\sum_{i=2}^{n-1}\left[(\epsilon_{i+1}-\epsilon_i)-(\epsilon_i-\epsilon_{i-1})\right]^2\right\}+Cov\left(A,B\right),$$

where

$$A=\sum_{i=2}^{n-1}\left[c_{i+1(1)}-c_{i(1)}\right]\left[(\epsilon_{i+1}-\epsilon_i)-(\epsilon_i-\epsilon_{i-1})\right]/3n(n-2)$$

and

$$B=\sum_{i=2}^{n-1}\left[(\epsilon_{i+1}-\epsilon_i)-(\epsilon_i-\epsilon_{i-1})\right]^2/6(n-2).$$

$$\begin{aligned}&Var\left\{\frac{1}{3n(n-2)}\sum_{i=2}^{n-1}\left[c_{i+1(1)}-c_{i(1)}\right]\left[(\epsilon_{i+1}-\epsilon_i)-(\epsilon_i-\epsilon_{i-1})\right]\right\}\\&=\frac{\sigma^2}{9n^2(n-2)^2}\left[c_2^2+(c_3-2c_2)^2+\sum_{i=3}^{n-3}(c_{i+1}-2c_i+c_{i-1})^2+(c_{n-2}-2c_{n-1})^2+c_{n-1}^2\right],\end{aligned}$$

for $c_i=c_{i+1(1)}-c_{i(1)}$,

$$\begin{aligned}&Var\left\{\frac{1}{6(n-2)}\sum_{i=2}^{n-1}\left[(\epsilon_{i+1}-\epsilon_i)-(\epsilon_i-\epsilon_{i-1})\right]^2\right\}\\&=\frac{1}{36(n-2)^2}Var\left\{\sum_{i=2}^{n-1}\left[\epsilon_{i+1}-2\epsilon_i+\epsilon_{i-1})\right]^2\right\}\\&=\frac{\sigma^4}{3(n-2)},\end{aligned}$$

and

$$\begin{aligned}Cov\Bigg\{&\frac{1}{3n(n-2)}\sum_{i=2}^{n-1}\left[c_{i+1(1)}-c_{i(1)}\right]\left[(\epsilon_{i+1}-\epsilon_i)-(\epsilon_i-\epsilon_{i-1})\right],\\&\frac{1}{6(n-2)}\sum_{i=2}^{n-1}\left[(\epsilon_{i+1}-\epsilon_i)-(\epsilon_i-\epsilon_{i-1})\right]^2\Bigg\}=0\end{aligned}$$

by $E(\epsilon_i)=0$, $E(\epsilon_i^3)=0$, $\forall i$ and $E(\epsilon_i\epsilon_j)=0$, $E(\epsilon_i^2\epsilon_j)=0$, $\forall i\neq j$.

Therefore

$$Var\left(\widehat{\sigma^2_{GSJ}}\right)$$
$$= \frac{\sigma^2}{9n^2(n-2)^2}\left[c_2^2+(c_3-2c_2)^2+\sum_{i=3}^{n-3}(c_{i+1}-2c_i+c_{i-1})^2+(c_{n-2}-2c_{n-1})^2+c_{n-1}^2\right]$$
$$+\frac{\sigma^4}{3(n-2)}.$$

Proof of Corollary 3.1
When $g(x) = ax$, we have the following;

$$g(x_i) - g(x_{i-k}) = c_{i(k)}(x_i - x_{i-k})$$
$$a\frac{i}{n} - a\frac{i-k}{n} = c_{i(k)}\left(\frac{i}{n} - \frac{i-k}{n}\right)$$
$$a = c_{i(k)}$$

$$\because c_{i(k)} = a,$$

$$E\left(\widehat{\sigma^2_R}(k)\right) = \sigma^2 + \frac{k^2}{2n^2(n-k)}\sum_{i=1+k}^{n} c_{i(k)}^2 = \sigma^2 + \frac{a^2k^2}{2n^2},$$
$$Var\left(\widehat{\sigma^2_R}(k)\right) = \frac{2a^2\sigma^2k^3}{n^2(n-k)^2} + \frac{(3n-4k)\sigma^4}{(n-k)^2},$$

and

$$\begin{aligned} mse\left(\widehat{\sigma^2_R}(k)\right) &= \text{Bias}^2\left(\widehat{\sigma^2_R}(k)\right) + Var\left(\widehat{\sigma^2_R}(k)\right) \\ &= \frac{a^4k^4}{4n^4} + \frac{2a^2\sigma^2k^3}{n^2(n-k)^2} + \frac{(3n-4k)\sigma^4}{(n-k)^2} \\ &= \frac{a^4k^4}{4n^4} + \frac{2a^2\sigma^2k^3}{n^2(n-k)^2} + \frac{3\sigma^4}{n-k} - \frac{k\sigma^4}{(n-k)^2} \\ &= \frac{a^4k^4}{4n^4} + \frac{3\sigma^4}{n-k} + \frac{2a^2\sigma^2k}{n^2(n-k)^2}\left(k^2 - \frac{\sigma^2n^2}{2a^2}\right) \\ &= \frac{a^4k^4}{4n^4} + \frac{3\sigma^4}{n-k} + \frac{2a^2\sigma^2k}{n^2(n-k)^2}\left(k + \frac{\sigma n}{\sqrt{2}a}\right)\left(k - \frac{\sigma n}{\sqrt{2}a}\right). \end{aligned}$$

The proof of those of GSJ's estimator is omitted as it is straightforward.

Proof of Corollary 3.2
When $g(x) = ax^2$, we have the following;

$$\begin{aligned} g(x_i) - g(x_{i-k}) &= c_{i(k)}(x_i - x_{i-k}) \\ a\frac{i^2}{n^2} - a\frac{(i-k)^2}{n^2} &= c_{i(k)}\left(\frac{i}{n} - \frac{i-k}{n}\right) \\ \frac{n}{k}\left(a\frac{i^2}{n^2} - a\frac{(i-k)^2}{n^2}\right) &= c_{i(k)} \\ a\frac{2i-k}{n} &= c_{i(k)} \end{aligned}$$

$$\because c_{i(k)} = a\frac{2i-k}{n}.$$

(i) Rice's estimator

$$E\left(\widehat{\sigma^2_R}(k)\right) = \sigma^2 + \frac{k^2}{2n^2(n-k)}\sum_{i=1+k}^{n} c^2_{i(k)} = \sigma^2 + \frac{a^2k^2}{2n^4(n-k)}\sum_{i=1+k}^{n}(2i-k)^2,$$

and

$$\begin{aligned} Var\left(\widehat{\sigma^2_R}(k)\right) &= 2\sigma^2\frac{k^2}{n^2(n-k)^2}\left(\sum_{i=1+k}^{n} c^2_{i(k)} - \sum_{i=1+k}^{n-k} c_{i(k)}c_{i+k(k)}\right) + \frac{(3n-4k)\sigma^4}{(n-k)^2} \\ &= 2\sigma^2\frac{a^2k^2}{n^4(n-k)^2}\left(\sum_{i=1+k}^{n}(2i-k)^2 - \sum_{i=1+k}^{n-k}(2i-k)(2i+k)\right) + \frac{(3n-4k)\sigma^4}{(n-k)^2}, \end{aligned}$$

where

$$\begin{aligned} \sum_{i=1+k}^{n}(2i-k)^2 &= \sum_{i=1}^{n}(2i-k)^2 - \sum_{i=1}^{k}(2i-k)^2 \\ &= \frac{2}{3}n(n+1)(2n+1) - 2n(n+1)k + nk^2 - \frac{2}{3}k(k+1)(2k+1) \\ &\quad + 2k(k+1)k - k^3, \end{aligned}$$

and

$$\begin{aligned} \sum_{i=1+k}^{n}(2i-k) &= \sum_{i=1}^{n}(2i-k) - \sum_{i=1}^{k}(2i-k) \\ &= n(n+1) - nk - k(k+1) + k^2. \end{aligned}$$

(ii) GSJ's estimator

$$
\begin{aligned}
&Var(\widehat{\sigma^2_{GSJ}})\\
=&\frac{\sigma^2}{9n^2(n-2)^2}\left[c_2^2+(c_3-2c_2)^2+\sum_{i=3}^{n-3}(c_{i+1}-2c_i+c_{i-1})^2+(c_{n-2}-2c_{n-1})^2+c_{n-1}^2\right]\\
&+\frac{\sigma^4}{3(n-2)}\\
=&\frac{16a^2\sigma^2}{9n^4(n-2)^2}+\frac{\sigma^4}{3(n-2)},
\end{aligned}
$$

where $c_i = c_{i+1(1)} - c_{i(1)} = 2a/n$.

Proof of Corollary 3.3
If $g(x)$ is a quadratic function, then

$$
\begin{aligned}
g(x_i)-g(x_{i-k}) &= c_{i(k)}(x_i-x_{i-k})\\
ax_i^2+bx_i-ax_{i-k}^2-bx_i &\approx c_{i(k)}(x_i-x_{i-k})\\
a\frac{i^2}{n^2}-a\frac{(i-k)^2}{n^2}+b\frac{i}{n}-b\frac{i-k}{n} &\approx c_{i(k)}\left(\frac{i}{n}-\frac{i-k}{n}\right)\\
a\frac{2i-k}{n}+b &\approx c_{i(k)}
\end{aligned}
$$

$$
\because c_{i(k)} \approx a\frac{2i-k}{n}+b
$$

(i) Rice's estimator

$$
\begin{aligned}
E\left(\widehat{\sigma^2_R}(k)\right) &= \sigma^2+\frac{k^2}{2n^2(n-k)}\sum_{i=1+k}^{n}c_{i(k)}^2\\
&\approx \sigma^2+\frac{k^2}{2n^2(n-k)}\left[\frac{a^2}{n^2}\sum_{i=1+k}^{n}(2i-k)^2+\frac{2ab}{n}\sum_{i=1+k}^{n}(2i-k)+(n-k)b^2\right],
\end{aligned}
$$

and

$$
\begin{aligned}
Var\left(\widehat{\sigma^2_R}(k)\right) &= \frac{2\sigma^2k^2}{n^2(n-k)^2}\left(\sum_{i=1+k}^{n}c_{i(k)}^2-\sum_{i=1+k}^{n-k}c_{i(k)}c_{i+k(k)}\right)+\frac{(3n-4k)\sigma^4}{(n-k)^2}\\
&\approx \frac{2\sigma^2k^2}{n^2(n-k)^2}\left(\frac{a^2}{n^2}\sum_{i=1+k}^{n}(2i-k)^2+\frac{2ab}{n}\sum_{i=1+k}^{n}(2i-k)\right.\\
&\quad\left.-\frac{a^2}{n^2}\sum_{i=1+k}^{n-k}(2i-k)(2i+k)-\frac{4ab}{n^2}\sum_{i=1+k}^{n-k}i+kb^2\right)+\frac{(3n-4k)\sigma^4}{(n-k)^2},
\end{aligned}
$$

(ii) GSJ's estimator

$$\begin{aligned} E\left(\widehat{\sigma^2_{GSJ}}\right) &= \sigma^2 + \frac{1}{6n^2(n-2)} \sum_{i=2}^{n-1} \left[(c_{i+1(1)} - c_{i(1)}) - (c_{i(1)} - c_{i-1(1)}) \right]^2 \\ &= \sigma^2 + \frac{1}{6n^2(n-2)} \sum_{i=2}^{n-1} \left(c_{i+1(1)} - 2c_{i(1)} + c_{i-1(1)} \right)^2 \\ &\approx \sigma^2 + \frac{1}{6n^2(n-2)} \sum_{i=2}^{n-1} \left[a\frac{2i+1}{n} + b - 2\left(a\frac{2i-1}{n} + b \right) + a\frac{2i-3}{n} + b \right]^2 \\ &= \sigma^2, \end{aligned}$$

the proof of the variance is omitted as it is straightforward.

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.

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

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

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

Neumann, M. H. (1994). Fully data-driven nonparametric variance estimators. *Statistics*, **25**, 189-212.

Park, C. G. (2009). An estimator of the mean of the squared functions for a nonparametric regression. *Journal of the Korean Data & Information Science Society*, **20**, 577-585.

Park, C. G. (2008). A note on a Bayesian approach to the choice of wavelet basis functions at each resolution level. *Journal of the Korean Data & Information Science Society*,**19**, 1465-1476.

Park, C. G., Kim, Y. H. and Yang, W. Y. (2004). Determinacy on a maximum resolution in wavelet series. *Journal of the Korean Data & Information Science Society*, **15**, 467-476.

Rice, J. A. (1984). Bandwidth choice for nonparametric regression. *Annals of Statistics*, **12**, 1215-1220.

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, Vol. 59, SIAM, Philadelphia.