Journal of the Korean Data & Information Science Society
2011, **22**(6), 1223–1232

한국데이터정보과학회지

Estimation of error variance in nonparametric regression under a finite sample using ridge regression

Chun Gun Park[1]

[1]Kyonggi University, Department of Mathematics

Received 18 July 2011, revised 29 August 2011, accepted 27 September 2011

Abstract

Tong and Wang's estimator (2005) is a new approach to estimate the error variance using least squares method such that a simple linear regression is asymptotically derived from Rice's lag- estimator (1984). Their estimator highly depends on the setting of a regressor and weights in small sample sizes. In this article, we propose a new approach via a local quadratic approximation to set regressors in a small sample case. We estimate the error variance as the intercept using a ridge regression because the regressors have the problem of multicollinearity. From the small simulation study, the performance of our approach with some existing methods is better in small sample cases and comparable in large cases. More research is required on unequally spaced points.

Keywords: Difference-based estimator, least squares, Lipschitz condition, nonparametric regression, ridge regression, Taylor expansion.

1. Introduction

We consider the nonparametric regression model

$$y_i = g(x_i) + \epsilon_i, i = 1, \cdots, n, \tag{1.1}$$

where y_i's are observations, g is an unknown mean function, and ϵ_i's are independent and identically distributed random error with zero mean and common variance σ^2. We assume that the design points x_i's are equally spaced. The problem we are interested in is estimating the error variance when the mean $g(x)$ is unknown and a sample size is small. In other word, the mean $g(x)$ plays the role of a nuisance parameter. The problem of variance estimation in nonparametric regression with large sample cases was seriously considered from the 1980s (see, Rice, 1984; Gasser *et al.*, 1986; Hall *et al.*, 1990; Dette *et al.*, 1998; Park, 2009). To overcome this problem, some methods of estimators use differences to remove trend in the mean, an idea originating from time series analysis. We refer to estimators in these types as difference-based estimators which are attractive from a practical point of view because they often have biases for small sample sizes (Dette *et al.*, 1998).

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

Recently, Tong and Wang (2005) proposed a new type of difference-based estimators for the error variance σ^2. This new estimator is a difference-based estimator using least squares. That is, the error variance is estimated as the intercept in a simple linear regression with squared differences of paired observations as the dependent variable and squared distances between the paired covariates as the regressor under asymptotical properties.

As mentioned, all of difference based methods have been developed in large sample cases but none of them has been analytically attempted in small sample cases. Tong and Wang's estimator is a new approach, but the proposed regressor is vague.

In this article, we present a multiple regression model of a lag- k Rice estimator. The proposed multiple regression model has the problem of multicollinearity when the unknown mean function satisfies Lipschitz condition and a sample size is small. To overcome this problem, we employ a ridge regression model. In Section 2, we review some statistical properties of Tong and Wang's estimator. In Section 3, we propose some multiple regression model of the lag- k Rice estimator under Lipschitz condition and a small sample size case. In Section 4, We compare the performance of our estimator with existing approaches. We conclude the paper with a brief discussion in Section 5. Some proofs of the technical results are deferred to Appendix.

2. Tong and Wang estimator

Rice (1984) proposed the lag-k difference-based estimator which is always positively biased.

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

Suppose that the mean function g has a bounded first derivate. Motivated by the expectation of the Rice estimator, Tong and Wang (2005) proposed

$$E\left(\widehat{\sigma_R^2}(k)\right) = \sigma^2 + Jd_k,\ 1 \le k \le o(n) \tag{2.1}$$

where $J = \int_0^1 \left[g'(x)\right]^2 dx/2$ and $d_k = k^2/n^2$. Treating (2.1) as a simple linear regression model with d_k as the independent variable, they considered the linear model

$$s_k = \sigma^2 + \beta d_k + \epsilon_k,\ k = 1, \cdots, m \tag{2.2}$$

and estimated σ^2 as the intercept, where $s_k = \widehat{\sigma_R^2}(k)$ and e_k's are dependent random variables. Since s_k is the average of $(n-k)$ lag-k differences, they assigned weight $w_k = (n-k)/N$ to the observation s_k, where $N = \sum_{k=1}^{n}(n-k)$. Using the weighted least square, the error variance σ^2 was estimated as $\widehat{\sigma^2} = \bar{s_w} - \widehat{\beta}\bar{d_w}$ where

$$\bar{s}_w = \sum_{k=1}^{m} w_k s_k,\ \bar{d_w} = \sum_{k=1}^{m} w_k d_k,\ \text{and}\ \widehat{\beta} = \frac{\sum_{k=1}^{m} w_k s_k (d_k - \bar{d_w})}{\sum_{k=1}^{m} w_k (d_k - \bar{d_w})^2}.$$

For selecting m, they proposed three types; $m = n^{1/2}$ for a small sample size, $m = n^{1/3}$ for a large sample size or $m = n^{1/3}$ regardless of sample sizes.

Theorem 2.1 Let $x_i = i/n$, $i = 1, 2..., n$. Then the following properties hold.
(i) $\sigma^2 = E\left(\widehat{\sigma^2}\right)$, if $g(x) = a + bx$ regardless of the choice of m where $a, b \in R$.
(ii) $\widehat{\sigma^2} = y^T Dy/tr(D)$, where $D = D_{(k)}^T D_{(k)}$,

$$D_{(k)} = \begin{pmatrix} -1 & 0_{(k-1)} & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & -1 & 0_{(k-1)} & 1 \end{pmatrix} \in \mathrm{R}^{(n-k)\times n}$$

and $0_{(k-1)}$ is denoted as the number of zeroes being $k-1$.

The proof of Theorem 2.1 is omitted as it is straightforward (Tong and Wang, 2005). Their experience indicated that negative estimates could happen for other functions. The reason it could happen might be that Tong and Wang's estimator was asymptotically developed and the regressors d_k and the weights w_k were fixed. Therefore we propose a new approach via a local quadratic approximation to set regressor(s) d_k in a finite sample.

3. Error variance estimator using a ridge regression model

3.1. Rice's lag-k estimator

Park (2011) derived some statistical properties from Rice's lag- k estimator (1984) in (2.1) under a small sample size and Lipschitz condition,

$$|g(x_i) - g(x_j)| <= c_{i(j-i)}(x_i - x_j), \quad x_j < x_i,$$

for some constant value $c_{i(j-i)}$ >0.

Theorem 3.1 When $g(\cdot) \in \mathrm{Lip}[0,1]$ and $x_i = i/n$, some statistical properties of Rice's lag-k estimator are
(i) $E\left(\widehat{\sigma^2_R}(k)\right) \leq \sigma^2 + \dfrac{k^2}{2n^2(n-k)} \sum_{i=1+k}^{n} c_{i(k)}^2$,
(ii) $Var\left(\widehat{\sigma^2_R}(k)\right) \leq \dfrac{2\sigma^2 k^2}{n^2(n-k)^2}\left(\sum_{i=1+k}^{n} c_{i(k)}^2 - \sum_{i=1+k}^{n-k} c_{i(k)} c_{i+k(k)}\right) + 2\sigma^4 \dfrac{tr(D^2)}{tr(D)^2}$.

The proof of Theorem 3.1 are provided from Park (2011). From the expectations in (2.1) and (i) of Theorem 3.1, Rice's lag- k estimator is always positively biased and the coefficients depend on the conditions of the unknown function and sample sizes.

Motivated by Theorem 3.1, we fit the linear model

$$s_k = \sigma^2 + \beta d_k + e_k, \quad k = 1, 2, ..., m, (m < n), \tag{3.1}$$

where

$$
\begin{aligned}
&s_k = \widehat{\sigma^2}(k) \\
&\beta = \frac{1}{2}, \\
&d_k = \frac{k^2}{n^2(n-k)} \sum_{i=1+k}^{n} c_{i(k)}^2, \\
&E(e_k) = 0, \\
&Cov(e_i, e_j) \neq 0, \quad \forall i \neq j,
\end{aligned}
$$

and

$$
Var\,(e_k) = \frac{2k^2\sigma^2}{n^2(n-k)^2}\left(\sum_{i=1+k}^{n} c_{i(k)}^2 - \sum_{i=1+k}^{n-k} c_{i(k)}c_{i+k(k)}\right) + 2\sigma^4\frac{tr(D^2)}{tr(D)^2}.
$$

Therefore the linear model (3.1) is not a simple regression model because the regressor is unknown. Tong and Wang (2005), however, treated the expectation of Rice's lag- k estimator as a simple linear regression model with $d_k = k^2/n^2$ and $\beta = \int_0^1 [g'(x)]^2/2dx$.

3.2. Local quadratic approximation

From (3.1), using ordinary least squares the regressors could not be estimated. Now we consider the unknown mean function as a local quadratic function to obtain the regressors d_k. To do this, we assume that $g(x)$ has a bounded second derivative. The regressors could be fixed and this result is explained in Theorem 3.2.

Theorem 3.2 Suppose that $x_i = i/n,\ i = 1, ..., n$.
(i) If $g(x) = ax^2 + bx + c$, then

$$
\beta d_k = \frac{a^2k^2}{2n^4(n-k)} \sum_{i=1+k}^{n} (2i-k)^2 + \frac{2abk^2(n+1)}{2n^3} + \frac{b^2k^2}{2n^2},
$$

(ii) If $g(x)$ has a bounded second derivative, then

$$
\beta d_k \leq \beta_1 \frac{k^2}{2n^4(n-k)} \sum_{i=1+k}^{n} (2i-k)^2 + \beta_2\frac{k^2(n+1)}{n^3} + \beta_3\frac{k^2}{n^2}.
$$

where β's are some constants.

The proof of Theorem 3.2 is provided in Appendix. The unknown mean function $g(x)$ can be locally approximated by a quadratic function as

$$
g(x) = g(t) + g'(t)(x-t) + \frac{g(t)}{2}(x-t)^2 + R_{q>2},
$$

where $R_{q>2}$ represents the remainder of Taylor expansion.

3.3. Ridge regression model

From Theorem 3.2, we fit the linear regression using an ordinary least square method (OLS)

$$s_k = \sigma^2 + \beta_1 d_{1,k} + \beta_2 d_{2,k} + \beta_3 d_{3,k} + e_k, \quad k = 1, 2, ..., m, (m < n), \tag{3.2}$$

where

$$d_{1,k} = \frac{k^2}{2n^4(n-k)} \sum_{i=1+k}^{n} (2i-k)^2, \quad d_{2,k} = \frac{k^2(n+1)}{n^3}, \quad d_{3,k} = \frac{k^2}{n^2}.$$

Since two or all regressors from the above model are highly correlated, the problem of multicollinearity should occur and OLS could be inappropriate to estimate σ^2 as the intercept. To overcome this problem, several methods have been studied; (1) variable(s) deletion (2) a regression on principal component (3) a ridge regression, etc. In this case, we employ the ridge regression to estimate σ^2, because all regressors are important for estimating the error variance, whereas the remaining methods are related with model selection.

To estimate the intercept from (3.2), we use ridge regression, proposed by Hoerl and Kennard (1970a, 1970b), which is an alternative to the principal components regression. In the linear regression that $Y = X\beta + \epsilon$ with the design matrix X consisting of the regressors and ϵ being an error term from (3.2), the ridge estimator is given by

$$\widehat{\beta}(t) = \left[X^T X + tI\right]^{-1} X^T Y,$$

with $t \geq 0$, the nonstochastic quantity, being the control parameter. Of course, $\widehat{\beta}(0)$ is the OLS estimator. The mean and variance of $\widehat{\beta}(t)$ are easily derived as the following;

$$E\left[\widehat{\beta}(t)\right] = \left(X^T X + tI_p\right)^{-1} \left(X^T X\right) \beta$$

$$Var\left[\widehat{\beta}(t)\right] = \sigma^2 \left(X^T X + tI_p\right)^{-1} \left(X^T X\right) \left(X^T X + tI_p\right)^{-1}.$$

One of the practical problems in the ridge regression is the choice of t. Popular methods are the ridge trace plot, cross-validation, generalized cross validation and so on. To choose the control parameter t, the generalized cross-validation (GCV) is employed which is quite robust to the violation of model assumptions. One chooses t such that it fulfills some optimality criteria, e.g. that it minimizes GCV score defined as

$$\frac{1}{n} \sum_{i=1}^{n} \left[\frac{y_i - x_i^T \widehat{\beta}(t)}{1 - \frac{tr(A(t))}{n}} \right]^2 \tag{3.3}$$

where x_i is the i th column vector in the design matrix X and $a_{ii}(t)$ is the i th diagonal element of

$$A(t) = X \left(X^T X + tI\right)^{-1} X^T.$$

Here we can estimate the error variance σ^2 using GCV from (3.3) in the ridge regression.

4. Simulation study

In our comparisons, we evaluate the performance of Gasser *et al.*, Hall *et al.*, Tong and Wang's and our estimator:

1. Estimator proposed by Gasser *et al.* (GJS):

$$\widehat{\sigma^2}_{GSJ} = \frac{2}{3(n-2)}\sum_{i=2}^{n-1}\left(\frac{1}{2}y_{i-1} - y_i + \frac{1}{2}y_{i+1}\right)^2$$

2. Estimator proposed by Hall *et al.* (HKT)

$$\widehat{\sigma^2}_{HKT} = \frac{1}{n-2}\sum_{i=1}^{n-2}(0.8090y_{i-1} - 0.5y_{i+1} - 0.3090y_{i+2})^2$$

3. Estimator proposed by Tong and Wang (TW)

$$\widehat{\sigma^2}_{TW} = \bar{s}_w - \widehat{\beta}\bar{d}_w,$$

where $w_k = (n-k)/N$, $\bar{s}_w = \sum_{k=1}^m w_k s_k$, $\widehat{\beta} = w_k s_k (d_k - \bar{d}_w)/\sum_{k=1}^m w_k (d_k - \bar{d}_w)^2$, $\bar{d}_w = \sum_{k=1}^m w_k d_k$, $d_k = k^2/n^2$, and $m = n^{1/3}$.

4. Our estimator is from the ridge regression in (3.2). We choose m to have the largest correlation between $d_{1,k}$'s and s_k's, that is $m = \arg\max_m corr(d_{1,k}\text{'s},\ s_k\text{'s})$, because of linearity.

Table 4.1 Mean squares errors (MSE) for $n = 10$ and $a = 1$

σ^2		$\widehat{\sigma}^2_{GSJ}$	$\widehat{\sigma}^2_{HKT}$	$\widehat{\sigma}^2_{TW}$	$\widehat{\sigma}^2_{RG}$
0.01^2	Bias2	6.10E-04	1.31E+00	1.41E-02	4.76E-04
	Variance	2.75E-08	3.46E-05	5.92E-06	1.67E-05
	MSE	6.10E-04	1.31E+00	1.41E-02	4.93E-04
0.1^2	Bias2	6.70E-04	1.31E+00	1.48E-02	6.77E-04
	Variance	5.59E-05	3.09E-03	5.78E-04	6.83E-04
	MSE	7.26E-04	1.32E+00	1.53E-02	1.36E-03
1^2	Bias2	5.08E-01	1.94E+00	4.43E-01	6.32E-01
	Variance	5.09E-01	6.16E-01	4.37E-01	6.38E-01
	MSE	1.02E+00	2.56E+00	8.80E-01	1.27E+00
5^2	Bias2	2.79E+02	1.73E+02	2.48E+02	3.82E+02
	Variance	2.82E+02	1.75E+02	2.50E+02	3.84E+02
	MSE	5.61E+02	3.48E+02	4.97E+02	7.66E+02

In a simulation study, we use the same simulation setting as in Seifert *et al.* (1993) and Dette *et al.* (1998): $g(x) = 5\sin(a\pi x)$, where a is the frequency of the unknown mean function, $x_i = i/n$ and $\epsilon_i \sim N(0,\sigma^2)$. We consider four different frequencies, $a = 1, 4, 10$ which correspond to low, median and high oscillation, respectively. The four different error

Table 4.2 MSEs for $n = 10, 15$

n	a	σ^2	$\widehat{\sigma}^2_{GSJ}$	$\widehat{\sigma}^2_{HKT}$	$\widehat{\sigma}^2_{TW}$	$\widehat{\sigma}^2_{RG}$
10	4	0.01^2	1.84E+01	2.14E+02	1.08E+01	**2.70E − 03**
		0.1^2	1.84E+01	2.14E+02	1.09E+01	**7.92E − 02**
		1^2	2.10E+01	2.29E+02	1.32E+01	**2.16E + 00**
		5^2	6.36E+02	9.56E+02	**5.64E + 02**	7.72E+02
	10	0.01^2	4.45E+03	**6.25E + 02**	4.45E+03	8.56E+03
		0.1^2	4.41E+03	**6.20E + 02**	4.41E+03	8.50E+03
		1^2	4.77E+03	**6.94E + 02**	4.74E+03	9.13E+03
		5^2	9.98E+03	**1.58E + 03**	9.73E+03	1.87E+04
15	1	0.01^2	**2.08E − 05**	3.31E-01	1.03E-03	2.81E-04
		0.1^2	**6.89E − 05**	3.27E-01	1.34E-03	2.16E-03
		1^2	**6.03E − 01**	8.72E-01	5.67E-01	6.85E-01
		5^2	4.80E+02	**3.16E + 02**	4.62E+02	6.19E+02
	4	0.01^2	1.01E+00	6.59E+01	9.84E-01	**3.06E − 03**
		0.1^2	1.01E+00	6.60E+01	9.75E-01	**6.10E − 03**
		1^2	1.59E+00	6.72E+01	1.76E+00	**1.22E + 00**
		5^2	4.92E+02	**4.85E + 02**	**4.85E + 02**	6.57E+02
	10	0.01^2	3.00E+02	3.65E+02	2.72E+02	**1.51E + 02**
		0.1^2	2.99E+02	3.65E+02	2.72E+02	**1.51E + 02**
		1^2	3.15E+02	3.83E+02	2.85E+02	**1.59E + 02**
		5^2	1.25E+03	1.12E+03	1.11E+03	**1.06E + 03**

variance are $\sigma^2 = 0.01^2, 0.1^2, 1, 5^2$ and the four different sample sizes are $n = 10, 15, 30, 100$. Therefore, we have 48 combinations of simulation settings. For each simulation setting, we generate observations and estimate all four estimators. We repeat this process 1,000 times and calculate mean squared errors (MSE) for each estimator.

For the small sample sizes, Table 4.1 and Table 4.2 show that our estimator is comparable to the others. In particular, for some high oscillation, our estimator performs better.

Table 4.3 also lists MSE when the sample sizes are large. The performance of the GSJ is better. Our estimator is comparable to the others.

5. Conclusion and further work

In this article, we derive the statistical properties of Tong and Wang's least square estimator for estimating the error variance in a nonparameric regression which satisfies Lipschitz condition and has a small sample size. None of the existing difference-based methods deals with some statistical properties in small sample cases. Under Lipschitz condition and a small sample size, we propose a new type of Tong and Wang's least square estimator which estimates the error variance as the intercept in a multiple regression which has some regressors highly correlated. To estimate the error variance with the problem of multicollinearity, the ridge regression can be employed and our estimator can be used in any sample size. The

Table 4.3 MSEs for $n = 30$ and $n = 100$

n	a	σ	$\widehat{\sigma}^2_{GSJ}$	$\widehat{\sigma}^2_{HKT}$	$\widehat{\sigma}^2_{TW}$	$\widehat{\sigma}^2_{RG}$
30	1	0.01^2	**7.30E − 08**	2.52E-02	9.29E-05	7.73E-05
		0.1^2	**1.94E − 05**	2.50E-02	1.13E-04	7.83E-05
		1^2	2.34E-01	1.58E-01	1.64E-01	**1.50E − 01**
		5^2	1.64E+02	1.54E+02	**1.49E + 02**	**1.49E + 02**
	4	0.01^2	**4.40E − 03**	5.96E+00	8.62E-02	7.92E-02
		0.1^2	**4.38E − 03**	5.98E+00	8.62E-02	7.92E-02
		1^2	**2.42E − 01**	6.27E+00	3.12E-01	2.67E-01
		5^2	1.64E+02	1.57E+02	1.50E+02	**1.49E + 02**
	10	0.01^2	**4.66E + 00**	1.47E+02	3.26E+01	3.22E+01
		0.1^2	**4.64E + 00**	1.46E+02	3.25E+01	3.21E+01
		1^2	**5.02E + 00**	1.46E+02	3.30E+01	3.26E+01
		5^2	**1.71E + 02**	4.27E+02	2.08E+02	2.09E+02
100	1	0.01^2	**7.97E − 10**	2.28E-04	7.28E-07	4.60E-08
		0.1^2	9.11E-06	2.36E-04	**6.62E − 0**	6.91E-06
		1^2	7.98E-02	5.17E-02	4.95E-02	**4.66E − 02**
		5^2	4.93E+01	**3.38E + 01**	3.49E+01	3.75E+01
	4	0.01^2	**2.84E − 07**	5.79E-02	5.04E-04	2.60E-05
		0.1^2	**9.66E − 06**	5.83E-02	5.31E-04	3.60E-05
		1^2	7.98E-02	1.04E-01	**4.86E − 0**	5.55E-02
		5^2	4.93E+01	**3.36E + 01**	3.51E+01	3.47E+01
	10	0.01^2	**4.14E − 04**	2.16E+00	1.83E-01	7.82E-03
		0.1^2	**4.34E − 04**	2.17E+00	1.84E-01	7.91E-03
		1^2	8.04E-02	2.21E+00	2.27E-01	**6.77E − 02**
		5^2	4.93E+01	3.55E+01	**3.52E + 01**	3.82E+01

results from Table 4.1 and Table 4.2 show that our estimator is better than others in some small sample cases and some conditions of unknown functions or comparable to others.

For future work, our approach will be developed based on the remedy of the multicollinearity problem and unequal space intervals in small sample problems.

Appendix

Proof of Theorem 3.2

Let $g(x) = ax^2 + bx + c$ for a, b, c constants and $x_i = i/n,\ i = 1, ..., n.$ Then we have the following by Lipschitz condition,

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

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

and

$$
\begin{aligned}
\beta d_k &= \frac{k^2}{2n^2(n-k)} \sum_{i=1+k}^{n} c_{i(k)}^2 \\
&= \frac{k^2}{2n^2(n-k)} \sum_{i=1+k}^{n} \left[a\frac{(2i-k)}{n} + b\right]^2 \\
&= \frac{k^2}{2n^2(n-k)} \sum_{i=1+k}^{n} \left[a^2\frac{(2i-k)^2}{n^2} + 2ab\frac{(2i-k)}{n} + b^2\right] \\
&= \frac{a^2k^2}{2n^4(n-k)} \sum_{i=1+k}^{n} (2i-k)^2 + \frac{abk^2(n+1)}{n^3} + \frac{b^2k^2}{2n^2}.
\end{aligned}
$$

When $g(x)$ has a bounded second derivative, we have the following;

$$
\begin{aligned}
& g(x_i) - g(x_{i-k}) \approx \left(\frac{i}{n} - \frac{i-k}{n}\right) g'(t) + \left\{\left(\frac{i}{n} - t\right)^2 - \left[\left(\frac{i-k}{n}\right) - t\right]^2\right\} \frac{g''(t)}{2} \\
\Leftrightarrow\ & g(x_i) - g(x_{i-k}) \approx \frac{k}{n}g'(t) + \left[\frac{i^2}{n^2} - \frac{(i-k)^2}{n^2} + \frac{i-k}{n}(2t) - \frac{i}{n}(2t)\right] \frac{g''(t)}{2} \\
\Leftrightarrow\ & g(x_i) - g(x_{i-k}) \approx \frac{k}{n}g'(t) + \frac{(2i-k)k}{n^2}\frac{g(t)''}{2} - \frac{k}{n}tg''(t), \\
& g(x_i) - g(x_{i-k}) = c_{i(k)}(x_i - x_{i-k})
\end{aligned}
$$

$$\Rightarrow \quad \frac{k}{n}g'(t_{i(k)}) + \frac{(2i-k)k}{n^2}\frac{g''(t_{i(k)})}{2} - \frac{k}{n}t_{i(k)}g''(t_{i(k)}) \approx c_{i(k)}\frac{k}{n}, \quad t_{i(k)} \in (x_{i-k},\ x_i)$$

$$\therefore c_{i(k)} \approx g'(t_{i(k)}) + \left(\frac{2i-k}{n} - 2t_{i(k)}\right)\frac{g''(t_{i(k)})}{2} \le a^*\frac{2i-k}{n} + b^*,$$

and

$$\begin{aligned}
\beta d_k &= \frac{k^2}{2n^2(n-k)}\sum_{i=1+k}^{n} c_{i(k)}^2 \\
&\approx \frac{k^2}{2n^2(n-k)}\sum_{i=1+k}^{n}\left[a_{i(k)}\frac{(2i-k)}{n} + b_{i(k)}\right]^2 \\
&\le \frac{k^2}{2n^2(n-k)}\sum_{i=1+k}^{n}\left[a^*\frac{(2i-k)}{n} + b^*\right]^2 \\
&= \beta_1\frac{k^2}{n^4(n-k)}\sum_{i=1+k}^{n}(2i-k)^2 + \beta_2\frac{k^2(n+1)}{n^3} + \beta_3\frac{k^2}{n^2},
\end{aligned}$$

where $a^* = \max\left\{a_{i(k)}{'s}\right\}$, $b^* = \max\left\{b_{i(k)}{'s}\right\}$ and β's are some constants.

References

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

Hoerl, A. E., and Kennard, R. W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics,* **12**, 55-67.

Hoerl, A. E., and Kennard, R. W. (1970b). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics,* **12**, 69-82.

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. (2011). On statistical properties of some difference-based error variance estimators in nonparametric regression with a finite sample. *Journal of the Korean Data & Information Science Society*, **22**, 575-587.

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**, 373-383.

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