Difference-based Variance Estimation in Nonparametric Regression with Repeated Measurements

Tiejun Tong[1], Yanyuan Ma[2], Wenlin Dai[1] and Lixing Zhu[1]

[1]Department of Mathematics, Hong Kong Baptist University, Hong Kong

[2]Department of Statistics, Texas A&M University, College Station, TX 77843, USA

Abstract

This work develop the difference-based estimators in the repeated measurements setting for nonparametric regression models. Three difference-based methods are proposed for the variance estimation under both balanced and unbalanced repeated measurements settings: the sample variance method, the partitioning method, and the sequencing method. Both their asymptotic properties and finite sample performance are explored. The sequencing method is shown to be the most adaptive while the sample variance method and the partitioning method are shown to outperform in certain cases. Finally, two real data examples are analyzed to demonstrate the practical use of the proposed methods.

KEY WORDS: Asymptotic normality; Difference-based estimator; Least squares; Nonparametric regression; Repeated measurements; Residual variance.

1 Introduction

Consider the nonparametric regression model with repeated measurements,

$$Y_{ij} = f(x_i) + \varepsilon_{ij}, \qquad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, m, \tag{1}$$

where Y_{ij} are observations, x_i are design points, f is an unknown mean function, and ε_{ij} are independent and identically distributed (i.i.d.) random errors with mean zero and variance σ^2. In this paper we are interested in estimating the residual variance σ^2. Needless to say, an accurate estimate of σ^2 is desired in many situations, e.g., in testing the goodness of fit and in deciding the amount of smoothing. Over the past three decades, interest in cheap yet competitive variance estimates in the nonparametric setting has grown tremendously. One family of estimators which has generated great

interest and has become an important tool for this purpose is the difference-based estimators. Unlike their residual-based counterparts, difference-based estimators do not require the estimation of the mean function, which involves nonparametric estimation procedures, and have therefore become quite popular in practice.

In the simple situation when $m = 1$, there already exist a large body of difference-based estimators in the literature (Dette, Munk & Wagner 1998). We note that very little attention has been paid to model (1) with $m \geq 2$, as pointed out in Xu & You (2007). Gasser, Sroka & Jennen-Steinmetz (1986) encountered the multiple measurements issue, but they decided to order the data sequentially and treat them as if they came from different design points. Thus, the multiple measurements feature is ignored. This is quite a pity, since intuitively the repeated measurements contain different type of information, and this new information should be taken into account in constructing estimators. We suspect that one reason very few work is available for treating multiple observations in difference based variance estimation literature is that it is not easy to combine the between-design-point difference and the within-design-point difference properly. In addition, even if a certain new treatment is proposed, it is not straightforward to analyze how effective this treatment is in theory. For example, it is difficult to know if the treatment has optimal large sample property, in other words, it is difficult to know if a better method can be found in treating the multiple measurements, either within the difference based method family or overall. In this paper, we will fill this literature gap in both aspects. Specifically, we will propose three new difference based methods to utilize the multiple measurements, respectively the sample variance method, the partitioning method and the sequencing method. We analyze these methods and illustrate the practical advantages of each method under different data structures and/or model assumptions. In addition, we will show that one of our proposals, the sequencing method is indeed optimal in that it is root-n consistent and it reaches the minimum asymptotic estimation variability among all possible consistent estimators, regardless whether the estimators are residual based, difference based or based on any other approaches. Thus, we will solve the multiple measurements problem thoroughly in this paper.

2 Main Results

Without loss of generality, we order the observations as $\{Y_{11}, \ldots, Y_{1m}, \ldots, Y_{n1}, \ldots, Y_{nm}\}$ and relabel the indices as $l = 1, 2, \ldots, nm$. With this notation, model (1) can be equivalently written as

$$Z_l = f(t_l) + \epsilon_l, \qquad l = 1, \ldots, nm, \tag{2}$$

where $\{Z_1, Z_2, \ldots, Z_{nm}\} = \{Y_{11}, \ldots, Y_{1m}, \ldots, Y_{n1}, \ldots, Y_{nm}\}$, $\{t_1, t_2, \ldots, t_{nm}\} = \{x_1, \ldots, x_1, \ldots, x_n, \ldots, x_n\}$, and $\{\epsilon_l, \epsilon_2, \ldots, \epsilon_{nm}\} = \{\varepsilon_{11}, \ldots, \varepsilon_{1m}, \ldots, \varepsilon_{n1}, \ldots, \varepsilon_{nm}\}$.

For model (2), we define the lag-p Rice estimator

$$\hat{\sigma}^2_{\mathrm{R}}(p) = \frac{1}{2(nm-p)} \sum_{l=p+1}^{nm} (Z_l - Z_{l-p})^2, \qquad \text{for } p = 1, \ldots, nm-1.$$

Note that the first m lag-p Rice estimators only use differences of the identical or consecutive design points, i.e., none of the $f(x_i)-f(x_{i-r})$ terms with $r \geq 2$ are involved in the first m lag-p Rice estimators. We thus combine them and define a new Rice-type estimator using the weighted average of the first m lag-p Rice estimators,

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{Rt}} &= \frac{1}{m^2n - m(m+1)/2}\sum_{p=1}^{m}(nm-p)\hat{\sigma}^2_{\mathrm{R}}(p) \\ &= \frac{1}{2m^2n - m(m+1)}\left\{\sum_{k=1}^{m-1}\sum_{i=1}^{n}\sum_{j=k+1}^{m}(Y_{ij}-Y_{i,j-k})^2 + \sum_{k=1}^{m}\sum_{i=2}^{n}\sum_{j=1}^{k}(Y_{ij}-Y_{i-1,m-k+j})^2\right\},\end{aligned}$$

where the weight for $\hat{\sigma}^2_{\mathrm{R}}(p)$ is assigned because the lag-p Rice estimator uses $(nm-p)$ pairs of data.

Some algebra yields

$$\begin{aligned}E(\hat{\sigma}^2_{\mathrm{Rt}}) &= \sigma^2 + \frac{1}{2m^2n - m(m+1)}\sum_{k=1}^{m}\sum_{i=2}^{n}\sum_{j=1}^{k}\{f(x_i)-f(x_{i-1})\}^2 \\ &= \sigma^2 + \frac{m(m+1)/2}{2m^2n - m(m+1)}\sum_{i=2}^{n}\{f(x_i)-f(x_{i-1})\}^2.\end{aligned}$$

This reveals that the Rice-type estimator $\hat{\sigma}^2_{\mathrm{Rt}}$ is always positively biased, unless f is a constant function. Suppose that f has a bounded first derivative. By the Taylor expansion we have

$$E(\hat{\sigma}^2_{\mathrm{Rt}}) = \sigma^2 + \frac{(n-1)m(m+1)}{n^2\{2m^2n - m(m+1)\}}J + o(\frac{1}{n^2}), \tag{3}$$

where $J = \int_0^1\{f'(x)\}^2dx/2$. To eliminate the bias term in (3), we further define the lag-r Rice-type estimators

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{Rt}}(r) &= \frac{1}{c_r}\sum_{p=(r-1)m+1}^{rm}(nm-p)\hat{\sigma}^2_{\mathrm{R}}(p) \qquad (4)\\ &= \frac{1}{2c_r}\left\{\sum_{k=1}^{m-1}\sum_{i=r}^{n}\sum_{j=k+1}^{m}(Y_{ij}-Y_{i-r+1,j-k})^2 + \sum_{k=1}^{m}\sum_{i=r+1}^{n}\sum_{j=1}^{k}(Y_{ij}-Y_{i-r,m-k+j})^2\right\},\end{aligned}$$

where $r = 1, 2, n-1$, and $c_r = \sum_{p=(r-1)m+1}^{rm}(nm-p) = m^2n - rm^2 + m(m-1)/2$. By definition, $\hat{\sigma}^2_{\mathrm{Rt}} = \hat{\sigma}^2_{\mathrm{Rt}}(1)$. Similar calculation at any fixed $r = o(n)$ yields

$$\begin{aligned}&E\left\{\hat{\sigma}^2_{\mathrm{Rt}}(r)\right\} \\ =\ &\sigma^2 + \frac{1}{2c_r}\left[\sum_{k=1}^{m-1}\sum_{i=r}^{n}\sum_{j=k+1}^{m}\{f(x_i)-f(x_{i-r+1})\}^2 + \sum_{k=1}^{m}\sum_{i=r+1}^{n}\sum_{j=1}^{k}\{f(x_i)-f(x_{i-r})\}^2\right] \\ =\ &\sigma^2 + Jd_r + o(r^2/n^2), \qquad (5)\end{aligned}$$

where

$$d_r = \frac{m\{(m-1)(n-r+1)(r-1)^2 + (m+1)(n-r)r^2\}}{2c_r n^2}. \tag{6}$$

The relation in (5) indicates that the lag-r Rice-type estimator $\hat{\sigma}^2_{\text{Rt}}(r)$ has a linear relationship with the quantity d_r. Taking advantage of this relation, we fit a linear regression model by treating $\hat{\sigma}^2_{\text{Rt}}(r)$ as the response variable and d_r as the covariate, and estimate σ^2 as the intercept of the linear model.

We choose the first b pairs of $\{d_r, \hat{\sigma}^2_{\text{Rt}}(r)\}$ to perform the regression, where $b = o(n)$. The choice of b will be discussed in Sections 2.3.2 and 2.3.3. In performing the linear regression estimation, because $\hat{\sigma}^2_{\text{Rt}}(r)$ involves c_r pairs of data, we assign weight $w_r = c_r/s_b$ to the rth observation, where $s_b = \sum_{r=1}^{b} c_r = m^2nb - m^2b(b+1)/2 + m(m-1)b/2$. The advantage of such weight assignment will be investigated in Section 2.3.5. We then minimize the weighted sum of squares $\sum_{r=1}^{b} w_r\{\hat{\sigma}^2_{\text{Rt}}(r) - \alpha - \beta d_r\}^2$ to fit the linear model

$$\hat{\sigma}^2_{\text{Rt}}(r) = \alpha + \beta d_r + e_r, \qquad r = 1, \ldots, b. \tag{7}$$

For ease of notation, let $\bar{\sigma}^2_w = \sum_{r=1}^{b} w_r \hat{\sigma}^2_{\text{Rt}}(r)$ and $\bar{d}_w = \sum_{r=1}^{b} w_r d_r$. Then the sequencing estimator of σ^2 is given as

$$\hat{\sigma}^2_3 = \hat{\alpha} = \bar{\sigma}^2_w - \hat{\beta}\bar{d}_w, \tag{8}$$

where $\hat{\beta} = \sum_{r=1}^{b} w_r \hat{\sigma}^2_{\text{Rt}}(r)(d_r - \bar{d}_w) / \sum_{r=1}^{b} w_r (d_r - \bar{d}_w)^2$ is the fitted slope. In Appendix 1 we prove that

Theorem 1. *For the equally spaced design, $\hat{\sigma}^2_3$ is an unbiased estimator of σ^2 when f is a linear function, regardless of the choice of b.*

In what follows we establish further statistical properties of the sequencing estimator $\hat{\sigma}^2_3$. For notational convenience, we let $\boldsymbol{Y} = (Y_{11}, \ldots, Y_{1m}, \ldots, Y_{n1}, \ldots, Y_{nm})^T$, $\boldsymbol{f} = \{f(x_1), \ldots, f(x_1), \ldots, f(x_n), \ldots, f(x_n)\}^T$, and $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \ldots, \varepsilon_{1m}, \ldots, \varepsilon_{n1}, \ldots, \varepsilon_{nm})^T$. Then $\boldsymbol{Y} = \boldsymbol{f} + \boldsymbol{\varepsilon}$. Also let $\boldsymbol{u} = (1, \ldots, 1)^T$, $\gamma_i = E(\varepsilon^i/\sigma^i)$ for $i = 3, 4$, and assume that $\gamma_4 > 1$.

2.0.1 Quadratic Form Representation

Let $\tau_0 = 0$ and $\tau_r = 1 - \bar{d}_w(d_r - \bar{d}_w) / \sum_{r=1}^{b} w_r (d_r - \bar{d}_w)^2$, $r = 1, \ldots, b$. By (8),

$$\hat{\sigma}^2_3 = \sum_{r=1}^{b} \tau_r w_r \hat{\sigma}^2_{\text{Rt}}(r) = \frac{1}{2s_b} \sum_{r=1}^{b} \left\{ \tau_r \sum_{p=(r-1)m+1}^{rm} \sum_{l=p+1}^{nm} (Z_l - Z_{l-p})^2 \right\}.$$

With some algebra, we can write $\hat{\sigma}^2_3$ as

$$\hat{\sigma}^2_3 = \frac{1}{2s_b} \boldsymbol{Y}^T \boldsymbol{D} \boldsymbol{Y},$$

where $\boldsymbol{D}$ is an $(nm) \times (nm)$ symmetric matrix with elements

$$\boldsymbol{D}_{ij} = \begin{cases} d_{ii}(a), & (a-1)m < i = j \le am \text{ with } a = 1, \ldots, n, \\ -\tau_a, & (a-1)m < |i-j| \le am \text{ with } a = 1, \ldots, b, \\ 0, & \text{otherwise,} \end{cases}$$

where $d_{ii}(a) = m\sum_{r=1}^{b}\tau_r + m\sum_{r=0}^{a-1}\tau_r + \{i-1-(a-1)m\}\tau_a$ for $a = 1, \ldots, b$; $d_{ii}(a) = 2m\sum_{r=1}^{b}\tau_r$ for $a = b+1, \ldots, n-b$; and $d_{ii}(a) = m\sum_{r=1}^{b}\tau_r + m\sum_{r=0}^{n-a}\tau_r + (am-i)\tau_{n+1-a}$ for $a = n-b+1, \ldots, n$.

Note that $\boldsymbol{D}$ depends on the design points only. By letting $f = 0$, we have

$$E(\hat{\sigma}_3^2) = \frac{1}{2s_b}E(\boldsymbol{Y}^T\boldsymbol{D}\boldsymbol{Y}) = \frac{1}{2s_b}E(\boldsymbol{\varepsilon}^T\boldsymbol{D}\boldsymbol{\varepsilon}) = \frac{\sigma^2}{2s_b}\mathrm{tr}(\boldsymbol{D}),$$

Now because of Theorem 1, $\hat{\sigma}_3^2$ is unbiased when $f = 0$, we have $\mathrm{tr}(\boldsymbol{D}) = 2s_b$. This shows that the proposed sequencing estimator possesses a quadratic form,

$$\hat{\sigma}_3^2 = \boldsymbol{Y}^T\boldsymbol{D}\boldsymbol{Y}/\mathrm{tr}(\boldsymbol{D}). \tag{9}$$

2.0.2 Asymptotic MSE and Optimal Bandwidth

The quadratic form representation (9) of σ_3^2 enables us to take advantage of the existing results in Dette et al. (1998) and directly obtain

$$\begin{aligned} \mathrm{MSE}(\hat{\sigma}_3^2) &= \big[(\boldsymbol{f}^T\boldsymbol{D}\boldsymbol{f})^2 + 4\sigma^2\boldsymbol{f}^T\boldsymbol{D}^2\boldsymbol{f} + 4\boldsymbol{f}^T\{\boldsymbol{D}\cdot\mathrm{diag}(\boldsymbol{D})\boldsymbol{u}\}\sigma^3\gamma_3 \\ &\quad + \sigma^4(\gamma_4 - 3)\mathrm{tr}[\mathrm{diag}(\boldsymbol{D})^2] + 2\sigma^4\mathrm{tr}(\boldsymbol{D}^2)\big]/\{\mathrm{tr}(\boldsymbol{D})\}^2, \end{aligned} \tag{10}$$

where $\mathrm{diag}(\boldsymbol{D})$ denotes the diagonal matrix of the diagonal elements of $\boldsymbol{D}$. The first term in (10) represents the squared bias, and the last four terms represent the variance term of the estimator. In the case when the random errors are normally distributed, $\gamma_3 = 0$ and $\gamma_4 = 3$ so that the third and fourth terms vanish.

Theorem 2. *Assume that f has a bounded second derivative. For the equally spaced design with $b \to \infty$ and $b/n \to 0$, we have*

$$\mathrm{Bias}(\hat{\sigma}_3^2) = O(b^3n^{-3}), \tag{11}$$

$$\mathrm{Var}(\hat{\sigma}_3^2) = \frac{\mathrm{Var}(\varepsilon^2)}{mn} + \frac{9\sigma^4}{4m^2nb} + \frac{9b\mathrm{Var}(\varepsilon^2)}{112mn^2} + o\{(nb)^{-1} + bn^{-2}\}, \tag{12}$$

$$\mathrm{MSE}(\hat{\sigma}_3^2) = \frac{\mathrm{Var}(\varepsilon^2)}{mn} + \frac{9\sigma^4}{4m^2nb} + \frac{9b\mathrm{Var}(\varepsilon^2)}{112mn^2} + o\{(nb)^{-1} + bn^{-2}\} + O(b^6n^{-6}). \tag{13}$$

Theorem 2 indicates that $\hat{\sigma}_3^2$ is a consistent estimator of σ^2, and its MSE reaches the asymptotically optimal rate (Dette et al. 1998). By (13), the asymptotically optimal bandwidth in terms of minimizing the MSE is given as

$$b_{opt} = \left\{\frac{28n\sigma^4}{m\mathrm{Var}(\varepsilon^2)}\right\}^{1/2}. \tag{14}$$

It is interesting to point out that b_{opt} does not depend on the mean function f. We also note that b_{opt} is a decreasing function of m. Substituting (14) into (13) leads to

$$\mathrm{MSE}\{\hat{\sigma}_3^2(b_{opt})\} = \frac{1}{nm}\mathrm{Var}(\varepsilon^2) + \frac{9\sqrt{7}}{28m^{3/2}n^{3/2}}\sigma^2\{\mathrm{Var}(\varepsilon^2)\}^{1/2} + o(1/n^{3/2}). \qquad (15)$$

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, 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.

Xu, Q. and You, J. (2007). Difference-based estimation for error variances in repeated measurement regression models, *Statistics & Probability letter* **77**: 811–816.