2014

Variance estimation in nonparametric regression with jump discontinuities

Wenlin Dai
Hong Kong Baptist University

Tiejun Tong
Hong Kong Baptist University, tongt@hkbu.edu.hk

Variance estimation in nonparametric regression with jump discontinuities

Wenlin Dai and Tiejun Tong*

Department of Mathematics, Hong Kong Baptist University, Hong Kong

September 2, 2013

Abstract

Variance estimation is an important topic in nonparametric regression. In this paper, we propose a pairwise regression method for estimating the residual variance. Specifically, we regress the squared difference between observations on the squared distance between design points, and then estimate the residual variance as the intercept. Unlike most existing difference-based estimators that require a smooth regression function, our method applies to regression models with jump discontinuities. Our method also applies to the situations where the design points are unequally spaced. Finally, we conduct extensive simulation studies to evaluate the finite-sample performance of the proposed method and compare it with some existing competitors.

Key words: Difference-based estimator; Jump point; Nonparametric regression; Non-uniform design; Pairwise regression; Residual variance.

*Corresponding author. E-mail: tongt@hkbu.edu.hk

1 Introduction

Consider a nonparametric regression model with jump discontinuities,

$$y_i = g(x_i) + h(x_i) + \varepsilon_i, \qquad 1 \le i \le n, \tag{1}$$

where y_i are observations, g is a continuous function, h is a step function, and ε_i are independent and identically distributed random errors with zero mean and variance σ^2. To be specific, we write the step function h as

$$h(x) = \sum_{j=1}^{p} c_j I(x > t_j),$$

where p is the number of jumps, $I(\cdot)$ is the identify function with value 1 when $x > t_j$ and value 0 otherwise, and c_j are the magnitudes of jumps at the jump points $t_j \in (0,1)$, respectively. Note that $g + h$ is the mean function.

Model (1) has wide applications in statistical process control (Qiu & Hawkins 2001), piecewise linear regression (Hinkley 1969, Brown, Durbin & Evans 1975, Kim & Siegmund 1989), image processing (McDonald & Owen 1986, Hall & Titterington 1992, Qiu 2005), and other related areas. It has also applied to many real data examples such as the Nile river discharge data (Cobb 1978), the stock market return data (Wang 1995), the sea-level pressure data (Qiu & Yandell 1998), and the infants growth data (Müller & Stadtmüller 1999). There is an abundant literature for analyzing model (1) including the detection and estimation of the number, positions, and magnitudes of jump points (Müller 1992, Wu & Chu 1993a, Wu & Chu 1993b, Eubank & Speckman 1994, Loader 1996).

This paper considers the estimation of the residual variance σ^2 in model (1). Needless to say, an accurate estimate of σ^2 is very important in regression models with jump discontinuities. Usually, one applies a two-step procedure to estimate σ^2 in such models. The first step is to estimate the positions of change points and then divide the mean function into several continuous sections accordingly. The second step is to estimate the residual variance within each individual section and then use them to make a final estimate of σ^2. Note that one may apply the residual-based methods (Hall & Marron 1990) or apply the difference-based methods (Müller 1992, Wu & Chu 1993a) to estimate the residual variance within each individual section.

Apart from the above, Müller & Stadtmüller (1999) proposed a single-step method for estimating σ^2 in model (1). Consider the equally spaced design where $x_i = i/n$, $i = 1, \ldots, n$. Let

$$z_k = \sum_{i=1}^{n-L} (y_{i+k} - y_i)^2 / [2(n-L)],$$

where $k = 1, \ldots, L$ with $L = L(n) \geq 1$. Under certain conditions on the mean function and the bandwidth L, Müller and Stadtmüller showed that

$$E(z_k) \approx \sigma^2 + \gamma l_k + \delta l_k^2, \tag{2}$$

where $l_k = k/(n-L)$, $\gamma = \sum_{j=1}^{p-1} (c_{j+1} - c_j)^2/2$ is the amount of discontinuity in the data, and $\delta = \int_0^1 [g'(x)]^2 dx/2 + \sum_{j=1}^{p-1} g'(t_{j+1})(c_{j+1} - c_j)$ is the measurement of the interaction between continuous and discontinuous parts. By (2), they fitted a quadratic regression that regresses z_k on l_k and then estimate the residual variance as the intercept. Specifically, they estimated σ^2 by

$$\hat{\sigma}^2_{\text{MS}} = \frac{3\sum_{k=1}^{L} (3L^2 + 3L + 2 - 6(2L+1)k + 10k^2) z_k}{2L(L-1)(L-2)}. \tag{3}$$

This method does not require an estimate of the positions of change points and is popular in practice.

Note that z_k only uses the first $n - L$ pairs of observations for performing the quadratic regression. Ignoring the last $L - k$ terms can make z_k a less efficient representation for σ^2, especially when $L-k$ is large. In addition, Müller & Stadtmüller (1999) required that $\min_{1 \leq i \leq p-1} (t_{i+1} - t_i) \geq 2L/N$ for the possibility of change-points separation. In the special case when $\gamma = 0$, i.e., when $h(x) = 0$, Tong, Ma & Wang

(2012) have demonstrated that the least squares estimator in Tong & Wang (2005) provides a smaller mean squared error (MSE) than $\hat{\sigma}^2_{\text{MS}}$. In addition, the equally spaced design condition in Müller & Stadtmüller (1999) is somewhat strong and has limited the practical use of $\hat{\sigma}^2_{\text{MS}}$.

In this paper, we propose a pairwise regression method for estimating σ^2 in model (1). Specifically, we regress the squared difference between observations on the squared distance between design points, and then estimate the residual variance as the intercept. Our method generalizes the existing methods from the following perspectives: (i) it does not require to estimate the positions of change points compared to the two-step estimators in the literature; (ii) it does not require to estimate the discontinuity parameter γ compared to the single-step estimator in Müller & Stadtmüller (1999); and (iii) it also applies to the settings where the design points are unequally spaced.

The remainder of the paper is organized as follows. In Section 2.1, we review the difference-based methods in estimating the residual variance in continuous nonparametric regression. In Section 2.2, we propose a pairwise regression method that extends the least squares estimator in Tong & Wang (2005) to unequally spaced designs. In Section 2.3, we further extend the proposed pairwise regression method to adaptively estimate the residual variance in nonparametric regression with jump discontinuities. In Section 3, we conduct extensive simulation studies to evaluate the finite-sample performance of the proposed method with some existing competitors. We then apply the proposed method to a real data example in Section 4 and conclude the paper in Section 5 with some discussions.

2 Main Results

2.1 Difference-based Estimators

In the special case when $h(x) = 0$, model (1) reduces to

$$y_i = g(x_i) + \varepsilon_i, \qquad 1 \le i \le n. \tag{4}$$

Under model (4), there are many difference-based methods in the literature for estimating σ^2. Assume that $0 \le x_1 \le \cdots \le x_n \le 1$. von Neumann (1941) and Rice (1984) proposed a first order difference-based estimator,

$$\hat{\sigma}^2_{\mathrm{R}} = \frac{1}{2(n-1)} \sum_{i=2}^{n} (y_i - y_{i-1})^2.$$

Gasser, Sroka & Jennen-Steinmetz (1986) and Hall, Kay & Titterington (1990) extended the idea and proposed some higher order difference-based estimators. In addition, Müller, Schick & Wefelmeyer (2003), Tong, Liu & Wang (2008) and Du & Schick (2009) proposed covariate-matched U-statistic estimators for the residual variance.

Apart form them, Tong & Wang (2005) and Tong et al. (2012) proposed a variation of the difference-based estimator in nonparametric regression. Let $x_i = i/n$ and $s_k = \sum_{i=k+1}^{n} (y_i - y_{i-k})^2/[2(n-k)]$. Suppose that g has a bounded first derivative. Tong & Wang (2005) showed that for any fixed $m = o(n)$,

$$E(s_k) \approx \sigma^2 + d_k J, \qquad k = 1, \dots, m, \tag{5}$$

where $d_k = k^2/n^2$ and $J = \int_0^1 [g'(x)]^2 dx/2$. By (5), they regressed s_k on d_k and then estimated the residual variance as the intercept. Specifically, their least squares estimator is given as

$$\hat{\sigma}^2_{\mathrm{TW}} = \sum_{k=1}^{m} w_k s_k - \hat{\beta}\bar{d}_w, \tag{6}$$

where $N_1 = mn - m(m+1)/2$, $w_k = (n-k)/N_1$, $\bar{d}_w = \sum_{k=1}^{m} w_k s_k$, and $\hat{\beta} = \sum_{k=1}^{m} w_k s_k (d_k - \bar{d}_w)/\sum_{k=1}^{m} w_k (d_k - \bar{d}_w)^2$.

Recall that $\hat{\sigma}^2_{\text{TW}}$ is developed under model (4) with a continuous mean function. When $h(x) \neq 0$, $\hat{\sigma}^2_{\text{TW}}$ may not perform well in model (1). To illustrate this, we consider the following regression model with a single jump at $t = 0.5$,

$$y_i = g(x_i) + cI(x_i > 0.5) + \varepsilon_i, \qquad c > 0. \tag{7}$$

Assume that J and c are both finite values. We have

$$\begin{aligned} E(s_k) &= \sigma^2 + \frac{1}{2(n-k)} \sum_{i=k+1}^{n} \{[g(x_i) + cI(x_i > 0.5)] \\ &\quad -[g(x_{i-k}) + cI(x_{i-k} > 0.5)]\}^2 \\ &= \sigma^2 + \left[d_k J + o(\frac{k^2}{n^2})\right] + \left[\frac{k}{n}c^2 + o(\frac{k}{n})\right]. \end{aligned}$$

Note that the bias owing to the jump, $(k/n)c^2$, dominates the bias owing to the continuous function, $d_k J = (k/n)^2 J$. This implies that $\hat{\sigma}^2_{\text{TW}}$ may suffer a severe bias for estimating σ^2, especially when c is large.

For a visualization of the bias pattern along with the c value, consider $g(x) = 5x(1-x)$ and $h(x) = cI(x > 0.5)$ with $0 < c < 20$. We let $n = 100$, $m = 10$ and $\sigma^2 = 1$ throughout the simulations. The estimated variance against the c value is plotted in Figure 1. We observe that $\hat{\sigma}^2_{\text{TW}}$ increases rapidly as c increases. As a consequence, $\hat{\sigma}^2_{\text{TW}}$ does not provide a satisfactory performance in this example.

2.2 Pairwise Regression

Recall that the least squares estimator in Tong & Wang (2005) only applies to the equally spaced design. This has largely restricted the usage of their method in practice. In this section, we introduce a pairwise regression method for estimating the residual variance that extends the least squares estimator from the equally spaced design to unequally spaced designs.

Let $s_{ij} = (y_j - y_i)^2/2$ be the half squared differences and $d_{ij} = (x_j - x_i)^2$ be the corresponding squared distances for any $1 \leq i < j \leq n$. Let $d = o(1)$ be the

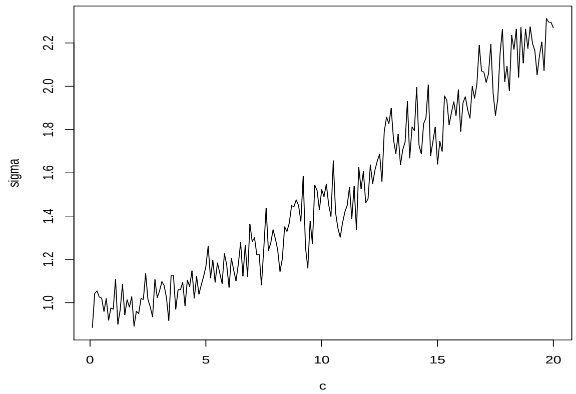

Figure 1: The estimated $\hat{\sigma}^2$ corresponding to different c values.

bandwidth. We collect all d_{ij} values that satisfy $d_{ij} \leq d$. For ease of notation, let $A = \{(i,j) : d_{ij} \leq d,\ 1 \leq i < j \leq n\}$ and $N = \#(A)$ be the total number of pairs in A. Correspondingly, we collect the s_{ij} values for all $(i,j) \in A$.

Note that $E(s_{ij}) = \sigma^2 + (g(x_j) - g(x_i))^2/2$. When g is a linear function with slope ψ, we have $E(s_{ij}) = \sigma^2 + d_{ij}\psi^2/2$. In view of this, for the paired data $\{(d_{ij}, s_{ij}) : (i,j) \in A\}$ with $d = o(1)$, we fit a simple linear regression model that regresses s_{ij} directly on d_{ij},

$$s_{ij} = \alpha + d_{ij}\beta + \eta_{ij}. \tag{8}$$

We then use the ordinary least squares method to estimate σ^2 using the fitted intercept. This leads to

$$\hat{\sigma}^2 = \hat{\alpha} = \frac{\sum_A (S_2 - S_1 d_{ij}) s_{ij}}{N S_2 - S_1^2}, \tag{9}$$

where $S_1 = \sum_A d_{ij}$ and $S_2 = \sum_A d_{ij}^2$. We refer to (9) as a pairwise regression estimator.

Let $c_{ij} = S_2 - S_1 d_{ij}$ and $y = (y_1, \ldots, y_n)^T$. The estimator (9) has a quadratic form $\hat{\sigma}^2 = y^T M y / \mathrm{tr}(M)$, where M is an $n \times n$ symmetric matrix with upper triangular elements

$$m_{ij} = \begin{cases} \sum_{(i,j)\in A_k} c_{ij}/2 & 1 \le i = j = k \le n \\ -c_{ij}/2 & (i,j) \in A \\ 0 & \text{otherwise} \end{cases}$$

where $A_k = \{(i,j) : i = k \text{ or } j = k,\ (i,j) \in A\}$ with $k = 1, 2, \ldots, n$.

In what follows we draw some connection between the pairwise regression estimator $\hat{\sigma}^2$ and the least squares estimator $\hat{\sigma}^2_{\mathrm{TW}}$. Let $x_i = i/n$ and $d = m^2/n^2$. Then $N = N_1 = mn - m(m+1)/2$ and $d_{ij} = d_{j-i}$. Also, it is easy to verify that $S_1 = N\bar{d}_w$, $S_2 = N\sum_{k=1}^m w_k d_k^2$, $\sum_A s_{ij} = N\sum_{k=1}^m w_k s_k$, and $\sum_A d_{ij}s_{ij} = N\sum_{k=1}^m w_k d_k s_k$. With the above equalities, we have

$$\begin{aligned} \hat{\sigma}^2 &= \frac{S_2\sum_A s_{ij} - S_1\sum_A d_{ij}s_{ij}}{NS_2 - S_1^2} \\ &= \frac{\sum_{k=1}^m w_k d_k^2 \sum_{k=1}^m w_k s_k - \bar{d}_w \sum_{k=1}^m w_k d_k s_k}{\sum_{k=1}^m w_k d_k^2 - \bar{d}_w^2} \\ &= \sum_{k=1}^m w_k s_k - \bar{d}_w \hat{\beta} \\ &= \hat{\sigma}^2_{\mathrm{TW}}. \end{aligned}$$

This shows that when the design points are equally spaced, $\hat{\sigma}^2$ and $\hat{\sigma}^2_{\mathrm{TW}}$ are equivalent to each other. From this point of view, we conclude that the pairwise regression estimator (9) generalized the least squares estimator $\hat{\sigma}^2_{\mathrm{TW}}$ from the equally spaced design to a general design.

2.3 Adaptive Pairwise Regression

As mentioned, most existing difference-based estimators were developed under model (4). In this section, we show that the pairwise regression method in Section 2.2 can be readily extended to model (1) with jump discontinuities.

To apply the pairwise regression to models with jump discontinuities, we revisit the simple regression model presented in (7). Let $O = \{(i,j) : 0.5 \in (x_i, x_j]\}$ be the pairs of design points that cross the jump point. By (9), we have

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{\sum_{A\backslash O}(S_2 - S_1 d_{ij})E(s_{ij})}{NS_2 - S_1^2} + \frac{\sum_O (S_2 - S_1 d_{ij})E(s_{ij})}{NS_2 - S_1^2} \\ &= \frac{\sum_{A\backslash O}(S_2 - S_1 d_{ij})(\sigma^2 + O(m^2/n^2))}{NS_2 - S_1^2} + \frac{\sum_O (S_2 - S_1 d_{ij})(\sigma^2 + c^2/2 + O(m/n))}{NS_2 - S_1^2} \\ &= \frac{\sum_A (S_2 - S_1 d_{ij})}{NS_2 - S_1^2}\sigma^2 + \frac{\sum_O (S_2 - S_1 d_{ij})}{NS_2 - S_1^2}\left(\frac{c^2}{2} + O(\frac{m}{n})\right) \\ &= \sigma^2 + \frac{\sum_O (S_2 - S_1 d_{ij})}{NS_2 - S_1^2}\frac{c^2}{2} + O(\frac{m^2}{n^2}), \end{aligned} \tag{10}$$

where

$$\frac{\sum_A (S_2 - S_1 d_{ij})}{NS_2 - S_1^2} = 1 \quad \text{and} \quad \frac{\sum_O (S_2 - S_1 d_{ij})}{NS_2 - S_1^2} = O(\frac{m}{n}).$$

By (10), to obtain a good estimate of σ^2, it is clear that the pairs in O should be excluded from the regression to eliminate the bias. Otherwise, given that the quantity c is large, the extra bias introduced by the jump can be very severe.

In what follows, we examine how excluding the pairs in O takes effect on the MSE of the estimator. We will also suggest ways to exclude certain pairs of data from the pairwise regression. Let $z_{ij} = y_j - y_i$ for any $1 \leq i < j \leq n$. For $d = o(1)$, we have $E(z_{ij}) \to 0$ for $(i,j) \in O$ and $E(z_{ij}) \to c$ for $(i,j) \in A\backslash O$. Whereas for any $(i,j) \in A$, $\mathrm{var}(z_{ij}) = 2\sigma^2$. To visualize the discrepancy between the two groups of z_{ij}, we consider $c = 0$, 2 and 5 for the example in Section 2.1. All other settings are kept the same as before except that now $\sigma = 0.5$.

We plot the histograms of the simulated z_{ij} values in the first column of Figure 2, respectively. When the mean function is continuous (i.e., $c = 0$), the histogram is unimodal and almost symmetric around zero. When c increases, the histogram tends to be right-skewed and eventually separates to two disjoint sections, one consisting of the pairs without jump and the other consisting of the pairs with jump. To eliminate the impact of the jump on the variance estimation, we can treat the extremely large $|z_{ij}|$ values, or correspondingly the extremely large s_{ij} values, as outliers and exclude them in the pairwise regression.

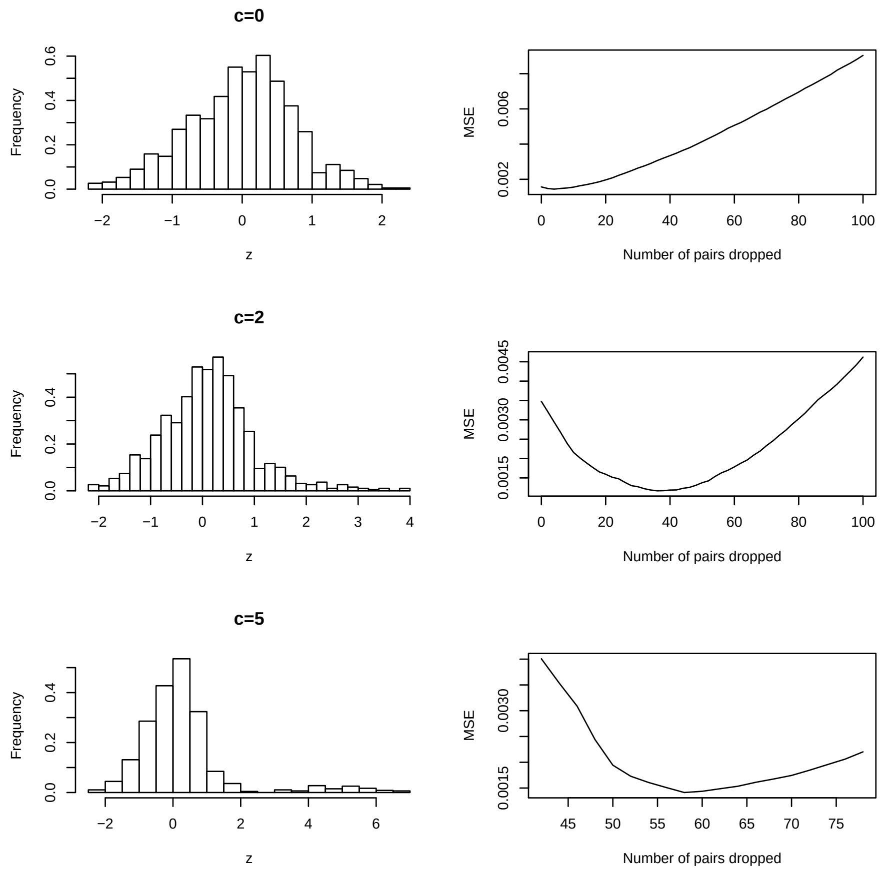

Figure 2: The histogram of z_{ij} and the change of MSE against the number of pairs dropped, where $c = 0$, 2 and 5, respectively.

Ideally, none of the z_{ij} values should be detected as outliers when the mean function is continuous. When c is non zero, to reduce the bias or essentially to gain a small MSE we may wish to drop the pairs that across the jump point. As an illustration, we also plot in Figure 2 the simulated MSE against the number of pairs dropped for the three c values, respectively. It suggests to drop few pairs for $c = 0$, drop around 40 pairs for $c = 2$, and drop around 55 pairs for $c = 5$, for estimating the residual variance with a minimum MSE. Finally, it is interesting to point out that for an equally spaced design with $m = 10$, there is a total of $m(m + 1)/2 = 55$ pairs across the jump point.

In what follows, we suggest two practical rules that identify certain z_{ij} values as outliers and then exclude them from the pairwise regression. The resulting methods are referred to as adaptive pairwise regression estimators.

2.3.1 Box Plot Method

The first method uses the box plot to detect certain z_{ij} values as outliers. Let $\mathrm{Q_L}(\{z_{ij}\})$ and $\mathrm{Q_U}(\{z_{ij}\})$ denote the lower quartile and the upper quartile of the observed z_{ij} values within the bandwidth, respectively. Follow the form of Sim, Gan & Chang (1994), we define $\mathrm{LB} = \mathrm{Q_L}(\{z_{ij}\}) - C \cdot \mathrm{IQR}$ and $\mathrm{UB} = \mathrm{Q_U}(\{z_{ij}\}) + C \cdot \mathrm{IQR}$, where $\mathrm{IQR} = \mathrm{Q_U}(\{z_{ij}\}) - \mathrm{Q_L}(\{z_{ij}\})$ is the interquartile range and C is an adjustment factor. Here, we assign a value of 2 or 3 to C. We then identify z_{ij} as an outlier if $z_{ij} \in (-\infty, \mathrm{LB})$ or $z_{ij} \in (\mathrm{UB}, \infty)$. We refer to the estimator by the box plot method as $\hat{\sigma}^2_{\mathrm{box}}$.

2.3.2 Cross-Validation Method

Note that the bandwidth d is also critical to the variance estimation. Our second method uses a V-fold cross-validation (CV) approach to simultaneously choose the bandwidth d and the adjustment factor C. Specifically, we first split the whole data set into V disjoint subsamples, S_1, ..., S_V as in Tong & Wang (2005). Second, for given d and C, we estimate σ^2 by $\hat{\sigma}_v^2(d, C)$ based on the subsample $\cup_{i \neq v} S_i$ and the pairs with $d_{ij} \leq d$ and $z_{ij} \in [\text{LB}(C), \text{UB}(C)]$. Finally, we choose the optimal tuning parameters d and C that minimize

$$\text{CV}(d, C) = \sum_{v=1}^{V} \left[\hat{\sigma}^2(d, C) - \hat{\sigma}_v^2(d, C)\right]^2,$$

where $\hat{\sigma}^2(d, C)$ is the estimate of σ^2 based on the whole data set with pairs $d_{ij} \leq d$ and $z_{ij} \in [\text{LB}(C), \text{UB}(C)]$. We refer to the estimator by the CV method as $\hat{\sigma}^2_{\text{CV}}$.

3 Simulations

In this section, we conduct extensive simulation studies to evaluate the finite-sample performance of the proposed estimators and compare them with some existing competitors.

3.1 Equidistant Design

The first study assumes an equally spaced design. Specifically, let $x_i = i/n$ with $i = 1, \ldots, n$. We consider the following four estimators for comparison: $\hat{\sigma}^2_{\text{box}}$, $\hat{\sigma}^2_{\text{CV}}$, $\hat{\sigma}^2_{\text{MS}}$ and $\hat{\sigma}^2_{\text{TW}}$. We consider a total of 9 mean functions with combinations $g_i + h_j$ from the following functions:

$$g_1(x) = 5x(1-x),$$
$$g_2(x) = 5\exp(1-x),$$
$$g_3(x) = 5\sin(2x),$$

and

$$h_1(x) = 4I(x > \sqrt{2}/2),$$
$$h_2(x) = 3I(x > \sqrt{2}/4) + 4I(x > \sqrt{2}/2),$$
$$h_3(x) = 0.$$

For each mean function, we consider $n = 30$, 100 and 500, ranging from small to large sample sizes respectively, and $\sigma = 0.2$, 0.5, 1, 2 and 5, ranging from small to large variances respectively. Finally, for given n and σ, we simulate the random errors ε_i independently from $N(0, \sigma^2)$.

For each simulation setting, we generate observations and compute the estimators $\hat{\sigma}^2_{\text{TW}}(m)$, $\hat{\sigma}^2_{\text{MS}}(L)$, $\hat{\sigma}^2_{\text{box}}(d, C)$ and $\hat{\sigma}^2_{\text{cv}}$. Note that the bandwidth L in Müller & Stadtmüller (1999) is not very sensitive to the estimation of σ^2. We consider both $L_s = m_s = n^{1/2}$ and $L_t = m_t = n^{1/3}$ as in Tong & Wang (2005). This leads to the corresponding d values as $d_s = (m_s/n)^2$ and $d_t = (m_t/n)^2$. Then together with $C = 2$ and 3, we have 4 different estimates for $\hat{\sigma}^2_{\text{box}}$. Recall that the CV estimator, $\hat{\sigma}^2_{\text{cv}}$, aims to figure out the best combination between d and C. We consider leave-one-out CV for $n = 30$, and 10-fold CV for $n = 100$ and $n = 500$, throughout the simulations.

We repeat the process 1000 times and compute the following relative MSEs, MSE/MSE$_{opt}$, for each method. Here, $\text{MSE}_{opt} = n^{-1}(\gamma_4 - 1)\sigma^4$ is specified as the optimal efficiency bound of all root-n consistent estimators of σ^2, and $\gamma_4 = E(\varepsilon^4)/\sigma^4$. For normal errors, we have $\gamma_4 = 3$ and $\text{MSE}_{opt} = 2\sigma^4/n$. We observe that negative estimates indicated by Tong & Wang (2005) and Müller & Stadtmüller (1999) do appear in certain simulations, though very rarely. We replace the negative estimates with zero when calculating the relative MSEs.

3.2 Non-equidistant Design

This section carries out simulation studies for unequally spaced designs. We generate design points from the beta distribution Beta$(3, 3)$. This is a bell shaped distribution on [0,1] with a mode at 0.5. Also for simplicity, we consider only three mean functions $g_3 + h_i$, where the first two functions are discontinuous and the last one is continuous. All other settings are kept the same as those in Section 3.1.

Finally, recall that Müller & Stadtmüller (1999) and Tong & Wang (2005) do not apply to unequally spaced designs. We thus omit both the estimators but add in the pairwise regression estimator $\hat{\sigma}^2$ in (9) for comparison. Then correspondingly, we compute the relative MSEs for $\hat{\sigma}^2$, $\hat{\sigma}^2_{\text{box}}(d, C)$, and $\hat{\sigma}^2_{\text{cv}}$, respectively.

3.3 Simulation Results

Tables 1 − 6 list the relative MSEs for the mean functions with jump points, respectively, under the equidistant design. In general, we observe that $\text{MSE}(\hat{\sigma}^2_{\text{CV}}) \doteq \text{MSE}(\hat{\sigma}^2_{\text{box}}) < \text{MSE}(\hat{\sigma}^2_{\text{MS}}) < \text{MSE}(\hat{\sigma}^2_{\text{TW}})$ for small and moderate σ values, and $\text{MSE}(\hat{\sigma}^2_{\text{CV}}) \doteq \text{MSE}(\hat{\sigma}^2_{\text{box}}) \doteq \text{MSE}(\hat{\sigma}^2_{\text{TW}}) < \text{MSE}(\hat{\sigma}^2_{\text{MS}})$ for large σ values. These results show that the proposed adaptive estimators outperform the existing estimators in the presence of jump discontinuities. We also observe that the comparative performance of $\hat{\sigma}^2_{\text{box}}(d_t, 2)$, $\hat{\sigma}^2_{\text{box}}(d_t, 3)$, $\hat{\sigma}^2_{\text{box}}(d_s, 2)$ and $\hat{\sigma}^2_{\text{box}}(d_s, 3)$ depends on the smoothness and continuity of the mean function, the sample size and the signal-to-noise ratio. As reported in Tong & Wang (2005), $\hat{\sigma}^2_{\text{box}}(d_s, \cdot)$ may not perform well when the sample size is small. As a compromise, $\hat{\sigma}^2_{\text{CV}}$ performs well in most settings.

In contrast, we list in Tables 7 − 9 the relative MSEs for the continuous mean functions $f_7(x)$ through $f_9(x)$, under the equidistant design. We observe that $\hat{\sigma}^2_{\text{box}}$, $\hat{\sigma}^2_{\text{CV}}$ and $\hat{\sigma}^2_{\text{TW}}$ perform very similar under various settings. More specifically, we observe that for a continuous mean function, very few z_{ij} values were detected from simulations as outliers. As a consequence, both $\hat{\sigma}^2_{\text{box}}(d_t, 2)$ and $\hat{\sigma}^2_{\text{box}}(d_t, 3)$ perform essentially the same as $\hat{\sigma}^2_{\text{TW}}(m_t)$, and both $\hat{\sigma}^2_{\text{box}}(d_s, 2)$ and $\hat{\sigma}^2_{\text{box}}(d_s, 3)$ perform essentially the same as $\hat{\sigma}^2_{\text{TW}}(m_s)$. Apart from them, $\hat{\sigma}^2_{\text{MS}}$ does not provide a comparable performance. This coincides the observation in Tong et al. (2012) that $\hat{\sigma}^2_{\text{MS}}$ is worse than $\hat{\sigma}^2_{\text{TW}}$ when the mean function is continuous.

Finally, we list in Tables 10 − 12 the relative MSEs for the settings with non-equidistant designs. Similarly as above, we observe that $\hat{\sigma}^2_{\text{box}}$ perform better than $\hat{\sigma}^2$ in the presence of jump discontinuities, and their performance are similar when the mean function is continuous. Meanwhile, $\hat{\sigma}^2_{\text{CV}}$ performs very well in most settings, especially when the sample size is small.

4 Case Study

For illustration, we apply the proposed methods to a real data example. The data were reported in Cobb (1978) on the annual volume of discharge in the Nile River from 1895 to 1934. In Figure 3, we find several observations with large variation and we suspect that the mean function might contain jump discontinuities. For this data with $n = 40$ observations, we choose $L_t = m_t = \lfloor n^{1/3} \rfloor = 3$ and $L_s = m_s = \lfloor n^{1/2} \rfloor = 6$ for $\hat{\sigma}^2_{\text{MS}}$ and $\hat{\sigma}^2_{\text{TW}}$, respectively. Here, $\lfloor a \rfloor$ denotes the largest integer smaller than or equal to a. For the proposed methods, correspondingly we choose $d_t = (m_t/n)^2 = 0.075^2$ and $d_s = (m_s/n)^2 = 0.15^2$. The estimated residual variances are as follows: $\hat{\sigma}^2_{\text{MS}}(L_t) = 126.9$, $\hat{\sigma}^2_{\text{MS}}(L_s) = 47.7$; $\hat{\sigma}^2_{\text{TW}}(m_t) = 119.9$ and $\hat{\sigma}^2_{\text{TW}}(m_s) = 144.8$; $\hat{\sigma}^2_{\text{box}}(d_t, 2) = 126.1$, $\hat{\sigma}^2_{\text{box}}(d_t, 3) = 119.9$, $\hat{\sigma}^2_{\text{box}}(d_s, 2) = 137.2$, $\hat{\sigma}^2_{\text{box}}(d_s, 3) = 144.8$ and $\hat{\sigma}^2_{\text{CV}} = 119.9$. We note that for a standard with $C = 3$, no outliers were identified so that $\hat{\sigma}^2_{\text{box}}(d_t, 3) = \hat{\sigma}^2_{\text{TW}}(m_t) = 119.9$ and $\hat{\sigma}^2_{\text{box}}(d_s, 3) = \hat{\sigma}^2_{\text{TW}}(m_s) = 144.8$. In addition, the cross validation method suggests to take $C = 3$ with a bandwidth at d_t and that results in the variance estimate as 119.9. Recall that the suggested value of σ^2 is 125 in Cobb (1978). We conclude that our pairwise regression method performs at least as well as the least squares estimator σ^2_{TW}. Nevertheless, the estimator σ^2_{MS} is very sensitive to the choice of the bandwidth and so is less reliable.

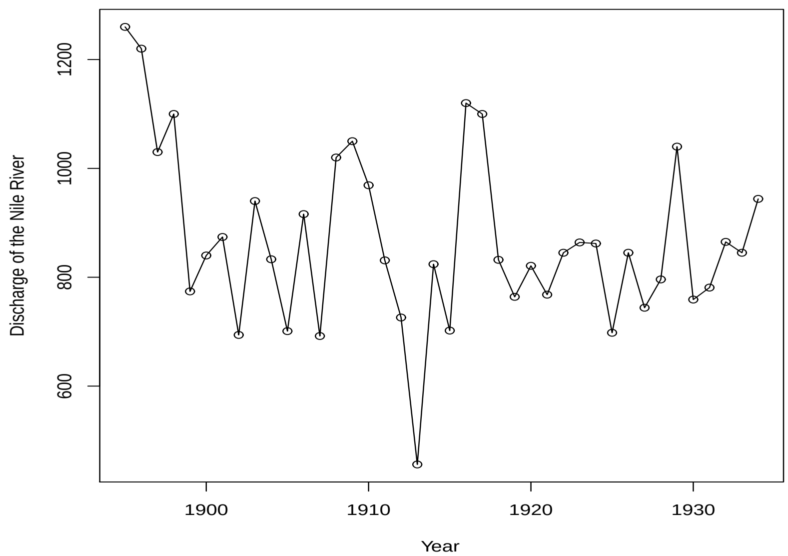

Figure 3: The Nile discharge data from 1895 to 1934.

5 Discussion

In this paper, we first introduced a pairwise regression method for estimating σ^2 in nonparametric regression models with continuous mean function. We further extended the pairwise regression method to model (1) with jump discontinuities via adaptation. As already mentioned in Section 1, the proposed adaptive method generalizes the existing methods from different points of view and has several important merits. In particular, our adaptive method turns out to be superior for its flexibility in eliminating the effect of potential jumps in the mean function and for its applicability in both equally and unequally design settings. In addition, compared with the residual-based estimators, our method provides a direct way to estimate the residual variance without the estimations of mean function and jump points. In conclusion, we recommend to use the estimator $\hat{\sigma}^2_{\mathrm{CV}}$ in practice. More work, though, is needed for demonstrating the theoretical results of our proposed estimators.

The proposed method can be readily extended to higher-dimensional regression models. Consider, for instance, the following bivariate nonparametric regression model with jump discontinuities,

$$y_i = g^*(x_{1i}, x_{2i}) + h^*(x_{1i}, x_{2i}) + \varepsilon_i, \qquad 1 \le i \le n,$$

We can define d_{ij} by $d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2}$ or by $d_{ij} = |x_{1i} - x_{1j}| + |x_{1i} - x_{1j}|$, and then proceed the estimation similarly as in Sections 2.2 and 2.3. Further research is necessary to investigate the practical rules for the corresponding adaptive method as well as to evaluate its finite-sample performance. Further, recall that the proposed method in this paper is restricted to a constant residual variance assumption. As this may not be realistic in applications, it should be of interest to propose new pairwise regression methods for estimating the variance function in regression models with jump discontinuities.

Acknowledgement

This research was supported by Hong Kong RGC grant HKBU202711, and Hong Kong Baptist University grants FRG2/10-11/020 and FRG2/11-12/110. The authors thank the editor, the associate editor, and two referees for their helpful comments and suggestions that have substantially improved the paper.

Table 1: Relative MSEs of various estimators for the mean function $f_1(x) = g_1(x) + h_1(x)$, under equidistant design.

		$\hat{\sigma}^2_{\text{MS}}$		$\hat{\sigma}^2_{\text{TW}}$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	L_t	L_s	m_t	m_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	65.5	34.6	513	1368	1.64	1.60	1.67	1.79	1.62
	0.5	20.8	9.58	16.8	39.0	2.41	7.78	6.12	24.4	2.82
	1	13.2	4.94	2.87	4.15	2.48	2.83	3.40	4.13	2.87
	2	11.0	3.61	1.66	1.54	1.63	1.66	1.49	1.53	1.47
	5	10.3	3.23	1.49	1.26	1.51	1.50	1.25	1.26	1.30
100	0.2	14.3	12.0	308	939	1.37	1.38	1.22	1.21	1.25
	0.5	5.38	3.76	9.75	26.0	1.41	2.80	1.39	5.22	1.33
	1	4.05	2.52	2.02	2.92	1.71	1.98	1.99	2.82	1.59
	2	3.73	2.19	1.43	1.34	1.37	1.44	1.28	1.34	1.26
	5	3.64	2.10	1.36	1.20	1.35	1.36	1.18	1.20	1.19
500	0.2	3.85	3.92	130	777	1.24	1.22	1.11	1.11	1.17
	0.5	2.36	1.79	4.46	20.8	1.25	1.43	1.15	1.92	1.20
	1	2.15	1.47	1.41	2.29	1.25	1.38	1.35	2.13	1.32
	2	2.10	1.38	1.23	1.17	1.22	1.23	1.11	1.17	1.13
	5	2.08	1.36	1.22	1.10	1.24	1.22	1.13	1.11	1.16

Table 2: Relative MSEs of various estimators for the mean function $f_2(x) = g_2(x) + h_1(x)$, under equidistant design.

		$\hat{\sigma}^2_{\text{MS}}$		$\hat{\sigma}^2_{\text{TW}}$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	L_t	L_s	m_t	m_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	61.1	23.2	526	1509	1.78	1.75	3.49	15.2	2.71
	0.5	20.8	9.40	17.1	42.6	1.90	5.64	4.50	26.7	2.66
	1	13.2	5.05	2.89	4.38	2.35	2.80	3.28	4.28	3.04
	2	11.0	3.65	1.66	1.55	1.64	1.66	1.49	1.55	1.53
	5	10.3	3.24	1.49	1.26	1.51	1.50	1.26	1.26	1.27
100	0.2	14.2	12.3	309	961	1.39	1.40	1.28	1.27	1.37
	0.5	5.36	3.81	9.81	26.5	1.39	2.29	1.25	3.73	1.28
	1	4.04	2.52	2.02	2.95	1.67	1.95	1.86	2.79	1.58
	2	6.72	2.20	1.43	1.34	1.40	1.44	1.28	1.34	1.28
	5	3.64	2.10	1.36	1.20	1.36	1.37	1.18	1.20	1.20
500	0.2	3.84	3.94	130	780	1.24	1.22	1.12	1.13	1.18
	0.5	2.36	1.79	4.47	20.9	1.24	1.39	1.14	1.62	1.22
	1	2.15	1.47	1.41	2.30	1.25	1.38	1.32	2.10	1.32
	2	2.10	1.38	1.23	1.17	1.22	1.23	1.11	1.17	1.13
	5	2.08	1.36	1.22	1.11	1.24	1.22	1.13	1.11	1.16

Table 3: Relative MSEs of various estimators for the mean function $f_3(x) = g_3(x) + h_1(x)$, under equidistant design.

		$\hat{\sigma}^2_{\text{MS}}$		$\hat{\sigma}^2_{\text{TW}}$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	L_t	L_s	m_t	m_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	61.4	22.7	506	1297	1.70	1.95	76.1	1427	1.69
	0.5	20.6	9.06	16.6	37.2	3.91	11.7	24.7	37.0	4.51
	1	13.2	4.90	2.86	4.05	2.60	2.83	3.81	4.05	2.71
	2	11.0	3.61	1.66	1.53	1.64	1.66	1.50	1.53	1.50
	5	10.3	3.23	1.49	1.26	1.51	1.50	1.26	1.26	1.30
100	0.2	14.3	12.0	306	927	1.39	1.38	1.28	1.27	1.35
	0.5	5.39	3.78	9.72	25.7	1.45	3.34	1.81	9.95	1.41
	1	4.05	2.52	2.01	2.90	1.72	1.98	2.15	2.86	1.60
	2	3.73	2.20	1.43	1.34	1.40	1.44	1.29	1.34	1.26
	5	3.64	2.10	1.36	1.20	1.35	1.36	1.18	1.20	1.19
500	0.2	3.85	3.93	130	775	1.23	1.22	1.11	1.12	1.17
	0.5	2.36	1.79	4.46	20.7	1.25	1.44	1.16	2.33	1.23
	1	2.15	1.47	1.41	2.29	1.25	1.38	1.38	2.15	1.33
	2	2.10	1.38	1.23	1.17	1.22	1.23	1.11	1.17	1.13
	5	2.08	1.36	1.22	1.11	1.24	1.22	1.13	1.11	1.16

Table 4: Relative MSEs of various estimators for the mean function $f_4(x) = g_1(x) + h_2(x)$, under equidistant design.

		$\hat{\sigma}^2_{\text{MS}}$		$\hat{\sigma}^2_{\text{TW}}$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	L_t	L_s	m_t	m_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	88.2	47.3	1217	3177	1.69	5.45	3177	3177	1.69
	0.5	24.6	12.3	36.1	86.7	9.65	24.2	86.7	86.7	18.3
	1	14.3	5.80	4.39	7.48	3.88	4.30	7.41	7.48	4.52
	2	11.3	3.85	1.84	1.84	1.79	1.84	1.81	1.84	1.71
	5	10.3	3.28	1.52	1.28	1.53	1.52	1.28	1.29	1.30
100	0.2	20.3	16.0	309	2274	1.35	1.15	1.23	1.28	1.29
	0.5	6.51	4.62	9.75	60.8	1.37	1.18	4.37	29.0	2.09
	1	1.33	2.76	1.94	5.21	1.34	1.33	4.00	5.17	2.24
	2	3.80	2.27	1.34	1.51	1.36	1.17	1.44	1.51	1.34
	5	3.65	2.12	1.32	1.20	1.38	1.22	1.19	1.20	1.19
500	0.2	4.82	5.52	318	1902	1.24	1.22	1.13	1.13	1.17
	0.5	2.51	2.08	9.37	49.8	1.28	2.10	1.39	7.98	1.37
	1	2.18	1.56	1.74	4.13	1.34	1.68	2.23	3.92	1.56
	2	2.10	1.41	1.26	1.29	1.23	1.26	1.18	1.29	1.18
	5	2.08	1.36	1.22	1.11	1.24	1.22	1.12	1.11	1.16

Table 5: Relative MSEs of various estimators for the mean function $f_5(x) = g_2(x) + h_2(x)$, under equidistant design.

		$\hat{\sigma}^2_{\text{MS}}$		$\hat{\sigma}^2_{\text{TW}}$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	L_t	L_s	m_t	m_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	82.8	30.0	1270	3770	1.94	4.73	3739	3770	1.93
	0.5	24.5	11.5	37.5	102	6.38	21.7	93.4	102	22.0
	1	14.3	5.78	4.47	8.39	3.70	4.36	7.95	8.37	5.64
	2	11.3	3.85	1.84	1.89	1.80	1.84	1.82	1.89	1.72
	5	10.3	3.28	1.52	1.28	1.53	1.52	1.27	1.29	1.30
100	0.2	20.2	16.0	751	2358	1.43	1.44	1.34	1.66	1.42
	0.5	6.50	4.66	21.3	63.0	1.78	5.65	3.28	25.8	1.80
	1	4.33	2.76	2.80	5.35	2.11	2.72	3.70	5.26	2.23
	2	3.79	2.27	1.50	1.51	1.45	1.50	1.43	1.51	1.35
	5	3.65	2.12	1.37	1.20	1.36	1.37	1.19	1.20	1.20
500	0.2	4.82	5.55	318	1913	1.24	1.23	1.14	1.15	1.20
	0.5	2.51	2.09	9.38	50.2	1.28	1.98	1.30	6.16	1.34
	1	2.18	1.56	1.74	4.15	1.34	1.68	2.09	3.89	1.52
	2	2.10	1.41	1.26	1.29	1.23	1.26	1.18	1.29	1.19
	5	2.08	1.36	1.22	1.11	1.24	1.22	1.12	1.11	1.17

Table 6: Relative MSEs of various estimators for the mean function $f_6(x) = g_3(x) + h_2(x)$, under equidistant design.

		$\hat{\sigma}^2_{\text{MS}}$		$\hat{\sigma}^2_{\text{TW}}$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	L_t	L_s	m_t	m_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	82.9	30.9	1190	2902	4.47	117	2902	2902	12.2
	0.5	24.3	11.5	35.5	79.7	15.8	30.7	79.7	79.7	23.1
	1	14.3	5.78	4.35	7.07	4.11	4.32	7.06	7.07	4.21
	2	11.3	3.85	1.84	1.82	1.81	1.84	1.79	1.82	1.70
	5	10.3	3.28	1.52	1.28	1.54	1.52	1.28	1.29	1.30
100	0.2	20.3	15.9	740	2229	1.41	1.40	1.30	9.69	1.39
	0.5	6.52	4.64	21.1	59.7	2.41	8.89	9.79	45.3	2.47
	1	4.34	2.76	2.78	5.14	2.25	2.76	4.30	5.13	2.26
	2	3.80	2.27	1.49	1.50	1.46	1.50	1.45	1.50	1.33
	5	3.66	2.12	1.37	1.20	1.35	1.37	1.18	1.20	1.19
500	0.2	4.83	5.53	317	1895	1.23	1.22	1.12	1.23	1.18
	0.5	2.51	2.09	9.36	49.7	1.28	2.19	1.56	10.6	1.39
	1	2.18	1.56	1.74	4.12	1.35	1.68	2.34	3.95	1.56
	2	2.10	1.41	1.26	1.29	1.23	1.26	1.18	1.29	1.18
	5	2.08	1.36	1.22	1.11	1.24	1.22	1.12	1.11	1.16

Table 7: Relative MSEs of various estimators for the mean function $f_7(x) = g_1(x) + h_3(x)$, under equidistant design.

		$\hat{\sigma}^2_{\text{MS}}$		$\hat{\sigma}^2_{\text{TW}}$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	L_t	L_s	m_t	m_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	10.2	3.54	1.53	1.64	1.56	1.53	1.72	1.64	1.58
	0.5	10.1	3.21	1.49	1.27	1.52	1.49	1.26	1.27	1.39
	1	10.1	3.17	1.48	1.25	1.51	1.49	1.24	1.25	1.40
	2	10.1	3.16	1.48	1.24	1.51	1.49	1.25	1.25	1.24
	5	10.1	3.16	1.48	1.24	1.51	1.49	1.25	1.25	1.29
100	0.2	3.65	2.10	1.35	1.21	1.34	1.36	1.21	1.21	1.25
	0.5	3.64	2.08	1.35	1.19	1.35	1.36	1.18	1.19	1.18
	1	3.63	2.08	1.35	1.19	1.35	1.36	1.18	1.19	1.17
	2	3.63	2.08	1.35	1.19	1.35	1.36	1.18	1.19	1.19
	5	3.63	2.08	1.35	1.19	1.35	1.36	1.18	1.19	1.18
500	0.2	2.08	1.35	1.22	1.11	1.24	1.22	1.12	1.11	1.17
	0.5	2.08	1.35	1.22	1.11	1.25	1.22	1.14	1.11	1.16
	1	2.08	1.35	1.22	1.12	1.25	1.22	1.14	1.12	1.16
	2	2.08	1.35	1.22	1.12	1.25	1.22	1.15	1.12	1.10
	5	2.08	1.35	1.22	1.12	1.25	1.22	1.15	1.12	1.15

Table 8: Relative MSEs of various estimators for the mean function $f_8(x) = g_2(x) + h_3(x)$, under equidistant design.

		$\hat{\sigma}^2_{\text{MS}}$		$\hat{\sigma}^2_{\text{TW}}$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	L_t	L_s	m_t	m_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	11.1	9.56	1.62	2.23	1.74	1.64	2.43	2.23	1.68
	0.5	10.3	3.75	1.50	1.34	1.49	1.50	1.36	1.34	1.46
	1	10.2	3.27	1.48	1.26	1.49	1.48	1.27	1.26	1.37
	2	10.1	3.19	1.48	1.25	1.51	1.48	1.26	1.25	1.23
	5	10.1	3.17	1.48	1.24	1.51	1.48	1.25	1.25	1.29
100	0.2	3.68	2.56	1.37	1.25	1.35	1.38	1.33	1.25	1.35
	0.5	3.63	2.11	1.36	1.19	1.34	1.36	1.17	1.19	1.18
	1	3.63	2.08	1.36	1.19	1.34	1.36	1.18	1.19	1.17
	2	3.63	2.08	1.35	1.19	1.35	1.36	1.18	1.19	1.20
	5	3.63	2.08	1.35	1.19	1.35	1.36	1.18	1.19	1.17
500	0.2	2.08	1.36	1.22	1.12	1.24	1.22	1.11	1.12	1.18
	0.5	2.08	1.35	1.22	1.12	1.24	1.22	1.14	1.12	1.16
	1	2.08	1.35	1.22	1.12	1.24	1.22	1.15	1.12	1.16
	2	2.08	1.35	1.22	1.12	1.24	1.22	1.15	1.12	1.09
	5	2.08	1.35	1.22	1.12	1.25	1.22	1.15	1.12	1.16

Table 9: Relative MSEs of various estimators for the mean function $f_9(x) = g_3(x) + h_3(x)$, under equidistant design.

		$\hat{\sigma}^2_{\text{MS}}$		$\hat{\sigma}^2_{\text{TW}}$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	L_t	L_s	m_t	m_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	10.4	8.34	1.61	2.26	1.68	1.61	2.48	2.26	1.70
	0.5	10.1	3.48	1.50	1.32	1.54	1.50	1.34	1.32	1.44
	1	10.1	3.21	1.49	1.26	1.51	1.50	1.24	1.26	1.42
	2	10.1	3.17	1.48	1.25	1.51	1.49	1.24	1.25	1.24
	5	10.1	3.16	1.48	1.24	1.50	1.49	1.24	1.25	1.28
100	0.2	3.70	2.30	1.35	1.24	1.36	1.36	1.29	1.24	1.34
	0.5	3.65	2.09	1.35	1.20	1.34	1.35	1.19	1.20	1.19
	1	3.64	2.08	1.35	1.19	1.34	1.35	1.18	1.19	1.18
	2	3.64	2.08	1.35	1.19	1.34	1.36	1.18	1.19	1.19
	5	3.63	2.08	1.35	1.19	1.35	1.36	1.18	1.19	1.18
500	0.2	2.08	1.36	1.22	1.11	1.24	1.22	1.10	1.11	1.16
	0.5	2.08	1.35	1.22	1.11	1.25	1.22	1.13	1.11	1.16
	1	2.08	1.35	1.22	1.11	1.25	1.22	1.14	1.12	1.16
	2	2.08	1.35	1.22	1.12	1.25	1.22	1.14	1.12	1.09
	5	2.08	1.35	1.22	1.12	1.25	1.22	1.15	1.12	1.15

Table 10: Relative MSEs of various estimators for the mean function $f_3(x) = g_3(x) + h_1(x)$, under non-equidistant design.

		$\hat{\sigma}^2$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	d_t	d_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	1174	2875	34.7	34.7	659	1647	50.2
	0.5	35.2	81.6	7.41	21.7	49.1	76.4	8.94
	1	4.28	7.45	3.74	4.17	6.89	7.40	4.16
	2	1.85	1.92	1.81	1.85	1.88	1.92	1.84
	5	1.56	1.35	1.57	1.57	1.35	1.35	1.42
100	0.2	654	1726	1.60	1.61	1.57	1.59	1.57
	0.5	18.8	46.2	1.64	4.31	2.30	16.7	1.72
	1	2.63	4.23	1.99	2.54	2.97	4.12	2.00
	2	1.54	1.50	1.50	1.54	1.45	1.50	1.43
	5	1.45	1.31	1.45	1.45	1.30	1.31	1.36
500	0.2	252	1233	1.43	1.44	1.33	1.33	1.40
	0.5	7.90	33.1	1.43	1.80	1.36	3.48	1.41
	1	1.86	3.34	1.52	1.82	1.94	3.17	1.58
	2	1.46	1.44	1.44	1.46	1.36	1.43	1.44
	5	1.43	1.30	1.43	1.43	1.30	1.30	1.38

Table 11: Relative MSEs of various estimators for the mean function $f_6(x) = g_3(x) + h_2(x)$, under non-equidistant design.

		$\hat{\sigma}^2$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	d_t	d_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	2435	6727	624	827	6411	6701	853
	0.5	70.0	182	43.1	60.7	181	182	62.4
	1	6.99	14.4	6.38	6.94	14.4	14.4	7.06
	2	2.17	2.59	2.14	2.17	2.55	2.59	2.47
	5	1.61	1.43	1.62	1.61	1.42	1.43	1.51
100	0.2	1477	4583	1.60	1.62	121	296	1.59
	0.5	41.0	121	4.98	20.9	47.4	103	5.82
	1	4.28	9.15	3.41	4.19	8.12	9.13	3.71
	2	1.72	1.87	1.66	1.72	1.79	1.87	1.67
	5	1.48	1.34	1.46	1.48	1.32	1.34	1.37
500	0.2	567	3458	1.44	1.45	1.32	1.33	1.40
	0.5	16.1	90.2	1.56	3.93	3.01	27.5	1.66
	1	2.41	6.97	1.75	2.35	4.20	6.77	1.89
	2	1.51	1.68	1.46	1.51	1.52	1.68	1.53
	5	1.44	1.31	1.43	1.44	1.29	1.31	1.38

Table 12: Relative MSEs of various estimators for the mean function $f_9(x) = g_3(x) + h_3(x)$, under non-equidistant design.

		$\hat{\sigma}^2$		$\hat{\sigma}^2_{\text{box}}$				$\hat{\sigma}^2_{\text{CV}}$
n	σ	d_t	d_s	$(d_t, 2)$	$(d_t, 3)$	$(d_s, 2)$	$(d_s, 3)$	
30	0.2	2.29	3.55	2.21	2.28	3.39	3.54	2.22
	0.5	1.63	1.51	1.62	1.63	1.49	1.51	1.52
	1	1.56	1.35	1.57	1.56	1.34	1.35	1.40
	2	1.54	1.31	1.55	1.54	1.30	1.31	1.38
	5	1.53	1.30	1.55	1.54	1.30	1.30	1.29
100	0.2	1.58	1.50	1.56	1.58	1.47	1.50	1.53
	0.5	1.48	1.33	1.47	1.48	1.33	1.33	1.38
	1	1.46	1.32	1.45	1.46	1.32	1.32	1.38
	2	1.46	1.32	1.44	1.46	1.32	1.32	1.38
	5	1.45	1.32	1.44	1.45	1.33	1.32	1.38
500	0.2	1.43	1.32	1.43	1.43	1.30	1.32	1.37
	0.5	1.43	1.30	1.43	1.43	1.30	1.30	1.38
	1	1.43	1.30	1.43	1.43	1.30	1.30	1.38
	2	1.43	1.30	1.43	1.43	1.31	1.30	1.38
	5	1.43	1.29	1.43	1.43	1.31	1.29	1.38

References

Brown, R. L., Durbin, J. & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society, Series B* **37**: 149–192.

Cobb, G. W. (1978). The problem of the Nile: Conditional solution to a changepoint problem, *Biometrika* **65**: 243–251.

Du, J. & Schick, A. (2009). A covariate-matched estimator of the error variance in nonparametric regression, *Journal of Nonparametric Statistics* **21**: 263–285.

Eubank, R. L. & Speckman, P. L. (1994). Nonparametric estimation of functions with jump discontinuities, *IMS Lecture Notes: Change-Point Problems* **23**: 130–144.

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

Hall, P., Kay, J. W. & Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression, *Biometrika* **77**: 521–528.

Hall, P. & Marron, J. S. (1990). On variance estimation in nonparametric regression, *Biometrika* **77**: 415–419.

Hall, P. & Titterington, D. M. (1992). Edge-preserving and peak-preserving smoothing, *Technometrics* **34**: 429–440.

Hinkley, D. V. (1969). Inference about the intersection in two-phase regression, *Biomatrika* **56**: 495–504.

Kim, H. J. & Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression, *Biometrika* **76**: 409–423.

Loader, C. R. (1996). Change point estimation using nonparametric regression, *The Annals of Statistics* **24**: 1667–1678.

McDonald, J. & Owen, A. (1986). Smothing with split linear fits, *Technometrics* **28**: 195–208.

Müller, H. (1992). Change-points in nonparametric regression analysis, *The Annals of Statistics* **20**: 737–761.

Müller, H. & Stadtmüller, U. (1999). Discontinuous versus smooth regression, *Annals of statistics* **27**: 299–337.

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

Qiu, P. (2005). *Image Processing and Jump Regression Analysis*, John Wiley & Sons, New York.

Qiu, P. & Hawkins, D. (2001). A rank-based multivariate cusum procedure, *Technometrics* **43**: 120–132.

Qiu, P. & Yandell, B. (1998). A local polynomial jump detection algorithm in nonparametric regression, *Technometrics* **40**: 141–152.

Rice, J. A. (1984). Bandwidth choice for nonparametric regression, *Annals of Statistics* **12**: 1215–1230.

Sim, C. H., Gan, F. F. & Chang, T. C. (1994). Outlier labeling with boxplot procedures, *Journal of the American Statistical Association,* **100**: 642–652.

Tong, T., Liu, A. & Wang, Y. (2008). Relative errors of difference-based variance estimators in nonparametric regression, *Communications in Statistics - Theory and Methods* **37**: 2890–2902.

Tong, T., Ma, Y. & Wang, Y. (2012). Optimal variance estimation without estimating the mean function, *Bernoulli,* in press .

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

von Neumann, J. (1941). Distribution of the ratio of the mean squared successive difference to the variance, *Annals of Mathematical Statistics* **12**: 367–395.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets, *Biomatrika* **82**: 385–397.

Wu, J. S. & Chu, C. K. (1993a). Kernel type estimators of jump points and valus of a regression function, *The Annals of Statistics* **21**: 1545–1566.

Wu, J. S. & Chu, C. K. (1993b). Nonparametric function estimation and bandwidth selection for discontinuous regression functions, *Statistica Sinica* **3**: 557–576.