JOURNAL OF APPLIED STATISTICS, 2018
VOL. 45, NO. 3, 450–473
http://dx.doi.org/10.1080/02664763.2016.1280004

Testing discontinuities in nonparametric regression

Wenlin Dai[a,b], Yuejin Zhou[c] and Tiejun Tong[d]

[a]HKBU Institute of Research and Continuing Education, Shenzhen, People's Republic of China; [b]CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; [c]School of Mathematics and Statistics, Zhejiang Gongshang University, Hangzhou, People's Republic of China; [d]Department of Mathematics, Hong Kong Baptist University, Hong Kong

ABSTRACT
In nonparametric regression, it is often needed to detect whether there are jump discontinuities in the mean function. In this paper, we revisit the difference-based method in [13] and propose to further improve it. To achieve the goal, we first reveal that their method is less efficient due to the inappropriate choice of the response variable in their linear regression model. We then propose a new regression model for estimating the residual variance and the total amount of discontinuities simultaneously. In both theory and simulation, we show that the proposed variance estimator has a smaller mean-squared error compared to the existing estimator, whereas the estimation efficiency for the total amount of discontinuities remains unchanged. Finally, we construct a new test procedure for detection of discontinuities using the proposed method; and via simulation studies, we demonstrate that our new test procedure outperforms the existing one in most settings.

ARTICLE HISTORY
Received 25 May 2015
Accepted 3 January 2017

KEYWORDS
Asymptotic normality; difference-based estimator; jump point; model selection; nonparametric regression; residual variance

1. Introduction

Consider the following nonparametric regression model:

$$y_i = g(x_i) + h(x_i) + \varepsilon_i, \quad 0 \le i \le n, \tag{1}$$

where y_i are observations, x_i are design points and ε_i are independent and identically distributed (i.i.d.) random errors with mean zero and variance σ^2. In model (1), $f(x) = g(x) + h(x)$ is the mean function. We assume that $g(x)$ is a smooth function and $h(x)$ is a step function with form

$$h(x) = \sum_{j=1}^{p} \psi_j I(x \ge t_j), \quad 0 < t_1 < \cdots < t_p < 1,$$

where p is the number of jumps, ψ_j are the jump magnitudes taking at positions t_j, and $I(x \ge t)$ is the indicator function with value 1 if $x \ge t$ and value 0 otherwise.

When $h(x) = 0$, there is no jump in the regression model and the mean function is smooth. In practical applications, however, such a smoothness assumption may be too

CONTACT Tiejun Tong tongt@hkbu.edu.hk

restrictive if the data include some change points, e.g. the sea-level pressure data [17], the infants growth data [13], and the annual temperature data [6]. It is known that the estimates of the mean function with or without the smoothness assumption are not only quantitatively but also qualitatively different [13]. This results in a fundamental model selection problem and it is often needed to detect whether there are jump discontinuities in the mean function before fitting the model. For this problem, a common approach is to estimate the left and right limits of the mean function and then use their difference to construct appropriate test statistics. As a common practice, a large absolute value of the difference indicates that the mean function may contain jump discontinuities [2–4,7,8,11,14–16].

Apart from the traditional methods, Müller and Stadtmüller [13] proposed a difference-based procedure for testing jump discontinuities in nonparametric regression. Consider an equivalent design with $x_i = i/n$ for $i = 1, \ldots, n$, and define $\gamma = \sum_{j=1}^{p} \psi_j^2$ as the total amount of discontinuities in the data. When $\gamma = 0$, we have $\psi_j = 0$ for all j and so the mean function is smooth; otherwise, the regression model may contain jump discontinuities. By this, the model selection problem is equivalent to testing

$$H_0 : \gamma = 0 \quad \text{versus} \quad H_1 : \gamma > 0. \tag{2}$$

Let $L = o(n) \geq 1$ and

$$Z_k = \frac{1}{2(n-L)} \sum_{i=1}^{n-L} (y_{i+k} - y_i)^2, \quad 1 \leq k \leq L.$$

Noting that $E(Z_k) \approx \sigma^2 + l_k\gamma$ where $l_k = k/2(n-L)$, the authors regressed Z_k on l_k through the linear model

$$Z_k = \sigma^2 + l_k\gamma + \xi_k, \quad 1 \leq k \leq L. \tag{3}$$

The intercept σ^2 and the slope γ were then estimated by least squares. We denote them by $\hat{\sigma}^2_{\text{MS}}$ and $\hat{\gamma}_{\text{MS}}$, and refer to their method as the MS method. Then for testing the hypotheses (2), they constructed a test statistic as

$$T_{\text{MS}} = \frac{\sqrt{m}\hat{\gamma}_{\text{MS}}}{\sqrt{12(\tilde{\mu}_4 - \tilde{\sigma}^4)/5}}, \tag{4}$$

where $\tilde{\mu}_4$ and $\tilde{\sigma}^4$ are consistent estimators for $\mu_4 = E(\varepsilon^4)$ and σ^4. When the random errors follow approximately a normal distribution, $\mu_4 - \sigma^4 = 2\sigma^4$ and so $\tilde{\mu}_4 - \tilde{\sigma}^4$ can be replaced by $2\hat{\sigma}^4_{\text{MS}}$. Under some regularity conditions, the null distribution of T_{MS} follows asymptotically a standard normal distribution.

Note that Z_k uses only the first $n-L$ pairs of differences out of a total of $n-k$ pairs in $\{y_{1+k} - y_1, \ldots, y_n - y_{n-k}\}$. By ignoring the last $L-k$ pairs, we will show in Sections 3 and 4 that their estimator $\hat{\sigma}^2_{\text{MS}}$ is a less efficient estimator for σ^2, especially when n is small or $L-k$ is large. As a consequence, the test statistic (4) is either less powerful or even invalid for testing jump discontinuities, that is, the type I error of the test may not be well controlled. For this point, we can refer to the simulation study in Section3, where the simulated type I error of T_{MS} will be as large as 0.237 at the significance level of 0.05 when $n = 30$. Note that a similar phenomenon was also observed in [18] where a smooth regression model is assumed.

In this paper, we propose to fully use the pairs of differences and develop new estimators for σ^2 and γ. A new test procedure will also be constructed and we show that the proposed test provides a better performance than T_{MS}. Specifically, we organize the rest of this paper as follows. In Section 2, we propose a new estimation method for σ^2 and γ. In Section 3, we study the theoretical properties of the new estimators and then propose a new procedure for testing whether the mean function contains jump discontinuities. In Section 4, we evaluate and compare the proposed estimators and the proposed test procedure with some existing methods. In Section 5, we applied the proposed testing method to a real example and demonstrate its usefulness in practice. Finally, we conclude the paper in Section 6 and provide the proofs of the theorems in Section 7.

2. Main results

To make full use of the available pairs of differences, we define

$$s_k = \frac{1}{2(n-k)} \sum_{i=1}^{n-k} (y_{i+k} - y_i)^2, \quad k = 1, \ldots, m.$$

The same as in [12,19], we refer to them as the lag-k Rice estimators. For the sake of fairness, we let $m = L$ and also assume the same conditions as in [13]. In particular, the following condition on the locations of jump points is assumed:

$$\min_{1 \le j \le p+1} (t_j - t_{j-1}) \ge 2m/n, \tag{5}$$

where $t_0 = 0$ and $t_{p+1} = 1$. This condition ensures that the different change points are not located too close to each other so that they can be separated by the proposed method.

For the equidistant design, it is easy to verify that

$$\begin{aligned} E(s_k) &= \sigma^2 + \frac{1}{2(n-k)} \sum_{i=1}^{n-k} [g(x_{i+k}) - g(x_i) + h(x_{i+k}) - h(x_i)]^2 \\ &= \sigma^2 + \frac{k}{2(n-k)}\gamma + \frac{k^2}{2n^2}\delta + o\left(\frac{k^2}{n^2}\right), \end{aligned} \tag{6}$$

where $\gamma = \sum_{j=1}^{p} \psi_j^2$ as in Section 1 and $\delta = \int_0^1 g'(t)^2 \, \mathrm{d}t + 2\int_0^1 g'(t) \, \mathrm{d}h(t)$. Let $\theta = (\theta_1, \theta_2) = (\sigma^2, \gamma/2)$ and $d_k = k/(n-k)$. Model (6) can be represented as

$$E(s_k) = \theta_1 + d_k\theta_2 + \eta_k,$$

where $\eta_k = k^2\delta/2n^2 + o(k^2/n^2)$. Treating η_k as a negligible term, we have a simple linear regression model with d_k being the independent variables and s_k the response variables. Unlike the Z_k values where exactly $n-L$ pairs are involved, we note that the lag-k Rice estimators s_k involve different number of pairs and this makes the computation much more challenging. To assign a same weight to each pair, we assign weights $w_k = (n-k)/N$ for $k = 1, \ldots, m$ to each response s_k accordingly, where $N = (2n - m - 1)m/2$ is the total number of pairs involved in the regression.

Note also that the responses s_k are correlated with each other. With some straightforward computation, the asymptotic covariance matrix of $\mathbf{s} = (s_1, \ldots, s_m)^{\mathrm{T}}$ is given by $\Sigma = (\Sigma_{ij})_{m\times m}$, where

$$\Sigma_{i,j} = \begin{cases} n^{-1}(\mu_4 - \sigma^4) + o(n^{-1}) & \text{if } i \neq j, \\ n^{-1}\mu_4 + o(n^{-1}) & \text{if } i = j. \end{cases}$$

We then apply the generalized least-squares method to estimate the parameters. Specifically by McElroy [10], we have the estimators of θ_1 and θ_2 as

$$\hat{\theta}_{\text{new},1} = \sum_{k=1}^{m} w_k s_k - \bar{d}_w \hat{\theta}_{\text{new},2},$$

$$\hat{\theta}_{\text{new},2} = \left(\sum_{k=1}^{m} w_k d_k^2 - \bar{d}_w^2\right)^{-1} \sum_{k=1}^{m} w_k (d_k - \bar{d}_w) s_k,$$

where $\bar{d}_w = \sum_{k=1}^{m} w_k d_k$. Then correspondingly, the new estimators of σ^2 and γ are $\hat{\sigma}^2_{\text{new}} = \hat{\theta}_{\text{new},1}$ and $\hat{\gamma}_{\text{new}} = 2\hat{\theta}_{\text{new},2}$. They are the best linear unbiased estimators of σ^2 and γ, respectively [9].

3. Asymptotic properties

In this section, we investigate the theoretical properties of the proposed estimators. For ease of notation, let

$$a_k = 1 - \bar{d}_w \frac{(d_k - \bar{d}_w)}{\sum_{k=1}^{m} w_k d_k^2 - \bar{d}_w^2} \quad \text{and} \quad b_k = \frac{d_k - \bar{d}_w}{\sum_{k=1}^{m} w_k d_k^2 - \bar{d}_w^2}.$$

Then for $\lambda, \rho \in \mathrm{R}$, we define

$$\hat{\theta}_{\text{new}}(\lambda, \rho) = \lambda \hat{\theta}_{\text{new},1} + \rho \hat{\theta}_{\text{new},2} = \sum_{k=1}^{m} (\lambda a_k + \rho b_k) w_k s_k.$$

Let also $y = (y_1, \ldots, y_n)^{\mathrm{T}}$, $c_0 = 0$ and $c_k = (\lambda a_k + \rho b_k)/2N$ for $k = 1, \ldots, m$. We note that $\hat{\theta}_{\text{new}}(\lambda, \rho)$ can be represented as a quadratic form $\hat{\theta}_{\text{new}}(\lambda, \rho) = Y^{\mathrm{T}} D Y$, where D is an $n \times n$ matrix with elements

$$d_{ij} = \begin{cases} \displaystyle\sum_{k=1}^{m} c_k + \sum_{k=0}^{\min\{i-1, n-i, m\}} c_k, & 1 \leq i = j \leq n, \\ -c_{|i-j|}, & 0 < |i-j| = k \leq m, \\ 0 & \text{otherwise.} \end{cases}$$

We can derive its variance as

$$\begin{aligned} \operatorname{var}[\hat{\theta}_{\text{new}}(\lambda, \rho)] &= (\operatorname{var}(\varepsilon^2) - 2\sigma^4)\operatorname{tr}(\operatorname{diag}(D)^2) + 2\sigma^4 \operatorname{tr}(D^2) \\ &\quad + 4\sigma^2 f^{\mathrm{T}} D^2 f + 4\sigma^3 \mu_3 (f^{\mathrm{T}} D \operatorname{diag}(D) \mathbf{1}), \end{aligned}$$

where $f = (f(x_1), \ldots, f(x_n))^{\mathrm{T}}$, $\mu_3 = E(\varepsilon^3)$ and $\operatorname{diag}(D)$ is the diagonal matrix of D. This form makes it convenient to calculate the theoretical results. The following theorem gives

the mean-squared errors (MSE) for the proposed estimators and also the estimators in [13] for comparison.

Theorem 3.1: *Assume that condition* (5) *holds,* $m \to \infty$ *and* $L = m = n^r$ *for* $1/2 < r < 2/3$. *Then for any mean function* $f = g + h$ *with the first derivative of* g *being bounded, we have*

$$\begin{aligned}
\mathrm{MSE}(\hat{\sigma}^2_{\mathrm{new}}) &= \frac{1}{n}\mathrm{var}(\varepsilon^2) + \frac{m}{15n^2}\mathrm{var}(\varepsilon^2) + \frac{4m}{15n^2}\sigma^2\gamma + o\left(\frac{m}{n^2}\right),\\
\mathrm{MSE}(\hat{\sigma}^2_{\mathrm{MS}}) &= \frac{1}{n}\mathrm{var}(\varepsilon^2) + \frac{16m}{15n^2}\mathrm{var}(\varepsilon^2) + \frac{4m}{15n^2}\sigma^2\gamma + o\left(\frac{m}{n^2}\right),\\
\mathrm{MSE}(\hat{\gamma}_{\mathrm{new}}) &= \frac{12}{5m}\mathrm{var}(\varepsilon^2) + \frac{48}{5m}\sigma^2\gamma + o\left(\frac{1}{m}\right),\\
\mathrm{MSE}(\hat{\gamma}_{\mathrm{MS}}) &= \frac{12}{5m}\mathrm{var}(\varepsilon^2) + \frac{48}{5m}\sigma^2\gamma + o\left(\frac{1}{m}\right).
\end{aligned}$$

The proof of Theorem 3.1 is given in Section 7.1. Theorem 3.1 indicates that $\hat{\gamma}_{\mathrm{new}}$ and $\hat{\gamma}_{\mathrm{MS}}$ are asymptotically equivalent. Both $\hat{\sigma}^2_{\mathrm{new}}$ and $\hat{\sigma}^2_{\mathrm{MS}}$ attain the root-n convergence rate. Compared to the MS estimator, the proposed new estimator $\hat{\sigma}^2_{\mathrm{new}}$ has a smaller second-order term in the MSE expression. This improvement results from the information of residual variance provided by $m(m-1)/2$ extra pairs in our model. The finite sample performance of the new estimators is presented in Section 4.

Theorem 3.2: *Assume that* $\mu_3 = 0$, $m \to \infty$ *and* $m = n^r$ *for* $1/2 < r < 2/3$. *We have the following asymptotic normality for the proposed estimators,*

$$\begin{pmatrix} \sqrt{n}(\hat{\sigma}^2_{\mathrm{new}} - \sigma^2) \\ \sqrt{m}(\hat{\gamma}_{\mathrm{new}} - \gamma) \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (\mu_4 - \sigma^4)\begin{pmatrix} 1 & 0 \\ 0 & 12/5 \end{pmatrix}\right) \quad \text{as } n \to \infty,$$

where $\xrightarrow{D}$ *denotes convergence in distribution.*

The proof of Theorem 3.2 is given in Section 7.2. Theorem 3.2 can be used to test whether or not the mean function is smooth. Specifically, to test $H_0 : \gamma = 0$ versus $H_1 : \gamma > 0$, we construct the test statistic as

$$T_{\mathrm{new}} = \frac{\sqrt{m}\hat{\gamma}_{\mathrm{new}}}{\sqrt{12(\tilde{\mu}_4 - \tilde{\sigma}^4)/5}},$$

where $\tilde{\mu}_4$ and $\tilde{\sigma}^2$ are consistent estimates of μ_4 and σ^2, respectively. If we treat η_k as normal errors, then $\mu_4 - \sigma^4 = 2\sigma^4$ and we can replace $\tilde{\mu}_4 - \tilde{\sigma}^4$ by $2\hat{\sigma}^4_{\mathrm{new}}$. Under H_0, the test statistic T_{new} follows a standard normal distribution. Then at the significance level of α, we reject H_0 if the observed T_{new} value is larger than z_α, where z_α is the upper αth percentile of the standard normal distribution.

4. Simulation studies

In this section, we first conduct Monte Carlo simulations to assess the finite sample performance of the proposed estimation and testing methods and compare them with the existing methods in [13]. Specifically, we consider respectively three smooth functions:

$$g_1(x) = 0,$$
$$g_2(x) = x/2,$$
$$g_3(x) = x/2 + x\sin(2\pi x)/4,$$

and three jump functions:

$$h_1(x) = 0,$$
$$h_2(x) = I(x \geq 0.5),$$
$$h_3(x) = I(x \geq 0.25) - 1.5I(x \geq 0.5).$$

In total, we have 9 combinations for the mean function. Note that $g_i + h_1$ for $i = 1,2,3$ are smooth functions with no jump points.

For the sample size, we consider $n = 30$, 100 and 500 that correspond to small, moderate and large sample sizes, respectively. The random errors are generated independently from the normal distribution $N(0, 0.25)$. For the bandwidth $\hat{m}$, we follow the selection method in [13]. Specifically, we let

$$\hat{m} = \underset{m}{\operatorname{argmin}} \left\{ \frac{1}{2m_0+1} \sum_{i=m-m_0}^{m+m_0} \hat{\gamma}^2(i) - \left[\frac{1}{2m_0+1} \sum_{i=m-m_0}^{m+m_0} \hat{\gamma}(i) \right]^2 \right\}, \tag{7}$$

where $m_0 = \max\{[n/50], 2\}$ with $[x]$ being the greatest integer smaller than or equal to x. This criterion suggests to choose the bandwidth that minimizes an approximation for the variance of estimator in the area of $[m - m_0, m + m_0]$.

For each simulated data set, we calculate the estimates of γ and σ^2 for both the new method and the MS method. We then repeat the procedure 1000 times for each simulation setting and report the MSE of $\hat{\gamma}$ and the relative difference of mean-squared errors (rdMSE) of $\hat{\sigma}^2_{\text{MS}}$ and $\hat{\sigma}^2_{\text{new}}$, $n^2[\text{MSE}(\hat{\sigma}^2_{\text{MS}}) - \text{MSE}(\hat{\sigma}^2_{\text{new}})]/[m\text{var}(\varepsilon^2)]$, in Table 1, respectively. From the results in Table 1, it is evident that $\hat{\sigma}^2_{\text{new}}$ outperforms $\hat{\sigma}^2_{\text{MS}}$ significantly in all settings. Note also that $\hat{\gamma}_{\text{new}}$ and $\hat{\gamma}_{\text{MS}}$ perform very similarly. Hence, together with the theoretical results in Theorem 3.1, we conclude that the additional pairs introduced into the regression does improve the overall performance of the estimators in both theory and simulations.

Our second simulation study is to compare the performance of the test statistics T_{new} and T_{MS} for testing $H_0 : \gamma = 0$ versus $H_1 : \gamma > 0$. For the smooth function, we consider the same three functions as in the previous simulation study. For the step function, we consider $h(x) = \psi I(x \geq 0.5)$ with the ψ value ranging from 0, 0.25, 0.5, 0.75 and 1. Note that $\psi = 0$ represents the null hypothesis, and vice versa. The sample size n is set to be 30, 100, 200 and 500, and the random errors are generated independently from $N(0, 0.25)$. We investigate the type I errors and the power function of the proposed test T_{new} and T_{MS} at $\alpha = 0.1$, 0.05 and 0.025 significance levels. With 1000 simulations for each setting, we report the simulation results in Table 2 for $f(x) = g_1(x) + h(x)$, in Table 3 for

Table 1. Simulation results for MSE of $\hat{\gamma}_{\text{new}}$, $\hat{\gamma}_{\text{MS}}$ and relative difference between MSEs (rdMSE) of $\hat{\sigma}^2_{\text{new}}$ and $\hat{\sigma}^2_{\text{MS}}$.

n	$h(x)$	$g(x)$	MSE($\hat{\gamma}_{\text{new}}$)	MSE($\hat{\gamma}_{\text{MS}}$)	rdMSE
30	1	1	0.12	0.18	1.98
	1	2	0.18	0.21	1.79
	1	3	0.17	0.22	1.59
	2	1	0.44	0.53	2.91
	2	2	0.67	0.77	2.51
	2	3	0.83	0.94	2.29
	3	1	2.45	2.61	4.11
	3	2	2.42	2.61	2.99
	3	3	2.84	3.15	2.74
100	1	1	0.027	0.025	1.37
	1	2	0.024	0.032	1.38
	1	3	0.032	0.039	1.40
	2	1	0.13	0.15	1.70
	2	2	0.23	0.24	1.78
	2	3	0.29	0.32	1.75
	3	1	0.56	0.58	2.13
	3	2	0.58	0.61	2.07
	3	3	0.81	0.84	2.10
500	1	1	0.0047	0.0061	1.55
	1	2	0.0048	0.0058	1.34
	1	3	0.0064	0.0068	1.27
	2	1	0.040	0.043	1.60
	2	2	0.12	0.12	1.82
	2	3	0.17	0.17	1.92
	3	1	0.12	0.13	2.25
	3	2	0.11	0.12	1.65
	3	3	0.35	0.36	2.55

Table 2. Simulated type I errors and powers of T_{new} and T_{MS} for $f(x) = g_1(x) + h(x)$, respectively.

		T_{new}			T_{MS}		
n	ψ	0.1	0.05	0.025	0.1	0.05	0.025
30	0	0.111	0.080	0.057	0.233	0.194	0.170
	0.25	0.271	0.190	0.147	0.357	0.298	0.265
	0.5	0.549	0.477	0.410	0.544	0.482	0.435
	0.75	0.828	0.775	0.717	0.755	0.702	0.660
	1	0.983	0.968	0.960	0.959	0.932	0.910
100	0	0.083	0.043	0.031	0.193	0.139	0.112
	0.25	0.361	0.273	0.198	0.409	0.335	0.274
	0.5	0.789	0.715	0.635	0.752	0.690	0.631
	0.75	0.978	0.956	0.936	0.954	0.933	0.912
	1	1.000	1.000	1.000	1.000	1.000	1.000
200	0	0.094	0.048	0.030	0.168	0.120	0.089
	0.25	0.454	0.345	0.284	0.472	0.375	0.302
	0.5	0.911	0.864	0.807	0.874	0.816	0.767
	0.75	0.999	0.996	0.994	0.997	0.995	0.993
	1	1.000	1.000	1.000	1.000	1.000	1.000
500	0	0.112	0.058	0.039	0.158	0.097	0.061
	0.25	0.553	0.447	0.354	0.552	0.439	0.370
	0.5	0.958	0.933	0.905	0.940	0.906	0.882
	0.75	1.000	0.999	0.999	0.999	0.998	0.994
	1	1.000	1.000	1.000	1.000	1.000	1.000

$f(x) = g_2(x) + h(x)$, and in Table 4 for $f(x) = g_3(x) + h(x)$, respectively. From the three tables, we observe that the simulated type I errors of T_{MS} always exceed the nominal levels. Accordingly, the simulated type I errors of the new method are much smaller and also more

Table 3. Simulated type I errors and powers of T_{new} and T_{MS} for $f(x) = g_2(x) + h(x)$, respectively.

		T_{new}			T_{MS}		
n	ψ	0.1	0.05	0.025	0.1	0.05	0.025
30	0	0.226	0.169	0.126	0.331	0.283	0.245
	0.25	0.658	0.554	0.489	0.625	0.561	0.517
	0.5	0.885	0.834	0.794	0.822	0.781	0.739
	0.75	0.977	0.959	0.946	0.946	0.922	0.903
	1	0.997	0.995	0.993	0.997	0.994	0.992
100	0	0.217	0.170	0.118	0.309	0.247	0.191
	0.25	0.685	0.589	0.517	0.660	0.589	0.542
	0.5	0.923	0.889	0.857	0.899	0.858	0.823
	0.75	0.987	0.984	0.980	0.982	0.978	0.965
	1	1.000	1.000	1.000	1.000	1.000	1.000
200	0	0.213	0.171	0.118	0.305	0.222	0.173
	0.25	0.889	0.837	0.791	0.842	0.778	0.729
	0.5	0.996	0.986	0.977	0.989	0.985	0.974
	0.75	1.000	0.999	0.999	0.999	0.999	0.999
	1	1.000	1.000	1.000	1.000	1.000	1.000
500	0	0.204	0.178	0.112	0.296	0.210	0.157
	0.25	0.884	0.836	0.785	0.853	0.808	0.745
	0.5	0.997	0.994	0.989	0.990	0.987	0.978
	0.75	1.000	0.999	0.999	0.999	0.999	0.999
	1	1.000	1.000	1.000	1.000	1.000	1.000

Table 4. Simulated type I errors and powers of T_{new} and T_{MS} for $f(x) = g_3(x) + h(x)$, respectively.

		T_{new}			T_{MS}		
n	ψ	0.1	0.05	0.025	0.1	0.05	0.025
30	0	0.149	0.103	0.079	0.264	0.220	0.196
	0.25	0.353	0.283	0.230	0.428	0.379	0.330
	0.5	0.658	0.556	0.493	0.621	0.560	0.514
	0.75	0.888	0.840	0.796	0.812	0.769	0.744
	1	0.987	0.983	0.969	0.973	0.956	0.938
100	0	0.111	0.059	0.044	0.246	0.182	0.143
	0.25	0.440	0.317	0.227	0.470	0.395	0.326
	0.5	0.931	0.890	0.843	0.869	0.831	0.786
	0.75	0.995	0.993	0.990	0.993	0.991	0.985
	1	1.000	1.000	1.000	1.000	1.000	1.000
200	0	0.112	0.068	0.044	0.235	0.169	0.132
	0.25	0.666	0.523	0.414	0.592	0.509	0.426
	0.5	0.995	0.989	0.973	0.976	0.955	0.932
	0.75	1.000	1.000	1.000	1.000	1.000	1.000
	1	1.000	1.000	1.000	1.000	1.000	1.000
500	0	0.225	0.136	0.092	0.297	0.215	0.158
	0.25	0.912	0.839	0.748	0.812	0.736	0.664
	0.5	1.000	1.000	1.000	0.999	0.996	0.995
	0.75	1.000	1.000	1.000	1.000	1.000	1.000
	1	1.000	1.000	1.000	1.000	1.000	1.000

close to the nominal levels in most settings. This shows that our proposed test provides a more accurate control than the test method in [13]. Furthermore, we note that the power of T_{new} can be larger than that of T_{MS} in a wide range of settings. To illustrate this, we plot in Figure 1 the power function for $f(x) = g_1(x) + h(x)$ at the 0.05 significance level. From the figure, the power function of T_{new} will always exceed that of T_{MS} as long as γ is not too small.

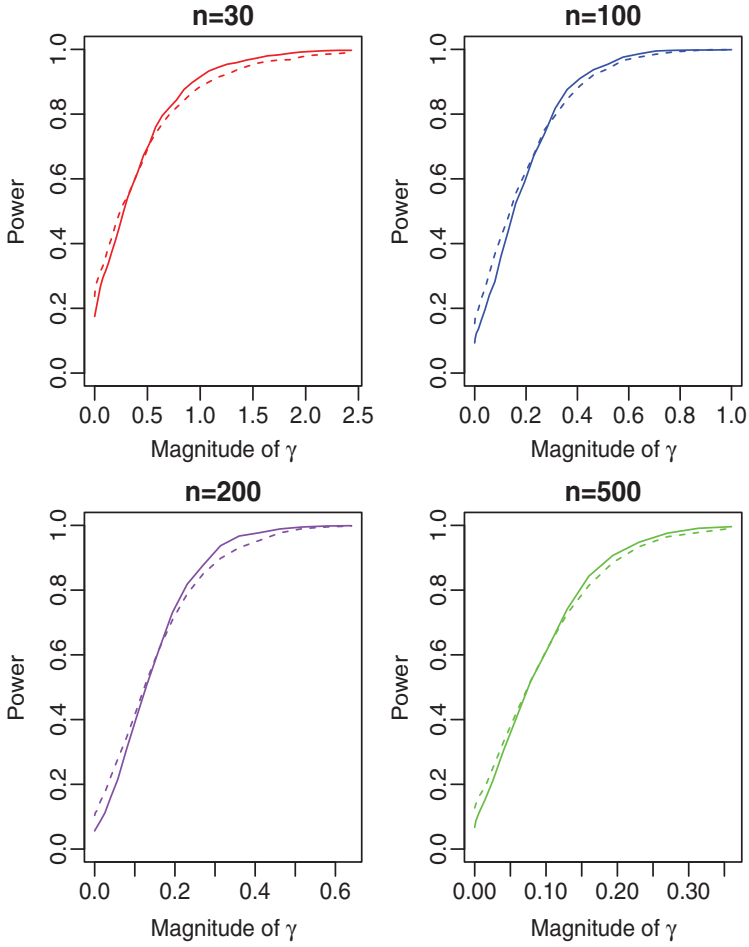

Figure 1. Test power of T_{MS} (dashed lines) and T_{new} (solid lines) against the magnitude of γ, where $g_1(x) = 0, h(x) = \psi I(x \geq 0.5)$, and $\alpha = 0.05$.

5. Real study

In this section, we apply our new testing method to the data of monthly gasoline prices of the United States for the recent 22 years, starting from January 1994 to December 2015. The data are freely available from the website of U.S. Energy Information Administration: http://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_nus_m.htm. For illustration, we display the 264 records for monthly gasoline prices in Figure 2.

We consider the null hypothesis being there is no jump in the gasoline prices data. To compare the behavior of the new and existing methods in practice, we set the bandwidth ranging from 16 to 132, with the two end points corresponding to $n^{1/2}$ and $n/2$, respectively. We then present the estimated results for $\hat{\gamma}$, $\hat{\sigma}^2$ and $\hat{T}$ in Figure 3, respectively. For γ in the left panel of Figure 3, the two methods provide comparable estimates, with the MS estimate slightly greater when the bandwidth is large. For σ^2 in the middle panel of Figure 3, we note that the MS method may provide negative variance estimates when the bandwidth is large, owing to the fact that many pairs of differences are prohibited from analysis. As

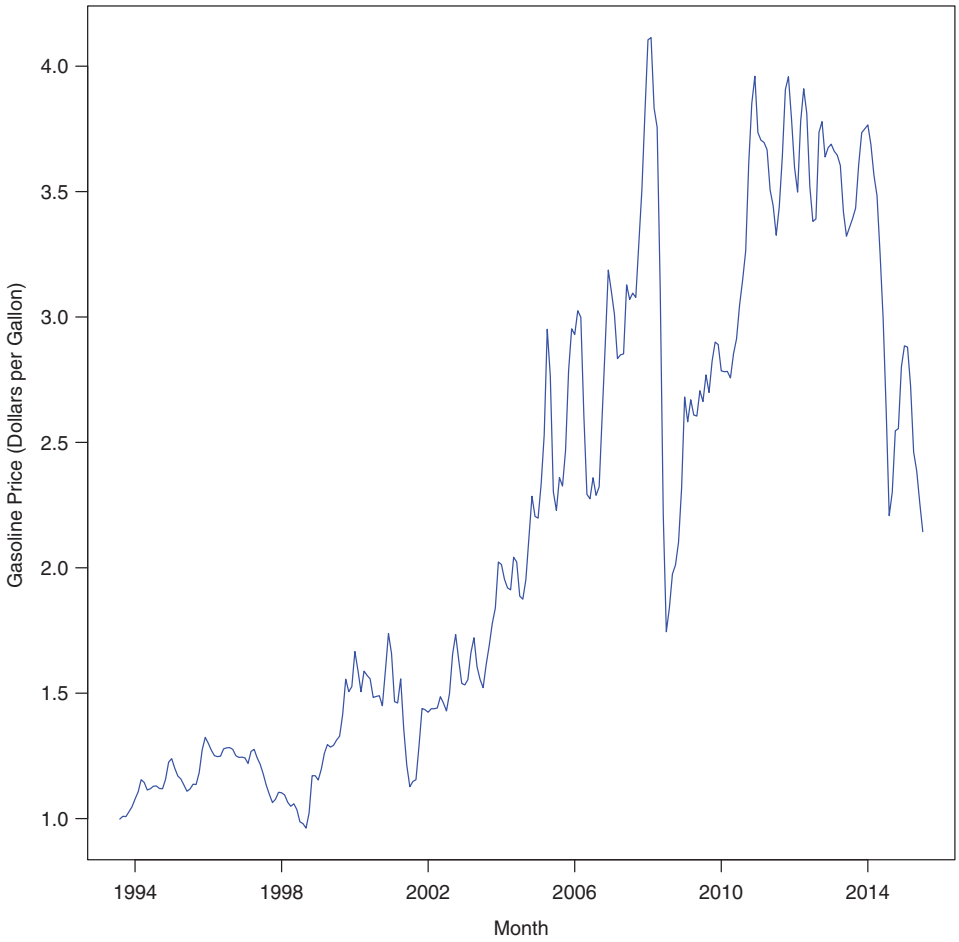

Figure 2. Monthly gasoline prices of the United States from January 1994 to December 2015.

a consequence, the MS method fails to provide a reliable testing result compared with our testing method. For details, see the right panel of Figure 3. Finally, we compare the two methods by using their respective bandwidths that minimize the local variation of $\hat{\gamma}$ in Equation (7). For our method, we get $\hat{m} = 131$ and $\hat{T}_{\text{new}} = 1342.2 > z_{0.95} = 1.64$; and for the MS method, we get $\hat{m} = 128$ and $\hat{T}_{\text{MS}} = 80.4 > z_{0.95} = 1.64$. It is evident that our new testing method provides a stronger evidence to reject the null hypothesis of no jump.

To summarize, there exists at least one jump discontinuity in the mean function of monthly gasoline prices. More specifically, we observe that there are two obvious jumps in the data: the first one is around 2008 in which the oil price went along with the subprime crisis; and the second one is around 2014 in which the oil price went down sharply. Finally, we note that the number and positions of jumps are quite obvious in our illustration example. In case if they are not that obvious, we recommend to apply some well-known methods such as [17,20] for further investigation upon the rejection of the null hypothesis.

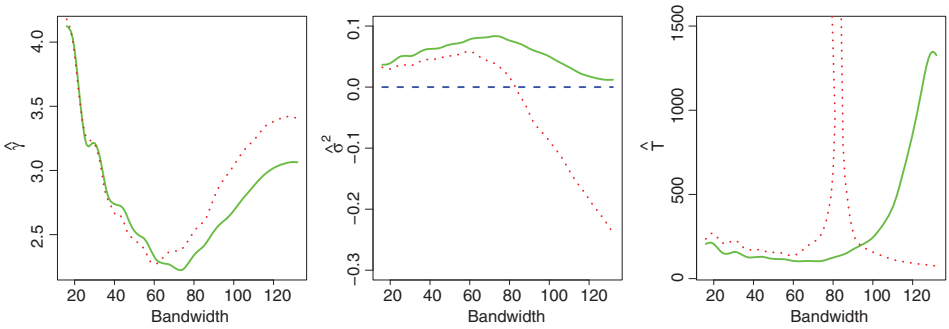

Figure 3. Estimation and testing results of the two methods with various bandwidths. Left panel: $\hat{\gamma}$, middle panel: $\hat{\sigma}^2$ and right panel: $\hat{T}$. Red dotted lines: the MS method. Green solid lines: our new method.

6. Conclusion

This paper considers the simultaneous estimation of the total amount of discontinuities γ and the residual variance σ^2 in nonparametric regression with jump continuities. It meanwhile considers the detection problem whether or not the mean function is a smooth function. To achieve the goals, we first reveal that the method in [13] is less efficient due to the inappropriate choice of the response variable in their linear regression model. We then propose a new regression model for estimating σ^2 and γ simultaneously, and also propose a new test procedure for detecting the smoothness. In both theory and simulations, we show that the proposed variance estimator has a smaller MSE compared to the estimator in [13], whereas the two estimators of γ remain a similar performance. Consequently, the proposed new test procedure also improves the existing test in [13]. An example of gasoline data is analyzed with both existing and proposed methods for practical implementation.

The current paper only considers the two-parameter asymptotic linear model for simplicity. In spirit, our proposed method in Section 2 can be applied to build a three-parameter asymptotic linear model using some higher order Taylor expansion, as in formula (4.11) in [13]. In addition, we note that the proposed method requires a strong condition on the locations of jump points, i.e. condition (5), that specifies no jump points in the boundaries. This requirement can be too restrictive in practice, especially when n is small or when L is chosen large. In such situations, our proposed method itself may not be satisfactory, although it is always better than the test method in [13]. Further research may be needed in this direction to overcome this limitation so that the proposed method is widely applicable.

7. Proofs

7.1. Proof of Theorem 3.1

We first present two lemmas. Lemma 7.1 can be derived with some tedious but simple calculation. For Lemma 7.2, we provide a detailed proof.

Lemma 7.1: *Assume that $m \to \infty$ and $m/n \to 0$. We have*

(a)

$$\sum_{k=1}^{m} c_k = \frac{1}{2N}\left[\left(m - \frac{m^2}{2n} + o\left(\frac{m^2}{n}\right)\right)\lambda + (m + o(m))\rho\right].$$

(b)

$$\sum_{k=1}^{l-1} c_k = \frac{1}{2N}\left[\left(4l - \frac{3l^2}{m} + o(l) + O(1)\right)\lambda \right.$$
$$\left. + \left(\frac{6nl^2}{m^2} - \frac{6nl}{m} - \frac{12ln}{m^2} + \frac{4l^3}{m^2} + 9l + \frac{6n}{m} - \frac{12l^2}{m} + o(l) + o\left(\frac{n}{m}\right)\right)\rho\right],$$
$$1 \le l \le m.$$

(c)

$$\sum_{k=1}^{m} c_k^2 = \frac{1}{4N^2}\left[\left(4m - \frac{4m^2}{n} - 12 + o\left(\frac{m^2}{n}\right) + o(1)\right)\lambda^2 \right.$$
$$+ \left(\frac{12n^2}{m} - 24n - \frac{72n^2}{m^2} + o(n) + o\left(\frac{n^2}{m^2}\right)\right)\rho^2$$
$$\left. + \left(-12n + 20m + \frac{60n}{m} + o(m) + o\left(\frac{n}{m}\right)\right)\lambda\rho\right].$$

(d)

$$\sum_{k=l}^{m} kc_k = \frac{1}{2N}[\lambda(O(m^2)) + \rho(O(mn))], \quad 1 \le l \le m.$$

(e)

$$\sum_{k=1}^{l-1} k^2 c_k = \frac{1}{2N}[\lambda(O(l^3)) + \rho(O(l^3 n/m))], \quad 1 \le l \le m.$$

(f)

$$\sum_{k=1}^{m} k^2 c_k = \frac{1}{2N}[\lambda(O(m^3)) + \rho(O(m^2 n))].$$

Lemma 7.2: *Assume that $m \to \infty$ and $m/n \to 0$. We have*

(I)

$$\begin{aligned}\operatorname{tr}(D^2) = \frac{1}{4N^2}\Bigg[&\lambda^2\left(4nm^2 - \frac{56}{15}m^3 + 8nm + o(m^3) + o(nm)\right)\\ &+ \rho^2\left(\frac{12}{5}n^2m + 24\frac{n^3}{m} - 60n^2 - \frac{34}{5}nm^2 - 144\frac{n^3}{m^2}\right.\\ &\left.+ o(nm^2) + o(n^2) + o\left(\frac{n^3}{m^2}\right)\right)\\ &+ \lambda\rho\left(-\frac{2}{5}nm^2 - 24n^2 + o(nm^2) + o(n^2)\right)\Bigg].\end{aligned}$$

(II)

$$\begin{aligned}\operatorname{tr}(\operatorname{diag}(D)^2) = \frac{1}{4N^2}\Bigg[&\lambda^2\left(4nm^2 - \frac{56}{15}m^3 + o(m^3) + o(nm)\right)\\ &+ \rho^2\left(\frac{12}{5}n^2m - \frac{34}{5}nm^2 + o(nm^2) + o(n^2) + o\left(\frac{n^3}{m^2}\right)\right)\\ &+ \lambda\rho\left(-\frac{2}{5}nm^2 + o(nm^2)\right)\Bigg].\end{aligned}$$

(III)

$$\begin{aligned}f^{\mathrm{T}}D^2f = \frac{1}{4N^2}\Bigg[&\lambda^2\left(\frac{4}{15}m^3 + o(m^3)\right)\\ &+ \rho^2\left(\frac{12}{5}n^2m + 24n^2 - \frac{14}{5}nm^2 + o(n^2) + o(nm^2)\right)\\ &+ \lambda\rho\left(-\frac{2}{5}nm^2 + o(nm^2)\right)\Bigg].\end{aligned}$$

(IV)

$$f^{\mathrm{T}}D\operatorname{diag}(D)\boldsymbol{1} = O\left(\frac{m^2}{n^3}\right) + O\left(\frac{1}{n^2}\right).$$

Proof of Lemma 7.2: We first prove (I). Note that

$$\begin{aligned}\operatorname{tr}(D^2) = (n-2m)&\left[\left(2\sum_{k=1}^{m}c_k\right)^2 + 2\sum_{k=1}^{m}c_k^2\right]\\ &+ 2\sum_{l=1}^{m}\left[\left(\sum_{k=1}^{m}c_k + \sum_{k=1}^{l-1}c_k\right)^2 + \sum_{k=1}^{m}c_k^2 + \sum_{k-1}^{l-1}c_k^2\right].\end{aligned} \tag{8}$$

By Lemma 7.1 (a) and (c), we have

$$\left(\sum_{k=1}^{m} c_k\right)^2 = \frac{1}{4N^2}\left[\lambda^2\left(m^2 - \frac{m^3}{n} + o\left(\frac{m^3}{n}\right)\right) + \rho^2(m^2 + o(m^2)) + \lambda\rho(2m^2 + o(m^2))\right],$$

$$\sum_{k=1}^{m} c_k^2 = \frac{1}{4N^2}\left[\lambda^2(4m + o(m)) + \rho^2\left(\frac{12n^2}{m} - 24n - \frac{72n^2}{m^2} + o(n) + o\left(\frac{n^2}{m^2}\right)\right) + \lambda\rho(-12n + o(n))\right].$$

This leads to

$$\begin{aligned}(n-2m)&\left[\left(2\sum_{k=1}^{m} c_k\right)^2 + 2\sum_{k=1}^{m} c_k^2\right] \\ &= \frac{1}{4N^2}\left[\lambda^2(4nm^2 - 12m^3 + 8nm + o(m^3) + o(nm)) \right. \\ &\quad + \rho^2\left(\frac{24n^3}{m} - 96n^2 - \frac{144n^3}{m^2} + 4nm^2 + o(n^2) + o\left(\frac{n^3}{m^2}\right) + o(nm^2)\right) \\ &\quad \left. + \lambda\rho(8nm^2 - 24n^2 + o(m^2) + o(n))\right]. \end{aligned} \tag{9}$$

In addition, by Lemma 7.1 (a), (b) and (c) we have

$$\begin{aligned}\sum_{l=1}^{m}\left(\sum_{k=1}^{m} c_k + \sum_{k=1}^{l-1} c_k\right)^2 &= \frac{1}{4N^2}\left[\lambda^2\left(\frac{62}{15}m^3 + o(m^3)\right) + \lambda\rho\left(-\frac{21}{5}nm^2 + o(nm^2)\right)\right. \\ &\quad \left. + \rho^2\left(\frac{6}{5}n^2m - \frac{27}{5}nm^2 + o(nm^2) + o(n^2)\right)\right] \end{aligned} \tag{10}$$

and

$$\sum_{l=1}^{m}\left(\sum_{k=1}^{m} c_k^2 + \sum_{k=1}^{l-1} c_k^2\right) = \frac{1}{4N^2}(\lambda^2 O(m^2) + \rho^2(18n^2 + o(n^2)) + \lambda\rho O(mn)). \tag{11}$$

Now plugging (9)–(11) into (8), we have $\mathrm{tr}(D^2)$ as shown in (I).

For (II), we have

$$\begin{aligned}\operatorname{tr}(\operatorname{diag}(D)^2) &= \frac{1}{4N^2}\left[(n-2m)\left(2\sum_{k=1}^{m}c_k\right)^2 + 2\sum_{l=1}^{m}\left(\sum_{k=1}^{m}c_k+\sum_{k=1}^{l-1}c_k\right)^2\right]\\ &= \frac{1}{4N^2}\left[\lambda^2\left(4nm^2-\frac{56}{15}m^3+o(m^3)+o(nm)\right)\right.\\ &\quad+\rho^2\left(\frac{12}{5}n^2m-\frac{34}{5}nm^2+o(nm^2)+o(n^2)+o\left(\frac{n^3}{m^2}\right)\right)\\ &\quad\left.+\lambda\rho\left(-\frac{2}{5}nm^2+o(nm^2)\right)\right]\end{aligned}$$

Now we prove (III). Let $f_i = f(x_i)$, $g_i = g(x_i)$, $h_i = h(x_i)$, $g_i' = g'(x_i)$ and $g_i'' = g''(x_i)$. Noting that D is a symmetric matrix, we have

$$f^{\mathrm{T}}D^2f = g^{\mathrm{T}}D^{\mathrm{T}}Dg + 2g^{\mathrm{T}}D^{\mathrm{T}}Dh + h^{\mathrm{T}}D^{\mathrm{T}}Dh = p^{\mathrm{T}}p + 2p^{\mathrm{T}}q + q^{\mathrm{T}}q,$$

where $p = Dg = (p_1, p_2, \ldots, p_n)^{\mathrm{T}}$ and $q = Dh = (q_1, q_2, \ldots, q_n)^{\mathrm{T}}$. For $i \in [m+1, n-m]$, by Lemma 7.1 (f) we have

$$\begin{aligned}p_i &= \sum_{k=1}^{m}c_k[(g_i-g_{i-k})-(g_{i+k}-g_i)]\\ &= -\frac{2g_i''}{n^2}\sum_{k=1}^{m}k^2c_k(1+o(1))\\ &= \frac{1}{2N}\left[\lambda\left(O\left(\frac{m^3}{n^2}\right)\right)+\rho\left(O\left(\frac{m^2}{n}\right)\right)\right]\end{aligned}$$

and

$$\begin{aligned}q_i &= \sum_{k=1}^{m}c_k[(h_i-h_{i-k})-(h_{i+k}-h_i)]\\ &= \sum_{k=1}^{m}c_k[\psi_jI(x_{i-k}<t_j\le x_i)-\psi_{j+1}I(x_i<t_j\le x_{i+k})].\end{aligned}$$

For $i \in [1, m]$, by Lemma 7.1 (d), (e) and (f), we have

$$\begin{aligned}p_i &= \sum_{k=1}^{i-1}c_k(g_i-g_{i-k})-\sum_{k=1}^{m}c_k(g_{i+k}-g_i)\\ &= -\frac{g_i'}{n}\sum_{k=i}^{m}kc_k(1+o(1))\\ &= \frac{1}{2N}\left[\lambda\left(O\left(\frac{m^2}{n}\right)\right)+\rho(O(m))\right].\end{aligned}$$

Similar, for $i \in [n-m+1, n]$, we have

$$p_i = \frac{1}{2N}\left[\lambda\left(O\left(\frac{m^2}{n}\right)\right) + \rho(O(m))\right].$$

With assumption (5), it is easy to check that $q_i = 0$ for $i \in [1, m]$ or $[n-m+1, n]$. Thus we have the following results for p_i and q_i:

$$g^{\mathrm{T}}D^2g = p^{\mathrm{T}}p = \sum_{i=1}^{n} p_i^2 = \frac{1}{4N^2}\left[\lambda^2 O\left(\frac{m^5}{n^2}\right) + \rho^2 O(m^3) + \lambda\rho O\left(\frac{m^4}{n}\right)\right], \tag{12}$$

$$h^{\mathrm{T}}D^2h = q^{\mathrm{T}}q = \sum_{i=1}^{n} q_i^2 = 2\gamma \sum_{l=1}^{m}\left(\sum_{k=l}^{m} c_k\right)^2, \tag{13}$$

$$g^{\mathrm{T}}D^2h = p^{\mathrm{T}}q = \sum_{i=1}^{n} p_i q_i = \frac{1}{4N^2}\left[\lambda^2 O\left(\frac{m^4}{n}\right) + \rho^2 O(m^3) + \lambda\rho O\left(\frac{m^4}{n}\right)\right]. \tag{14}$$

Note also that

$$\begin{aligned}\sum_{l=1}^{m}\left(\sum_{k=l}^{m} c_k\right)^2 = \frac{1}{4N^2}\Bigg[&\lambda^2\left(\frac{2}{15}m^3 + o(m^3)\right) + \lambda\rho\left(-\frac{1}{5}nm^2 + o(nm^2)\right)\\ &+ \rho^2\left(\frac{6}{5}n^2m + 12n^2 - \frac{7}{5}nm^2 + o(n^2) + o(nm^2)\right)\Bigg].\end{aligned}$$

Therefore, by Equations (12)–(14), we have (III).

For (IV), it is an immediate result from Lemma 7.1 (a), Equations (12) and (13). This finishes the proof of Lemma 7.2. ■

Proof of Theorem 3.1: By Lemma 7.2, we have the variance of $\hat{\theta}_{\mathrm{new}}(\lambda, \rho)$ as

$$\begin{aligned}&\mathrm{var}[\hat{\theta}_{\mathrm{new}}(\lambda, \rho)]\\ &= (\mathrm{var}(\varepsilon^2) - 2\sigma^4)\mathrm{tr}\{\mathrm{diag}(D)^2\} + 2\sigma^4\mathrm{tr}(D^2) + 4\sigma^2 f^{\mathrm{T}}D^2 f + 4\sigma^3\mu_3(f^{\mathrm{T}}D\mathrm{diag}(D)\mathbf{1})\\ &= \lambda^2\left[\frac{1}{n}\mathrm{var}(\varepsilon^2) + \frac{m}{15n^2}\mathrm{var}(\varepsilon^2) + \frac{4}{nm}\sigma^4 + \frac{4m}{15n^2}\sigma^2\gamma + o\left(\frac{1}{nm}\right) + o\left(\frac{m}{n^2}\right)\right]\\ &\quad + \rho^2\left[\left(\frac{3}{5m} - \frac{11}{10n}\right)\mathrm{var}(\varepsilon^2) + \left(\frac{12n}{m^3} - \frac{18}{m^2} - \frac{72n}{m^4}\right)\sigma^4\right.\\ &\quad \left. + \left(\frac{12}{5m} + \frac{24}{m^2} - \frac{2}{5n}\right)\sigma^2\gamma + o\left(\frac{1}{n}\right) + o\left(\frac{1}{m^2}\right) + o\left(\frac{n}{m^4}\right)\right]\\ &\quad + \lambda\rho\left[-\frac{1}{10n}\mathrm{var}(\varepsilon^2) - \frac{2}{5n}\sigma^2\gamma - \frac{12}{m^2}\sigma^4 + o\left(\frac{1}{n}\right) + o\left(\frac{1}{m^2}\right)\right].\end{aligned}$$

Since λ and ρ are arbitrary, $\mathrm{var}(\hat{\theta}_1)$ and $\mathrm{var}(\hat{\theta}_2)$ can be derived directly from the above result. Using the conditions in Theorem 3.1, we have

$$\mathrm{var}(\hat{\theta}_{\mathrm{new},1}) = \frac{1}{n}\mathrm{var}(\varepsilon^2) + \frac{m}{15n^2}\mathrm{var}(\varepsilon^2) + \frac{4m}{15n^2}\sigma^2\gamma + o\left(\frac{m}{n^2}\right), \tag{15}$$

$$\mathrm{var}(\hat{\theta}_{\mathrm{new},2}) = \frac{3}{5m}\mathrm{var}(\varepsilon^2) + \frac{12}{5m}\sigma^2\gamma + o\left(\frac{1}{m}\right). \tag{16}$$

Simple calculation shows that

$$\sum_{k=1}^{m} k^2 a_k w_k = -\frac{1}{6}m^2(1+o(1)), \quad \sum_{k=1}^{m} k^2 b_k w_k = nm(1+o(1))$$

and

$$\sum_{k=1}^{m} a_k w_k = 1, \quad \sum_{k=1}^{m} b_k w_k = 0, \quad \sum_{k=1}^{m} \frac{k a_k w_k}{n-k} = 0, \quad \sum_{k=1}^{m} \frac{k b_k w_k}{n-k} = 1.$$

This leads to the biases of $\hat{\theta}_{\mathrm{new},1}$ and $\hat{\theta}_{\mathrm{new},2}$ as

$$\begin{aligned} E(\hat{\theta}_{\mathrm{new},1}) &= \sum_{k=1}^{m} a_k w_k E(s_k) = \sum_{k=1}^{m} a_k w_k \left(\theta_1 + \frac{k}{(n-k)}\theta_2 + \frac{k^2}{2n^2}\delta + o\left(\frac{k^2}{n^2}\right)\right) \\ &= \theta_1 + \left[\delta \sum_{k=1}^{m} \frac{k^2 a_k w_k}{2n^2}\right](1+o(1)) \\ &= \theta_1 - \frac{m^2}{12n^2}\delta + o\left(\frac{m^2}{n^2}\right) \end{aligned} \tag{17}$$

and

$$\begin{aligned} E(\hat{\theta}_{\mathrm{new},2}) &= \sum_{k=1}^{m} b_k w_k E(s_k) = \sum_{k=1}^{m} b_k w_k \left(\theta_1 + \frac{k}{(n-k)}\theta_2 + \frac{k^2}{2n^2}\delta + o\left(\frac{k^2}{n^2}\right)\right) \\ &= \theta_2 + \left[\delta \sum_{k=1}^{m} \frac{k^2 b_k w_k}{2n^2}\right](1+o(1)) \\ &= \theta_2 + \frac{m}{2n}\delta + o\left(\frac{m}{n}\right). \end{aligned} \tag{18}$$

Finally, by Equations (15)–(18), we can derive the MSEs of $\hat{\sigma}^2_{\mathrm{new}}$ and $\hat{\gamma}_{\mathrm{new}}$ as in Theorem 3.1. The derivation of MSEs for $\hat{\sigma}^2_{\mathrm{MS}}$ and $\hat{\gamma}_{\mathrm{MS}}$ is similar and so is omitted. ■

7.2. Proof of Theorem 3.2

For simplicity, we give the proof for the case $\gamma = 0$ only as in [13]. The proof for the case $\gamma \neq 0$ is similar and so is omitted. To derive the asymptotic normality for $(\hat{\theta}_{\mathrm{new},1}, \hat{\theta}_{\mathrm{new},2})^{\mathrm{T}}$,

by the Cramér–Wold device it is sufficient to show that for any pair of $(\lambda,\ \rho) \in R^2$, the following X_n is asymptotically normal:

$$X_n = \lambda\sqrt{n}(\hat{\theta}_{\text{new},1} - E(\hat{\theta}_{\text{new},1})) + \rho\sqrt{m}(\hat{\theta}_{\text{new},2} - E(\hat{\theta}_{\text{new},2})).$$

Let $\alpha_k = w_k a_k$ and $\beta_k = w_k b_k$ for $1 \le k \le m$. We rewrite X_n as follows:

$$\begin{aligned}
X_n &= \lambda\sqrt{n}(\hat{\theta}_{\text{new},1} - E(\hat{\theta}_{\text{new},1})) + \rho\sqrt{m}(\hat{\theta}_{\text{new},2} - E(\hat{\theta}_{\text{new},2})) \\
&= \lambda\sqrt{n}\left(\sum_{k=1}^{m}\alpha_k s_k - \theta_1 + O\left(\frac{m^2}{n^2}\right)\right) + \rho\sqrt{m}\left(\sum_{k=1}^{m}\beta_k s_k + O\left(\frac{m}{n}\right)\right) \\
&= \sum_{k=1}^{m}(\lambda\sqrt{n}\alpha_k + \rho\sqrt{m}\beta_k)s_k - \lambda\sqrt{n}\sigma^2 + O(m^{3/2}/n) \\
&= \sum_{k=1}^{m}(\lambda\sqrt{n}\alpha_k + \rho\sqrt{m}\beta_k)(s_k - \sigma^2) + O(m^{3/2}/n) \\
&= \sum_{k=1}^{m}\omega_{k,n}\eta_k + O(m^{3/2}/n),
\end{aligned}$$

where $\omega_{k,n} = (\lambda\sqrt{n}\alpha_k + \rho\sqrt{m}\beta_k)$ and $\eta_k = s_k - \sigma^2$. For $\omega_{k,n}$, it is easy to show that

$$\sum_{k=1}^{m}\omega_{k,n} \sim \lambda\sqrt{n} \quad \text{and} \quad \sum_{k=1}^{m}\omega_{k,n}^2 = O\left(\frac{n^2}{m^2}\right).$$

For η_k, we divide it into two parts: $\eta_k = \tilde{\eta}_{k,n} + r_{k,n}$, where

$$\begin{aligned}
\tilde{\eta}_{k,n} &= \frac{1}{2(n-k)}\sum_{i=1}^{n-k}[(\varepsilon_{i+k} - \varepsilon_i)^2 - 2\sigma^2], \\
r_{k,n} &= \frac{1}{2(n-k)}\sum_{i=1}^{n-k}[(g_{i+k} - g_i)^2 + 2(g_{i+k} - g_i)(\varepsilon_{i+k} - \varepsilon_i)].
\end{aligned}$$

It is easy to check that

$$E\left(\sum_{k=1}^{m}\omega_{k,n}r_{k,n}\right) = O(m^{3/2}/n) \to 0. \tag{19}$$

In addition, we have

$$\begin{aligned}
\operatorname{cov}(r_{k,n}, r_{l,n}) &= \frac{1}{4(n-k)(n-l)}\sum_{i=1}^{n-k}\sum_{j=1}^{n-l}(g_{i+k} - g_i)(g_{j+l} - g_j)\operatorname{cov}(\varepsilon_{i+k} - \varepsilon_i, \varepsilon_{j+h} - \varepsilon_j) \\
&= O\left(\frac{m^3}{n^2}\right)\frac{\sigma^2}{4(n-k)(n-l)} \\
&= O\left(\frac{m^3}{n^4}\right).
\end{aligned}$$

Therefore,

$$\operatorname{var}\left(\sum_{k=1}^{m}\omega_{k,n}r_{k,n}\right)=\sum_{k,l=1}^{m}\omega_{k,n}\omega_{l,n}\operatorname{cov}(r_{k,n},r_{l,n})=O\left(\frac{m^3}{n^4}\right)\sum_{k,l=1}^{m}|\omega_{k,n}\omega_{l,n}|=o(1). \quad (20)$$

Define $\tilde{X}_n=\sum_{k=1}^{m}\omega_{k,n}\tilde{\eta}_{k,n}$. By Equations (19) and (20), we have

$$E(X_n-\tilde{X}_n)^2=E\left(\sum_{k=1}^{m}\omega_{k,n}r_{k,n}\right)^2=\operatorname{var}\left(\sum_{k=1}^{m}\omega_{k,n}r_{k,n}\right)+\left[E\left(\sum_{k=1}^{m}\omega_{k,n}r_{k,n}\right)\right]^2\to 0.$$

This shows that X_n and $\tilde{X}_n$ are asymptotically equivalent to each other. Now to derive the asymptotic normality of $\tilde{X}_n$, we introduce a further simplified statistic $\hat{T}_n$. Specifically, it is

$$\hat{T}_n=\sum_{\mu=1}^{n}E(\tilde{X}_n\mid\varepsilon_\mu).$$

By Lemma 7.4 as described below, we know that $\tilde{X}_n$ and $\hat{T}_n$ are also asymptotically equivalent to each other. Then by Lemma 7.5 and together with the Cramér–Wold device, we have

$$\begin{pmatrix}\sqrt{n}(\hat{\theta}_{\text{new},1}-\theta_1)\\ \sqrt{m}(\hat{\theta}_{\text{new},2}-\theta_2)\end{pmatrix}\xrightarrow{D}N\left(\begin{pmatrix}0\\0\end{pmatrix},\quad(\mu_4-\sigma^4)\begin{pmatrix}1&0\\0&3/5\end{pmatrix}\right)\quad\text{as } n\to\infty.$$

Finally, noting that $\hat{\sigma}^2_{\text{new}}=\hat{\theta}_{\text{new},1}$ and $\hat{\gamma}_{\text{new}}=2\hat{\theta}_{\text{new},2}$, the proof of Theorem 3.2 is finished.

Lemma 7.3: *Let $c_{v,n}=\sum_{k=1}^{v-1}\omega_{k,n}/(n-k)$. Assume that $m\to\infty$ and $m/n\to 0$. Then for $1\le v\le m$, we have*

(i) $c_{m+1,n}=\lambda/\sqrt{n}(1+o(1))$.
(ii) $\sum_{v=1}^{m}c_{v,n}=\lambda m/\sqrt{n}(1+o(1))-\rho\sqrt{m}(1+o(1))$.
(iii)

$$c_{v,n}=\lambda\frac{\sqrt{n}(v-1)(4m-3v)}{nm^2}(1+o(1))+\rho\frac{6\sqrt{m}(v-1)(v-m)}{m^3}(1+o(1)).$$

(iv)

$$\sum_{v=1}^{m}c_{v,n}^2=\frac{6}{5}\rho^2(1+o(1)).$$

(v) $\max_{1\le v\le m}\{|c_{v,n}|\}=O(m^{-1/2})$.

Proof of Lemma 7.3: This lemma can be readily achieved by Lemma 7.1 and the following four identities:

$$\sum_{k=1}^{v-1} \frac{\alpha_k}{n-k} = \frac{v-1}{mn}(1+o(1)) - \frac{m}{2n}\frac{6(v-1)(v-m)}{m^3}(1+o(1))$$
$$= \frac{(v-1)(4m-3v)}{nm^2}(1+o(1))$$

and

$$\sum_{k=1}^{v-1} \frac{\beta_k}{n-k} = \frac{1}{N(\sum_{k=1}^{m} w_k d_k^2 - \bar{d}_w^2)}\left[\frac{(v-1)^2}{2n}(1+o(1)) - \frac{m(v-1)}{2n}(1+o(1))\right]$$
$$= \frac{6(v-1)(v-m)}{m^3}(1+o(1))$$

and

$$\sum_{v=1}^{m} c_{v,n} = \lambda\sqrt{n}\sum_{v=1}^{m}\sum_{k=1}^{v-1}\frac{\alpha_k}{n-k} + \rho\sqrt{m}\sum_{v=1}^{m}\sum_{k=1}^{v-1}\frac{\beta_k}{n-k}$$
$$= \frac{\lambda\sqrt{n}m^3}{nm^2}(2-1)(1+o(1)) + \frac{6\rho\sqrt{m}m^3}{m^3}\left(\frac{1}{3}-\frac{1}{2}\right)(1+o(1))$$
$$= \lambda m/\sqrt{n}(1+o(1)) - \rho\sqrt{m}(1+o(1))$$

and

$$\sum_{v=1}^{m} c_{v,n}^2 = \sum_{v=1}^{m}\left(\lambda\sqrt{n}\sum_{k=1}^{v-1}\frac{\alpha_k}{n-k} + \rho\sqrt{m}\sum_{k=1}^{v-1}\frac{\beta_k}{n-k}\right)^2$$
$$= \sum_{v=1}^{m}\left(\lambda^2 n\left(\sum_{k=1}^{v-1}\frac{\alpha_k}{n-k}\right)^2 + \rho^2 m\left(\sum_{k=1}^{v-1}\frac{\beta_k}{n-k}\right)^2\right.$$
$$\left. + 2\lambda\rho\sqrt{mn}\left(\sum_{k=1}^{v-1}\frac{\alpha_k}{n-k}\right)\left(\sum_{k=1}^{v-1}\frac{\beta_k}{n-k}\right)\right)$$
$$= \frac{6}{5}\rho^2(1+o(1)).$$

■

Lemma 7.4: *For $\hat{T}_n$ and $\tilde{X}_n$, we have*

$$E(\hat{T}_n - \tilde{X}_n)^2 \to 0 \quad \text{as } n \to \infty.$$

Proof of Lemma 7.4: Note that $E(\hat{T}_n) = E(\tilde{X}_n) = 0$ and $\mathrm{var}(\hat{T}_n) \sim \mathrm{var}(\tilde{X}_n) = O(1)$. We have

$$\begin{aligned} E(\hat{T}_n - \tilde{X}_n)^2 &= E(\hat{T}_n^2) + E(\tilde{X}_n^2) - 2E(\hat{T}_n\tilde{X}_n) \\ &\sim 2[\mathrm{var}(\hat{T}_n) - E(\hat{T}_n\tilde{X}_n)]. \end{aligned}$$

For ease of notation, let $s_\mu^2 = \varepsilon_\mu^2 - \sigma^2$. We then rewrite $\hat{T}_n$ as follows:

$$\begin{aligned} \hat{T}_n &= \sum_{\mu=1}^{n}\sum_{k=1}^{m}\frac{\omega_{k,n}}{2(n-k)}\sum_{i=1}^{n-k}E[(\varepsilon_{i+k}-\varepsilon_i)^2 - 2\sigma^2 \mid \varepsilon_\mu] \\ &= \sum_{\mu=1}^{n}\sum_{k=1}^{m}\frac{\omega_{k,n}}{2(n-k)}\sum_{i=1}^{n-k}(\delta_{i+k,\mu}+\delta_{i,\mu})s_\mu^2 \\ &= \sum_{\mu=m+1}^{n-m}\left[\sum_{k=1}^{m}\frac{\omega_{k,n}}{n-k}\right]s_\mu^2 + \sum_{\mu=1}^{m}\left[\sum_{k=1}^{\mu-1}\frac{\omega_{k,n}}{n-k} + \sum_{k=\mu}^{m}\frac{\omega_{k,n}}{2(n-k)}\right]s_\mu^2 \\ &\quad + \sum_{\mu=n-m+1}^{n}\left[\sum_{k=1}^{n-\mu}\frac{\omega_{k,n}}{n-k} + \sum_{k=n-\mu+1}^{m}\frac{\omega_{k,n}}{2(n-k)}\right]s_\mu^2 \\ &= c_{m+1,n}\sum_{\mu=m+1}^{n-m}s_\mu^2 + \frac{1}{2}\sum_{\mu=1}^{m}(c_{m+1,n}+c_{\mu,n})s_\mu^2 + \frac{1}{2}\sum_{\mu=1}^{m}(c_{m+1,n}+c_{\mu,n})s_{n-\mu+1}^2 \\ &= c_{m+1,n}\sum_{\mu=m+1}^{n-m}s_\mu^2 + \frac{1}{2}\sum_{\mu=1}^{m}c_{\mu,n}(s_\mu^2+s_{n-\mu+1}^2) + \frac{c_{m+1,n}}{2}\sum_{\mu=1}^{m}(s_\mu^2+s_{n-\mu+1}^2). \end{aligned} \tag{21}$$

For $\hat{T}_n\tilde{X}_n$, with Equation (21), we have

$$\begin{aligned} E(\hat{T}_n\tilde{X}_n) &= E\left[\left(c_{m+1,n}\sum_{\mu=m+1}^{n-m}s_\mu^2 + \frac{1}{2}\sum_{\mu=1}^{m}c_{\mu,n}(s_\mu^2+s_{n-\mu+1}^2)\right.\right. \\ &\qquad \left.\left. + \frac{c_{m+1,n}}{2}\sum_{\mu=1}^{m}(s_\mu^2+s_{n-\mu+1}^2)\right)\tilde{X}_n\right] \\ &= c_{m+1,n}\sum_{\mu=m+1}^{n-m}E(s_\mu^2\tilde{X}_n) + \frac{1}{2}\sum_{\mu=1}^{m}c_{\mu,n}[E(s_\mu^2\tilde{X}_n) + E(s_{n-\mu+1}^2\tilde{X}_n)] \\ &\quad + \frac{c_{m+1,n}}{2}\sum_{\mu=1}^{m}[E(s_\mu^2\tilde{X}_n) + E(s_{n-\mu+1}^2\tilde{X}_n)]. \end{aligned}$$

Note that

$$\begin{aligned}
E(s_\mu^2 \tilde{X}_n) &= E(s_\mu^2 \sum_{k=1}^{m} \frac{\omega_{k,n}}{2(n-k)} \sum_{i=1}^{n-k} [(\varepsilon_{i+k} - \varepsilon_i)^2 - 2\sigma^2]) \\
&= \sum_{k=1}^{m} \frac{\omega_{k,n}}{2(n-k)} \sum_{i=1}^{n-k} E[s_\mu^2 (\varepsilon_{i+k} - \varepsilon_i)^2] \\
&= (\mu_4 - \sigma^4) \sum_{k=1}^{m} \frac{\omega_{k,n}}{2(n-k)} \sum_{i=1}^{n-k} (\delta_{\mu,i} + \delta_{\mu,i+k}).
\end{aligned}$$

For $m+1 \le \mu \le n-m$,

$$E(s_\mu^2 \tilde{X}_n) = c_{m+1,n}(\mu_4 - \sigma^4).$$

For $1 \le \mu \le m$ or $n-m+1 \le \mu \le n$,

$$E(s_\mu^2 \tilde{X}_n) = \tfrac{1}{2}(c_{m+1,n} + c_{\mu,n})(\mu_4 - \sigma^4).$$

Hence by Lemma 7.3,

$$\begin{aligned}
E(\hat{T}_n \tilde{X}_n) &= c_{m+1,n} \sum_{\mu=m+1}^{n-m} E(s_\mu^2 \tilde{X}_n) + \frac{1}{2} \sum_{\mu=1}^{m} c_{\mu,n} [E(s_\mu^2 \tilde{X}_n) + E(s_{n-\mu+1}^2 \tilde{X}_n)] \\
&\quad + \frac{c_{m+1,n}}{2} \sum_{\mu=1}^{m} [E(s_\mu^2 \tilde{X}_n) + E(s_{n-\mu+1}^2 \tilde{X}_n)] \\
&= (\mu_4 - \sigma^4) \left[(n-2m) c_{m+1,n}^2 + \frac{1}{2} \sum_{\mu=1}^{m} c_{\mu,n}^2 + c_{m+1,n} \sum_{\mu=1}^{m} c_{\mu,n} + \frac{m}{2} c_{m+1,n}^2 \right] \\
&\sim (\mu_4 - \sigma^4) \left(\lambda^2 + \frac{3}{5} \rho^2 \right) \\
&\sim \mathrm{var}(\hat{T}_n).
\end{aligned}$$

This shows that $E(\hat{T}_n \tilde{X}_n) \sim \mathrm{var}(\hat{T}_n)$ and so proves Lemma 7.4. ■

Lemma 7.5: *Assume that condition* (5) *holds,* $m \to \infty$ *and* $L = m = n^r$ *for* $1/2 < r < 2/3$. *For* $\hat{T}_n$, *we have*

$$\hat{T}_n \xrightarrow{D} N(0, \lambda^2(\mu_4 - \sigma^4) + \tfrac{3}{5}\rho^2(\mu_4 - \sigma^4)) \quad \text{as } n \to \infty.$$

Proof of Lemma 7.5: For $\hat{T}_n$, we apply the following notations for ease of presentation:

$$\hat{T}_n = c_{m+1,n} \sum_{\mu=m+1}^{n-m} s_\mu^2 + \frac{1}{2} \sum_{\mu=1}^{m} c_{\mu,n}(s_\mu^2 + s_{n-\mu+1}^2) + \frac{c_{m+1,n}}{2} \sum_{\mu=1}^{m} (s_\mu^2 + s_{n-\mu+1}^2)$$
$$= P_{1,n} + P_{2,n} + P_{3,n}.$$

For $P_{1,n}$, by the central limit theorem and Lemma 7.3 (i), we have

$$P_{1,n} \xrightarrow{D} N(0, \lambda^2(\mu_4 - \sigma^4)).$$

For $P_{2,n}$, we have the following fact,

$$\max_{1 \le v \le m} \{|c_{v,n}|\} = O(m^{-1/2}) \to 0.$$

This implies that $\{c_{\mu,n} s_\mu^2\}$ satisfy Lindeberg's condition [1]. Thus,

$$\sum_{\mu=1}^{m} c_{\mu,n} s_\mu^2 / ((\mu_4 - \sigma^4) \sum_{v=1}^{m} c_{v,n}^2)^{1/2} \xrightarrow{D} N(0, 1).$$

Note that $\sum_{v=1}^{m} c_{v,n}^2 \to 6\rho^2/5$. We have

$$\frac{1}{2} \sum_{\mu=1}^{m} c_{\mu,n} s_\mu^2 \xrightarrow{D} N\left(0, \frac{3}{10}\rho^2(\mu_4 - \sigma^4)\right).$$

By the same procedure, we have

$$\frac{1}{2} \sum_{\mu=1}^{m} c_{\mu,n} s_{n-\mu+1}^2 \xrightarrow{D} N\left(0, \frac{3}{10}\rho^2(\mu_4 - \sigma^4)\right).$$

For $P_{3,n}$, we have

$$\mathrm{Var}(P_{3,n}) = \frac{1}{2} m c_{m+1,n}^2 (\mu_4 - \sigma^4) = O\left(\frac{m}{n}\right) \to 0.$$

Finally, by Slutsky's theorem and the fact that $P_{1,n}$, $P_{2,n}$ and $P_{3,n}$ are independent of each other, we have

$$\hat{T}_n \xrightarrow{D} N(0, \lambda^2(\mu_4 - \sigma^4) + \tfrac{3}{5}\rho^2(\mu_4 - \sigma^4)). \qquad \blacksquare$$

Acknowledgments

The authors claim that the partial results in this paper follow the Chapter 5 of the first author's PhD dissertation at Hong Kong Baptist University [5], and they have not been published in any other scientific journals. The authors thank the editor, the associate editor and the referee for their constructive comments that led to a substantial improvement of the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Tiejun Tong's research was supported by the Hong Kong Baptist University grants FRG1/14-15/044, FRG2/15-16/019 and FRG2/15-16/038, and the National Natural Science Foundation of China grant (No. 11671338).

References

[1] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley, New York, 1995.
[2] A.W. Bowman, A. Pope, and B. Ismail, *Detecting discontinuities in nonparametric regression curves and surfaces*, Stat. Comput. 16 (2006), pp. 377–390.
[3] M.-Y. Cheng and M. Raimondo, *Kernel methods for optimal change-points estimation in derivatives*, J. Comput. Graph. Stat. 17 (2008), pp. 56–75.
[4] C.-K. Chu, J.-S. Siao, L.-C. Wang, and W.-S. Deng, *Estimation of 2D jump location curve and 3D jump location surface in nonparametric regression*, Stat. Comput. 22 (2012), pp. 17–31.
[5] W. Dai, *Different-based methods in nonparametric regression models*, Open Access Theses and Dissertations, Paper 40, 2014.
[6] I. Gijbels and A.-C. Goderniaux, *Bootstrap test for change-points in nonparametric regression*, J. Nonparametric Stat. 16 (2004), pp. 591–611.
[7] G. Grégoire and Z. Hamrouni, *Two non-parametric tests for change-point problems*, J. Nonparametric Stat. 14 (2002), pp. 87–112.
[8] J.-H. Joo and P. Qiu, *Jump detection in a regression curve and its derivative*, Technometrics 51 (2009), pp. 289–305.
[9] T. Kariya and H. Kurata, *Generalized Least Squares*, Wiley, Chichester, 2004.
[10] F.W. McElroy, *A necessary and sufficient condition that ordinary least-squares estimators be best linear unbiased*, J. Amer. Statist. Assoc. 62 (1967), pp. 1302–1304.
[11] H.-G. Müller, *Change-points in nonparametric regression analysis*, Ann. Stat. 20 (1992), pp. 737–761.
[12] H.G. Müller and U. Stadtmüller, *Detecting dependencies in smooth regression models*, Biometrika 75 (1988), pp. 639–650.
[13] H.-G. Müller and U. Stadtmüller, *Discontinuous versus smooth regression*, Ann. Stat. 27 (1999), pp. 299–337.
[14] N. Neumeyer and I. Van Keilegom, *Change-point tests for the error distribution in nonparametric regression*, Scand. J. Stat, 36 (2009), pp. 518–541.
[15] P. Qiu, *Image Processing and Jump Regression Analysis*, Wiley, New York, 2005.
[16] P. Qiu, *Jump surface estimation, edge detection, and image restoration*, J. Amer. Statist. Assoc. 102 (2007), pp. 745–756.
[17] P. Qiu and B. Yandell, *A local polynomial jump detection algorithm in nonparametric regression*, Technometrics 40 (1998), pp. 141–152.
[18] T. Tong, Y. Ma, and Y. Wang, *Optimal variance estimation without estimating the mean function*, Bernoulli 19 (2013), pp. 1839–1854.
[19] T. Tong and Y. Wang, *Estimating residual variance in nonparametric regression using least squares*, Biometrika 92 (2005), pp. 821–830.
[20] Y. Wang, *Jump and sharp cusp detection by wavelets*, Biometrika 82 (1995), pp. 385–397.