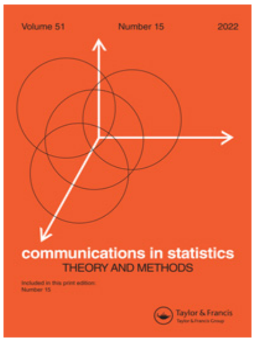

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Comparison of difference based variance estimators for partially linear models

Guoyi Zhang & Yan Lu

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2022.2064498

Taylor & Francis
Taylor & Francis Group

Check for updates

Comparison of difference based variance estimators for partially linear models

Guoyi Zhang and Yan Lu

Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico, USA

ABSTRACT
In this research, we evaluated two difference based variance estimators: one by Gasser, Sroka, and Jennen-Steinmetz, and another by Hall, Kay, and Titterington for use in partially linear models. Under various settings, we compared power of tests for heteroskedasticity, and other finite population properties of the estimators using simulation studies. We also proved that under regularity conditions, the estimator from Hall, Kay, and Titterington provides larger power of the tests for heteroskedasticity. A real example is given to illustrate the usage of the estimators.

ARTICLE HISTORY
Received 11 June 2021
Accepted 5 April 2022

1. Introduction

A partially linear model (PLM) is a regression function that contains a parametric linear component and a nonparametric component that involves an additional predictor variable. Suppose $y_1, ..., y_n$ are responses, and t is the predictor variable with $0 \le t_1 < t_2 \cdots < t_n \le 1$. A PLM can be written as follows:

$$y_i = \sum_{j=1}^{k} u_{ij}\gamma_j + f(t_i) + \epsilon_i, i = 1, ..., n \quad (1)$$

where $\gamma_1, ..., \gamma_k$ are unknown parameters; u_{ij}'s are known constants; f is an unknown smooth function; and $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. For example, both indexes: nasdaq composite and S&P 500 are in the same market environment. Their percentage of return (see detailed description in Section 4) can be modeled by the estimated smooth function $\hat{f}(\cdot)$ as the two green curves in Figure 1. Note that these two green curves are exactly the same, indicating the same market environment. The indexes' specific performance is modeled by $\sum_{j=1}^{k} u_{ij}\gamma_j$. Their difference is reflected by the different intercepts.

It is natural to ask questions such as: is the variation of the percentage of return over time a constant? Are the intercepts of the two indexes significantly different from each other? To answer these questions, we first need to estimate the variance σ^2. Difference-based variance estimators are commonly used in nonparametric regression. A benefit of

CONTACT Guoyi Zhang gzhang123@gmail.com Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131-0001, USA.

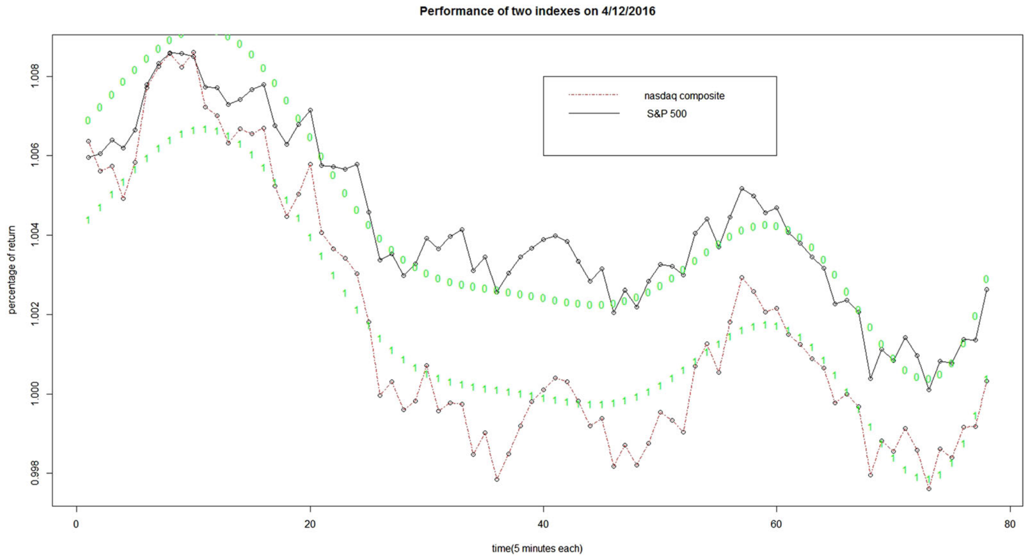

Figure 1. Performance of two indexes, modeled by partially linear models.

difference-based approaches is that the underlying function does not need to be estimated to obtain a useful variance estimator. Interested readers can refer to Von Neumann (1941), Rice (1984), Gasser, Sroka, and Jennen-Steinmetz (1986) (GSJS), Buckley, Eagleson, and Silverman (1988), Hall, Kay, and Titterington (1990) (HKT), Zhou et al. (2015), Dai et al. (2015), and Lu (2014) etc.

Methods for estimating PLM type models have been studied by Green, Jennison, and Seheuh (1985), Speckman (1988), Wahba (1990), etc. R. Eubank et al. (1998), and Klipple and Eubank (2006) extended work from Gasser, Sroka, and Jennen-Steinmetz (1986) to partially linear models. They developed efficient algorithms to compute fitted values, regression coefficients, standard errors, smoothing parameter selection criteria for the Speckman smoothing spline estimator, and proposed an estimator for the residual variance.

In this research, we apply GSJS and HKT variance estimators to PLMs using results by R. Eubank et al. (1998), and investigate their performance and properties. In Section 2, we review background knowledge related to the research. In Section 3, we perform simulation studies to compare performance of GSJS and HKT in PLMs, and to compare statistical power in testing heteroskedasticity using these two estimators. We also examine asymptotic properties of the test statistics. Section 4 gives a real example to illustrate usage of the variance estimators. Finally, Section 5 concludes the research.

2. Background

In this section, we briefly review GSJS and HKT variance estimators, and PLM estimators by R. Eubank et al. (1998) and Speckman (1988). In a matrix form, model (1) can be written as

$$\mathbf{y} = \mathbf{U}\gamma + \mathbf{f} + \epsilon, \tag{2}$$

where $\mathbf{y}$ is the vector of the responses $y_1, y_2, \ldots, y_n$, $\mathbf{U}$ is the specified design matrix, γ is the vector of unknown parameter γ_i's, $\mathbf{f} = (f(t_1), f(t_2), \ldots, f(t_n))'$, and ϵ is the vector of error terms. In model (2), we want to estimate the tuning parameter λ related to nonparametric function f, the expected value $\mu_\lambda = E(\mathbf{y})$, the parametric parameters γ, and variance σ^2.

2.1. GSJS difference based variance estimators for a simple nonparametric regression model

Gasser, Sroka, and Jennen-Steinmetz (1986) proposed second order difference based variance estimators for a simple nonparametric regression model as $\hat{\sigma}^2_{GSJS} = \sum_{i=1}^{n-2} \tilde{\epsilon}_i^2/(n-2)$, where $\tilde{\epsilon}_i$'s are called pseudo-residuals defined as

$$\tilde{\epsilon}_i = d_{i0}y_i + d_{i1}y_{i+1} + d_{i2}y_{i+2}, \tag{3}$$

with $d_{i0} = -a_i/\sqrt{1+a_i^2+b_i^2}$, $d_{i1} = 1/\sqrt{1+a_i^2+b_i^2}$, $d_{i2} = -b_i/\sqrt{1+a_i^2+b_i^2}$ for $a_i = (t_{i+2}-t_{i+1})/(t_{i+2}-t_i)$, and $b_i = (t_{i+1}-t_i)/(t_{i+2}-t_i)$.

The pseudo residuals $\tilde{\epsilon}_i^2$ are from straight line fits involving triples of points in lieu of successive differences. They can be considered as weighted average of the observations that are asymptotically free of response means. Gasser, Sroka, and Jennen-Steinmetz (1986) showed that when ϵ_i's are independent and identically distributed, $\sqrt{n}(\hat{\sigma}^2_{GSJS} - \sigma^2)$ has a limiting normal distribution.

The difference based variance estimator GSJS of σ^2 can be written in a matrix form as follows:

$$\hat{\sigma}^2_{GSJS} = \frac{\mathbf{y}^T\mathbf{A}^T_{GSJS}\mathbf{A}_{GSJS}\mathbf{y}}{n-2},$$

where

$$\mathbf{A}_{GSJS} = \begin{bmatrix} d_{10} & d_{11} & d_{12} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & d_{20} & d_{21} & d_{22} & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & d_{30} & d_{31} & d_{32} & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & d_{n-2,0} & d_{n-2,1} & d_{n-2,2} \end{bmatrix}.$$

2.2. HKT difference based variance estimators for a simple nonparametric regression model

Hall, Kay, and Titterington (1990) proposed a difference based variance estimator $\hat{\sigma}^2_{HKT}$ for use in a simple nonparametric regression model. Let m be the order of the sequence. A difference sequence $\{d_j\}$ is subject to the constraints $\sum_{j=1}^m d_j = 0$, and $\sum_{j=1}^m d_j^2 = 1$, where $d_j = 0$ for $j<0$ and $j>m$, and $d_0 d_m \neq 0$. The estimator based on order m sequence is as follows:

$$\hat{\sigma}^2_{HKT} = \frac{1}{n-m}\sum_{k=1}^{n-m}\left(\sum_{j=1}^{m} d_j y_{j+k}\right)^2,$$

or in a matrix form as

$$\hat{\sigma}^2_{HKT} = \frac{\mathbf{y}^T \mathbf{A}^T_{HKT} \mathbf{A}_{HKT} \mathbf{y}}{n-2}.$$

For a sequence with $m = 3$,

$$\mathbf{A}_{HKT} = \begin{bmatrix} d_0 & d_1 & d_2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & d_0 & d_1 & d_2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & d_0 & d_1 & d_2 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & d_0 & d_1 & d_2 \end{bmatrix}$$

with $d_0 = 0.8090, d_1 = -0.5,$ and $d_2 = -0.3090.$

The HKT optimal difference sequences provide substantial improvements over other commonly used sequences, and are free of unknown parameters. The efficiency of an optimal mth-order difference estimator relative to the error sample variance is $2m/(2m+1)$. Hall, Kay, and Titterington (1990) showed that the mean squared error (MSE) of $\hat{\sigma}^2_{HKT}$ is

$$MSE(\hat{\sigma}^2_{HKT}) = \frac{1}{n}\left(E(\epsilon_i^4) - \sigma^4 + \frac{\sigma^4}{m}\right) + o(n^{-1}).$$

Later, Dette, Munk, and Wagner (1998) derived a more detailed MSE formula for $\hat{\sigma}^2_{HKT}$ as follows:

$$\begin{aligned} MSE(\hat{\sigma}^2_{HKT}) =& \frac{1}{n}\left(E(\epsilon_i^4) - \sigma^4 + \frac{\sigma^4}{m}\right) \\ &+ \frac{(2m+1)^2(m+1)^2}{144n^2}\left(\|f'\|_2^4 + \frac{4\sigma^2}{n}\|f'\|_2^2\right) + o(n^{-5}). \end{aligned} \tag{4}$$

Equation (4) shows that the degree of smoothness of the nonparametric function f has an effect on estimating $MSE(\hat{\sigma}^2_{HKT})$.

2.3. Estimators of PLM

For the PLM model (2), R. Eubank et al. (1998) and Speckman (1988) suggested the following estimator for mean of $\mathbf{y}$:

$$\mu_\lambda = \mathbf{U}\gamma_\lambda + \mathbf{f}_\lambda, \tag{5}$$

with

$$\gamma_\lambda = (\mathbf{U}^T(\mathbf{I} - \mathbf{S}_\lambda)\mathbf{U})^{-1}\mathbf{U}^T(\mathbf{I} - \mathbf{S}_\lambda)\mathbf{y}, \tag{6}$$

and $\mathbf{f}_\lambda = \mathbf{S}_\lambda(\mathbf{y} - \mathbf{U}\gamma_\lambda),$ where $\mathbf{S}_\lambda = \mathbf{X}_\lambda(\mathbf{X}_\lambda^T\mathbf{X}_\lambda)^{-1}\mathbf{X}_\lambda^T, \mathbf{X}_\lambda = x_j(t_i)_{i=1,\dots,n, j=1,\dots,\lambda},$ x_j's are the basis functions, and λ is the tuning parameter to be estimated via generalized cross-validation (GCV) method.

R. Eubank et al. (1998) also suggested a difference based estimator of σ^2 for use in a PLM as follows:

$$\tilde{\sigma}^2 = \frac{\mathbf{y}^T \mathbf{A}^T (\mathbf{I} - \mathbf{P}) \mathbf{A} \mathbf{y}}{\mathrm{tr}(\mathbf{A}^T (\mathbf{I} - \mathbf{P}) \mathbf{A})}, \tag{7}$$

where $\mathbf{A}$ is a differencing matrix, and $\mathbf{P} = \mathbf{AU}(\mathbf{U}^T \mathbf{A}^T \mathbf{AU})^{-1} \mathbf{U}^T \mathbf{A}^T$. We apply differencing matrices $\mathbf{A}_{GSJS}$ and $\mathbf{A}_{HKT}$ for GSJS and HKT estimators respectively.

Under regularity conditions, R. Eubank et al. (1998) showed that

$$E(\tilde{\sigma}^2) = \sigma^2 + O(n^{-2}),$$

and

(1) $\sqrt{n}(\gamma_{\lambda_{opt}} - \gamma) \xrightarrow{d} N(\mathbf{0}, \mathbf{V})$. If $\gamma_\lambda = \mathbf{C}_\lambda \mathbf{y}$, the confidence interval for the ith unknown parameter γ_i is

$$\gamma_{i\hat{\lambda}} \pm 2\tilde{\sigma}^2 \sqrt{\left[\mathbf{C}_{\hat{\lambda}} \mathbf{C}_{\hat{\lambda}}^T\right]_{ii}}.$$

(2)

$$\sqrt{n}(\tilde{\sigma}^2 - \sigma^2) = Z_n + O_p\left(n^{-\frac{1}{2}}\right)$$

where $\tau^{-\frac{1}{2}} Z_n \xrightarrow{L} N(0, 1)$ (Klipple 2000) and

$$\tau = \sigma_{00} + \sum_{c=1}^{m} 2 \cdot \frac{(n - m - c)}{n - m} \sigma_{0c} \tag{8}$$

$$\sigma_{00} = \mathrm{Var}\left(\tilde{\epsilon}_i^2 | \mathrm{Var}(\epsilon_i) = \sigma^2\right) = E(\epsilon_1^4) \sum_{i=0}^{m} d_i^4 + 6\sigma^4 \sum_{i=0}^{m-1} \sum_{j=i+1}^{m} d_i^2 d_j^2 - \sigma^4 \tag{9}$$

$$\begin{aligned} \sigma_{0c} =& \mathrm{Cov}\left(\tilde{\epsilon}_i^2, \tilde{\epsilon}_{i+c}^2 | \mathrm{Var}(\epsilon_i) = \sigma^2\right) \\ =& (E\epsilon_1^4 - \sigma^4) \sum_{i=0}^{m-c} d_i^2 d_{i+c}^2 + 4\sigma^4 \sum_{i=0}^{m-c-1} \sum_{j=i+1}^{m-c} d_i d_j d_{i+c} d_{j+c} \text{ for } c = 1, \ldots, m. \end{aligned} \tag{10}$$

3. Comparison of difference based variance estimators (GSJS and HKT) in partially linear models

In this section, we conduct simulation studies to evaluate performance of GSJS and HKT in PLM under various settings. We also compare power of test for heteroskedasticity using GSJS and HKT estimators, and prove that HKT provides larger power of the tests for heteroskedasticity under certain conditions.

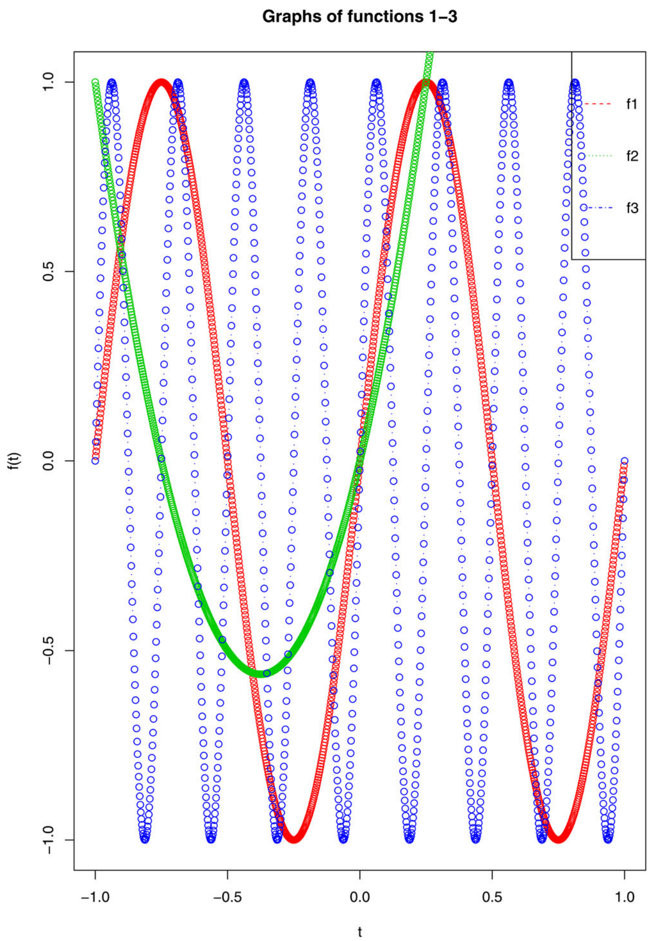

Figure 2. Plots of f_1, f_2, and f_3.

3.1. Performance comparison of GSJS and HKT

The partial linear model (1) for the two-group case can be written as follows

$$\begin{bmatrix} y_{1t_1} \\ y_{1t_2} \\ \vdots \\ y_{1t_{n_1}} \\ y_{2t_1} \\ y_{2t_2} \\ \vdots \\ y_{2t_{n_2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} f(1t_1) \\ f(1t_2) \\ \vdots \\ f(1t_{n_1}) \\ f(2t_1) \\ f(2t_2) \\ \vdots \\ f(2t_{n_2}) \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t_1} \\ \varepsilon_{1t_2} \\ \vdots \\ \varepsilon_{1t_{n_1}} \\ \varepsilon_{2t_1} \\ \varepsilon_{2t_2} \\ \vdots \\ \varepsilon_{2t_{n_2}} \end{bmatrix}. \tag{11}$$

Simulation study is performed with factors: (1) γ: the difference between γ_1 and γ_2 is set to be 0.5 or 1; (2) population standard deviation $\sigma = 0.4$ or 0.8; (3) sample size for the two groups $(n_1, n_2) = (75, 75), (n_1, n_2) = (25, 25)$; (4) functions considered $f_1(t) = sin\,(2\pi t), f_2(t) = 3t + 4t^2,$ and $f_3(t) = sin\,(8\pi t)$. Figure 2 shows that function f_3 has the

Table 1. Simulation results using function f_1 with different sizes $n_1 = 75, n_2 = 75$ and $n_1 = 25, n_2 = 25$. Index "G" represents GSJS, and index "H" represents HKT. Numbers reported in the table are the average, standard error (se), mean squared error (MSE) and bias square ($Bias^2$) of the GSJS and HKT variance estimators.

		$n_1 = 75, n_2 = 75$							
γ	σ	$\hat{\sigma}_G^2$	$se(\hat{\sigma}_G^2)$	MSE_G	$Bias_G^2$	$\hat{\sigma}_H^2$	$se(\hat{\sigma}_H^2)$	MSE_H	$Bias_H^2$
0.5	0.4	0.1604	0.0269	0.0007	0.0000	0.1612	0.0211	0.0004	0.0000
0.5	0.8	0.6425	0.1029	0.0106	0.0000	0.6430	0.0818	0.0067	0.0000
1	0.4	0.1602	0.0253	0.0006	0.0000	0.1609	0.0206	0.0004	0.0000
1	0.8	0.6398	0.1058	0.0111	0.0000	0.6427	0.0845	0.0071	0.0000
		$n_1 = 25, n_2 = 25$							
γ	σ	$\hat{\sigma}_G^2$	$se(\hat{\sigma}_G^2)$	MSE_G	$Bias_G^2$	$\hat{\sigma}_H^2$	$se(\hat{\sigma}_H^2)$	MSE_H	$Bias_H^2$
0.5	0.4	0.1601	0.0464	0.0021	0.0000	0.1688	0.0374	0.0014	0.0000
0.5	0.8	0.6403	0.1828	0.0334	0.0000	0.6472	0.1464	0.0214	0.0000
1	0.4	0.1600	0.0461	0.0021	0.0000	0.1690	0.0372	0.0014	0.0000
1	0.8	0.6379	0.1850	0.0342	0.0000	0.6469	0.1452	0.0211	0.0000

Table 2. Simulation results using function f_2 with different sizes $n_1 = 75, n_2 = 75$ and $n_1 = 25, n_2 = 25$. Index "G" represents GSJS, and index "H" represents HKT.

		$n_1 = 75, n_2 = 75$							
γ	σ	$\hat{\sigma}_G^2$	$se(\hat{\sigma}_G^2)$	MSE_G	$Bias_G^2$	$\hat{\sigma}_H^2$	$se(\hat{\sigma}_H^2)$	MSE_H	$Bias_H^2$
0.5	0.4	0.1592	0.0264	0.0007	0.0000	0.1627	0.0206	0.0004	0.0000
0.5	0.8	0.6349	0.1045	0.0109	0.0000	0.6398	0.0842	0.0070	0.0000
1	0.4	0.1597	0.0256	0.0006	0.0000	0.1631	0.0206	0.0004	0.0000
1	0.8	0.6403	0.1031	0.0106	0.0000	0.6426	0.0832	0.0069	0.0000
		$n_1 = 25, n_2 = 25$							
γ	σ	$\hat{\sigma}_G^2$	$se(\hat{\sigma}_G^2)$	MSE_G	$Bias_G^2$	$\hat{\sigma}_H^2$	$se(\hat{\sigma}_H^2)$	MSE_H	$Bias_H^2$
0.5	0.4	0.1594	0.0466	0.0021	0.0000	0.1874	0.0374	0.0021	0.0007
0.5	0.8	0.6378	0.1868	0.0348	0.0000	0.6654	0.1477	0.0224	0.0006
1	0.4	0.1599	0.0463	0.0021	0.0000	0.1875	0.0374	0.0021	0.0007
1	0.8	0.6377	0.1858	0.0345	0.0000	0.6651	0.1474	0.0223	0.0006

highest volatility. We expect that the two terms $\|f_3'\|_2^4$ and $\|f_3'\|_2^2$ in equation (4) are large, i.e., bias is not negligible. Functions f_1 and f_2 are relatively smooth.

For the given group size and parameter configuration, we generate $L = 2000$ samples according to model (11). For each sample, we calculate $\hat{\sigma}^2_{GSJS}$, $\hat{\sigma}^2_{HKT}$, and report averages of $\hat{\sigma}^2_{GSJS}$ and $\hat{\sigma}^2_{HKT}$, standard sampling variability $se(\hat{\sigma}^2)$, $Bias^2$ and MSE from the 2000 simulations.

Tables 1–3 give the simulation results. From Tables 1–3, we can see that variance estimates $\hat{\sigma}^2_{GSJS}$ and $\hat{\sigma}^2_{HKT}$ are close to each other under different settings. $se(\hat{\sigma}^2)$ by HKT are smaller than those of GSJS since HKT provides optimal difference sequences for estimating error variance. For functions f_1 and f_2, biases are almost close to 0, meaning that MSE of functions f_1 and f_2 are almost the same as $var(\hat{\sigma}^2)$. From Table 3, with the highest volatility function f_3 and large sample size of 75, $Bias^2$ is very small compared to $var(\hat{\sigma}^2)$ for both GSJS and HKT. Theoretically, for GSJS estimator, R. L. Eubank (1999, page 49) stated, "GSJS estimator is $\sqrt{n}$-consistent in that $\hat{\sigma}^2 - \sigma^2 = O_p(n^{-1/2})$, in other words, bias goes to 0 when n goes to infinity. For HKT estimator, theorem from Hall, Kay, and Titterington (1990, page 526) stated that $\text{var}(\hat{\sigma}^2) \sim E(\hat{\sigma}^2 - \sigma^2)^2 \sim n^{-1}\tau^2$ as n goes to infinity.

Table 3. Simulation results using function f_3 with different sizes $n_1 = 75, n_2 = 75$ and $n_1 = 25, n_2 = 25$. Index "G" represents GSJS, and index "H" represents HKT.

		$n_1 = 75, n_2 = 75$							
γ	σ^2	$\hat{\sigma}_G^2$	$se(\hat{\sigma}^2)$	MSE_G	$Bias_G^2$	$\hat{\sigma}_H^2$	$se(\hat{\sigma}_H^2)$	MSE_H	$Bias_H^2$
0.5	0.16	0.1593	0.00068	0.0006	0.0000	0.1774	0.00046	0.0007	0.0003
0.5	0.64	0.6378	0.0113	0.0113	0.0000	0.6555	0.0072	0.0074	0.0002
1	0.16	0.1613	0.00068	0.0006	0.0000	0.1784	0.00042	0.0007	0.0003
1	0.64	0.6392	0.0118	0.0117	0.0000	0.6558	0.0074	0.0076	0.0002
		$n_1 = 25, n_2 = 25$							
γ	σ^2	$\hat{\sigma}_G^2$	$se(\hat{\sigma}_G^2)$	MSE_G	$Bias_G^2$	$\hat{\sigma}_H^2$	$se(\hat{\sigma}_H^2)$	MSE_H	$Bias_H^2$
0.5	0.16	0.1677	0.0022	0.0022	0.0000	0.3122	0.0022	0.0254	0.0231
0.5	0.64	0.6509	0.0359	0.0360	0.0001	0.7950	0.0259	0.0499	0.0240
1	0.16	0.1681	0.0022	0.0022	0.0000	0.3126	0.0023	0.0255	0.0232
1	0.64	0.6467	0.0343	0.0343	0.0000	0.7908	0.0253	0.0480	0.0227

Table 4. True MSE calculated for $\sigma = 0.4$ and $\sigma = 0.8$ by τ/n for GSJS and HKT estimators under different settings, where τ is obtained by equation (8).

	GSJS		HKT	
n	$n = 50$	$n = 150$	$n = 50$	$n = 150$
$\sigma = 0.4$	0.001970	0.00066	0.00128	0.000426
$\sigma = 0.8$	0.03152	0.01058	0.02048	0.006826

However, with medium sample size of 25, bias of HKT is not negligible. For example, for the setting with $\gamma = 0.5$ and $\sigma^2 = 0.16$, bias is dominating with $Bias^2 = 0.0231$ compared to $var(\hat{\sigma}^2) = 0.0022^2 = 4.84 * 10^{-6}$. This can be explained by MSE formula (4) of HSK estimator, in which bias term involves first derivative of function f. When function f is with great volatility, bias term is not negligible, or sometimes is dominating. HKT may not be suitable for high volatility functions with medium sized sample. On the other hand, GSJS performs well with medium group size and function f_3.

Table 4 listed MSEs calculated by τ/n, where τ is defined in equation (8). These are treated as true MSE. We can see that MSE from the simulation studies are close to the MSE in Table 4.

3.2. Testing heteroskedasticity

In this section, we first stated two theorems. One theorem is about the asymptotic distribution of test statistic T in equation (12) under $H_0 : \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, and the other is regarding the statistical power of HKT and GSJS when testing for heteroskedasticity. Next, we did simulation studies on the power of test for heteroskedasticity.

3.2.1. Theorems

Theorem 1. *For a PLM under $H_0 : \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, assume that the nonparametric function $f \in C^2[0, 1]$ ($C^2[0, 1]$ denotes the set of all functions on $[0, 1]$ with second order continuous derivatives), $\max|t_i - t_{i-1}| = O(n^{-1})$, and $\mathbf{f}'\mathbf{A}'\mathbf{A}\mathbf{f} = O(n^{-1})$, the test statistic T converges to a standard normal distribution, i.e.,*

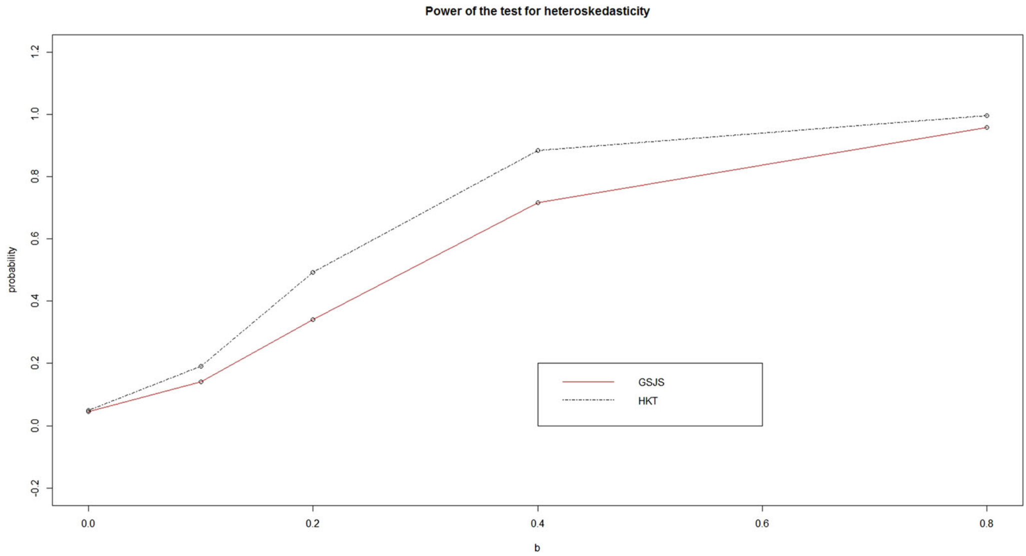

Figure 3. Power of tests for heteroskedasticity using function f_1, and variance estimators GSJS and HKT.

$$T = \frac{\sum_{i=1}^{n-m}(t_i - \bar{t})(\tilde{\epsilon}_i^2 - \sigma^2)}{\sqrt{\tau \sum_{i=1}^{n-m}(t_i - \bar{t})^2}} \xrightarrow{L} N(0,1), \tag{12}$$

where τ is defined in equation (8).

Proof is given in Appendix.

Theorem 2. *Consider $H_0 : \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, under regularity conditions given in Theorem 1, HKT using optimal difference sequences will provide larger power of the tests compared to GSJS.*

Proof is given in Appendix.

3.2.2. Simulation studies on testing heteroskedasticity using GSJS and HKT

In this section, we perform simulations to study power of testing heteroskedasticity by using GSJS and HKT variance estimators. By Theorem 1, under $H_0 : \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, test statistic T converges to a standard normal distribution.

Let $\sigma = 0.4 + bt$, where b is a coefficient that controls the magnitude of deviations from $\sigma = 0.4$. Consider $b = (0, 0.1, 0.2, 0.4, 0.8)$, $n_1 = 75, n_2 = 75$, and functions f_1, f_2 and f_3 in the simulation. We did $L = 2000$ simulations for each setting, and reported power of the tests in Figures 3–5. From Figures 3–5, we can see that tests by using GSJS and HKT variance estimators both control type I error well. Power of tests by using HKT estimator is larger than that by using GSJS estimator. In addition, when deviation from σ is large enough such that $b = 0.8$, power of tests by using both estimators approaches 1.

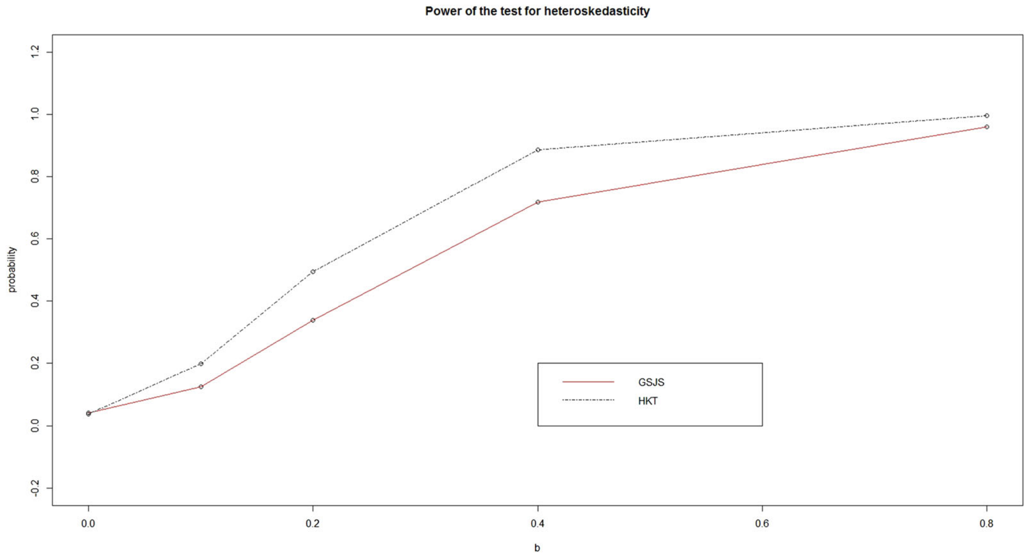

Figure 4. Power of tests for heteroskedasticity using function f_2, and variance estimators GSJS and HKT.

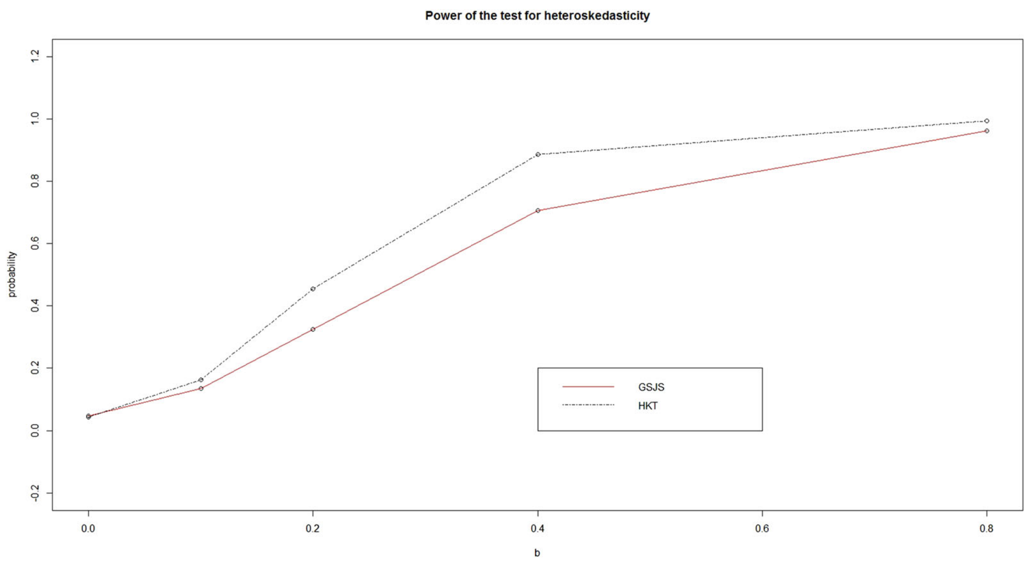

Figure 5. Power of tests for heteroskedasticity using function f_3, and variance estimators GSJS and HKT.

4. Application

In this section, we illustrate usage of the GSJS and HKT estimators in PLM using a real data example. Consider two indexes: nasdaq composite and S&P 500, which include different companies. To make the results comparable, we define percentage of return as the market value of the index every five minutes divided by the closing market value from the previous day. There are 78 data points collected during a six and half hour trading day. PLM model (11) is used for this example. Both indexes are in the same market environment which is modeled by the nonparametric function $f(\cdot)$. On the other hand, the two indexes include different public traded companies in open market, therefore the indexes have different performance that is modeled by γ_1 and γ_2.

In general, the two indexes have similar shapes and trends, but vary over time. Is the variation of percentage of return over time a constant? Is the percentage of return of the two indexes significantly different from each other? To answer these questions, we use GSJS and HKT to estimate the variance and perform tests of heteroskedasticity and difference of the means.

Using equation (6), the estimated difference of the slopes is $\hat{\gamma} = 0.001258$. Replace $\mathbf{A}$ by $\mathbf{A}_{GSJS}$ in equation (7), the GSJS variance estimate is calculated as $\hat{\sigma}^2_{GSJS} = 7.759 * 10^{-7}$. Similarly, replace $\mathbf{A}$ by $\mathbf{A}_{HKT}$, the HKT variance estimate is $\hat{\sigma}^2_{HKT} = 6.059 * 10^{-7}$.

To test heteroskedasticity, we obtain the test statistics by equation (12). Using GSJS estimate, $T_{GSJS} = -4.13$, and using HKT estimate, $T_{HKT} = -3.81$. Since T is asymptoticly normal under H_0, we compare T_{GSJS} and T_{HKT} to the critical value $Z_{0.975} = 1.96$, and reject the null hypothesis of constant variance. We conclude that at 5% significance level, the variations of percentage of return of the two indexes (nasdaq composite and S&P 500) on April 20th 2016 are not constants.

To test the difference γ of means γ_1 and γ_2, we first calculate standard er- rors. By the asymptotic normality property (1) in Section 2.3, standard error of $\hat{\gamma}$ from GSJS is $\hat{\sigma}_{GSJS}\sqrt{[\mathbf{C}_{\hat{\lambda}}\mathbf{C}_{\hat{\lambda}}^T]_{ii}} = 0.0000705$, and from HKT is $\hat{\sigma}_{HKT}\sqrt{[\mathbf{C}_{\hat{\lambda}}\mathbf{C}_{\hat{\lambda}}^T]_{ii}} = 0.0000623$. Test statistics are $T_{GSJS} = 0.001258/0.0000705 = 17.84$, and $T_{HKT} = 0.001257/0.0000623 = 20.17$. Compared to critical value of $Z_{0.975} = 1.96$, we reject the null hypothesis, and conclude that the mean percentage of return of two indexes are significantly different from each other.

5. Conclusions

In this research, we studied two difference based variance estimators GSJS and HKT in PLMs. Simulation studies show that both GSJS and HKT work well in estimating σ^2 for PLMs under various settings. HKT tends to have smaller MSE compared to GSJS for large samples. This is because HKT provides optimal difference sequences for estimating error variances. For medium and small samples with high volatility function such as function f_3, bias of HKT is not negligible. Therefore, we would suggest using GSJS for variance estimation for medium and small samples with high volatility function. In testing heteroskedasticity, HKT always provides larger power than GSJS.

Acknowledgements

The authors thank the editor and referee for their helpful comments and constructive suggestions to improve the manuscript.

References

Buckley, M., G. Eagleson, and B. Silverman. 1988. The estimation of residual variance in nonparametric regression. *Biometrika* 75 (2):189–99. doi:10.1093/biomet/75.2.189.

Dai, W., Y. Ma, T. Tong, and L. Zhu. 2015. Difference-based variance estimation in nonparametric regression with repeated measurement data. *Journal of Statistical Planning and Inference* 163:1–20. doi:10.1016/j.jspi.2015.02.010.

Dette, H., A. Munk, and T. Wagner. 1998. Estimating the variance in nonparametric regression - what is a reasonable choice? *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60 (4):751–64. doi:10.1111/1467-9868.00152.

Eubank, R. L. 1999. *Nonparametric regression and spline smoothing*. Boca Raton, Florida: CRC Press.

Eubank, R., E. Kambour, J. Kim, K. Klipple, C. Reese, and S. M. 1998. Estimation in partially linear models. *Computational Statistics & Data Analysis* 29 (1):27–34. doi:10.1016/S0167-9473(98)00054-1.

Gasser, T., L. Sroka, and C. Jennen-Steinmetz. 1986. Residual variance and residual pattern in nonlinear regression. *Biometrika* 73 (3):625–33. doi:10.1093/biomet/73.3.625.

Green, P., C. Jennison, and A. Seheuh. 1985. Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society: Series B* 47:299–315.

Hall, P., J. W. Kay, and D. M. Titterington. 1990. Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* 77 (3):521–8. doi:10.1093/biomet/77.3.521.

Klipple, K. 2000. Error variance estimation and testing for homoscedasticity in partially linear models. *Dissertation*.

Klipple, K., and R. L. Eubank. 2006. Difference based variance estimators for partially linear models. In *Festschrift in Honor of Distinguished Professor Mir Masoom Ali On the Occasion of his Retirement, Muncie, IN, USA. 18–19 May, 2007*, 313–23.

Lu, Y. 2014. Difference based variance estimator for nonparametric regression in complex survey. *Journal of Statistical Computation and Simulation* 84 (2):335–43. doi:10.1080/00949655.2012.708344.

Rice, J. 1984. Bandwidth choice for nonparametric regression. *The Annals of Statistics* 12 (4): 1215–30. doi:10.1214/aos/1176346788.

Speckman, P. 1988. Kernel smoothing in partial linear mocels. *Journal of the Royal Statistical Society, Series B* 50:413–36.

Von Neumann, J. 1941. Distribution of the ratio of the mean squared successive difference to the variance. *The Annals of Mathematical Statistics* 12 (4):367–95. doi:10.1214/aoms/1177731677.

Wahba, G. 1990. *Spline models for observational data*. Philadelphia, PA: SIAM.

Zhou, Y., Y. Cheng, L. Wang, and T. Tong. 2015. Optimal difference-based variance estimation in heteroscedastic nonparametric regression. *Statistica Sinica* 25:1377–97.

Appendix

Proof of Theorem 1 follows a similar argument as that of Theorem 2 by Klipple (2000).

Proof of Theorem 2 is as follows:

Proof. Let σ_i^2 be the variance of ϵ_i under alternative hypothesis. Without loss of generality, assume that $\sigma_i^2 = \sigma^2 g(t_i)$, where $g(t_i)$ is a smooth non-constant positive function.

Follow a similar argument as that of Lemma 3.4 (Klipple 2000), we have

$$E\left(\sum_{i=1}^{n-m}(t_i - \bar{t})\tilde{\epsilon}_i^2\right) = \sum_{i=1}^{n-m}(t_i - \bar{t})g(t_i)\sigma^2. \tag{13}$$

By the fact that $g(\cdot)$ is a smooth function and $\max|t_i - t_{i-1}| = O(n^{-1})$,

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{n-m}(t_i-\bar{t})\tilde{\epsilon}_i^2\right) &= \sum_{i=1}^{n-m}(t_i-\bar{t})^2 g^2(t_i)\sigma_{00} + 2\sum_{c=1}^{m}\sum_{k=1}^{n-m-c}(t_k-\bar{t})(t_{k+c}-\bar{t})g(t_k)g(t_{k+c})\sigma_{0c} \\ &= \sum_{i=1}^{n-m}(t_i-\bar{t})^2 g^2(t_i)\left(\sigma_{00}+2\sum_{c=1}^{m}\sigma_{0c}\right)+o(n) \\ &= \sum_{i=1}^{n-m}(t_i-\bar{t})^2 g^2(t_i)\tau + o(n),\end{aligned} \tag{14}$$

where σ_{00}, σ_{0c} are defined in equations (9) and (10) respectively.

Recall that the test statistic is $T = \sum_{i=1}^{n-m}(t_i-\bar{t})(\tilde{\epsilon}_i^2-\sigma^2)/\sqrt{\tau\sum_{i=1}^{n-m}(t_i-\bar{t})^2}$. By equation (13),

$$E(T|\sigma_i^2 = \sigma^2 g(t_i)) = E\left(\frac{\sum_{i=1}^{n-m}(t_i-\bar{t})\tilde{\epsilon}_i^2}{\sqrt{\tau\sum_{i=1}^{n-m}(t_i-\bar{t})^2}}\right) \tag{15}$$

$$= \frac{\sigma^2\sum_{i=1}^{n-m}(t_i-\bar{t})g(t_i)}{\sqrt{\tau\sum_{i=1}^{n-m}(t_i-\bar{t})^2}} \tag{16}$$

Since $\tau_{HKT} \le \tau_{GSJS}$, if t and $g(t)$ are positively correlated, i.e., $\sum_{i=1}^{n-m}(t_i-\bar{t})(g(t_i)-\bar{g}(t_i)) > 0$, we have

$$E(T_{HKT}|\sigma_i^2) \ge E(T_{GSJS}|\sigma_i^2) \ge 0; \tag{17}$$

On the other hand, if if t and $g(t)$ are negatively correlated,

$$E(T_{HKT}|\sigma_i^2) \le E(T_{GSJS}|\sigma_i^2) \le 0; \tag{18}$$

Now we want to prove that asymptotically the sampling distribution of T_{HKT} and T_{GSJS} has same variance. By result from equation (14),

$$\begin{aligned}\mathrm{Var}(T) &= \frac{1}{\tau\sum_{i=1}^{n-m}(t_i-\bar{t})^2}\sum_{i=1}^{n-m}(t_i-\bar{t})^2 g^2(t_i)\tau + o(1) \\ &= \frac{\sum_{i=1}^{n-m}(t_i-\bar{t})^2 g^2(t_i)}{\sum_{i=1}^{n-m}(t_i-\bar{t})^2} + o(1)\end{aligned}$$

Thus, asymptotically, the sampling distribution of T has constant variance for HKT and GSJS. The statistical power of the tests are determined by mean of the test statistics. By equations (17) and (18), $|E(T_{HKT}|\sigma_i^2)| \ge |E(T_{GSJS}|\sigma_i^2)|$, so that method HKT leads to more rejection, thus provides higher statistical power compared to method GSJS.