Hacettepe Journal of Mathematics and Statistics
Volume 44 (1) (2015), 239 – 245

A new nonparametric estimation method of the variance in a heteroscedastic model

Xiaoqin Zhang*[†] [‡], Hongxia Hao[§] and Jiye Liang[¶]

Abstract

In most economic phenomena, the assumption of homoscedasticity in the classic linear regression model is not necessarily true, which leads to heteroscedasticity. The heteroscedastic estimate is an important aspect for the problem of heteroscedasticity. For this hot issue, this paper proposes a nonparametric estimation method with simple calculation for the estimation of heteroscedasticity through orthogonal arrays, which does not rely on the distribution of data. The performance of the proposed method is investigated by prediction error in real data sets and simulations. The results suggest that this method offers substantial improvements over the existing tests.

2000 AMS Classification: 37L15, 37M05, 37M25

Keywords: Orthogonal array; heteroscedasticity; non-parametric estimation

Received 17/07/2013 : Accepted 08/01/2014 Doi: 10.15672/HJMS.201417452

1. Introduction

In model analysis in many areas such as sociology, economics and technology, homoscedastic assumption in classic linear regression model is not necessarily true. That is to say, the variance of random error term changes with the observed values. This model is called a heteroscedasticity model[1]. What leads to the heteroscedasticity? One reason is because the random error term includes the measurement error and the impact of some factors omitted in the model on the dependent variable, on the other hand, the value of the dependent variable in different sampling unit may be very different. If we use ordinary least squares (OLS) to estimate the parameters under heteroscedasticity model,

*School of Mathematical Science, Shanxi University, Taiyuan 030006, Shanxi, P.R. China.
†Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan, 030006, Shanxi, P.R. China.
‡Corresponding Author. Email: zhangxiaoqin@sxu.edu.cn
§Department of Mathematics, Southeast University, Nanjing, 210096, Jiangsu, P.R. China. Email: hong_xia_mm@163.com
¶Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan, 030006, Shanxi, P.R. China. Email: ljy@sxu.edu.cn

it will have serious consequences, for example, the estimated variance of the parameter does not have the validity (i.e. minimum variance), although with no bias, significant test and interval estimation will draw the wrong statistical inference, which directly leads to the decline of the accuracy and the prediction accuracy[1][2]. Therefore, the research for heteroscedastic linear regression models is of great significance.

Currently, there are many ways to solve the problem of heteroscedasticity. Early, the weighted least squares method can be used in the situation that the variance is not constant[3], but it often requires the mean and covariances of the dependent variable satisfy the linear relationship except for known variance. In general, it is very difficult to meet these two requirements. Thus literature [4] studied the model $y_i = m(x_i) + \sigma(x_i)\varepsilon_i$ and estimated the unknown function $m(\cdot)$ and $\sigma(\cdot)$, but their method can only handle the situation where the covariate x_i is one-dimensional. When the covariate is high-dimensional, the paper [5] discussed the heteroscedastic model $Y_i = m(X_i\beta)+\sigma(X_i\beta, \theta)\varepsilon_i$ where $m(\cdot)$ and $\sigma(\cdot)$ are known. A high-dimensional X_i is projected to a direction of where $X_i\beta$ is and the function $m(\cdot)$, $\sigma(\cdot)$ are changed to unary function. Thereby, the paper solved the problem of dimensionality reduction. The model requires a known contact function $m(\cdot)$ and a variance affected by the mean, however, these two points often can not be satisfied in practice. The article [6] studied two estimators, namely: the HC3 estimator and the weighted bootstrap estimator. Furthermore, it evaluated the finite sample behavior of two bootstrap tests and proposed a new estimator. The literature [7] employed the maximum likelihood method to study parameter estimation based on Lognormal distribution jointly logarithmic mean and logarithmic variance model. $y_i \sim LN(\mu_i, \sigma_i^2), \mu_i = x_i'\beta, \ln(\sigma_i^2) = z_i'\gamma, i = 1, 2, \cdots, n$. The model asked y_i to obey a Lognormal distribution. The literature [8] proposed a method to estimate the coefficient in heteroscedastic model, but it still has some disadvantages.

So, with aid of orthogonal arrays, this paper proposes a nonparametric estimation method with simple calculation for the estimation of heteroscedasticity, which has improved the method in literature [8]. Most importantly, this method does not rely on the specific distribution type for y_i. As a consequence, compared with the result in the literature [7], the proposed method in this paper has a a wider range in use.

The paper is structured as follows: Section 2 gives the steps for the estimation of heteroscedasticity by orthogonal table. In section 3, in conjunction with simulated and real data sets, we illustrate the validity of proposed method. Section 4 does a brief summary and points out the direction of future research.

2. Estimation for heteroscedasticity

2.1. Assumptions of the model. Assume that data $(x_{i1}, x_{i2}, \cdots, x_{ip}, y_i)$, $i = 1, 2, \cdots, n$, has the following linear relationship:

$$(2.1) \qquad \begin{cases} Y = X\beta + \varepsilon, \\ E(\varepsilon) = 0_n, D(\varepsilon) = \Sigma = diag(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2), \\ \sigma_i^2 \text{ is not the same, } \ i = 1, 2, \cdots, n. \end{cases}$$

where

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

β is parameter to be estimated. The model is called a heteroscedastic linear model[2].

When heteroscedasticity occurs in the model, if covariance matrix of random item ε is known, we can use generalized least squares estimation (GLSE) $\hat{\beta}$ as the estimation of model coefficients β, that is:

$$\hat{\beta} = (X^{'}\Sigma^{-1}X)^{-1}X^{'}\Sigma^{-1}Y. \tag{2.2}$$

However, in most practical problems, Σ is unknown. To solve this problem, it is necessary to estimate Σ. Therefore, combining with the nature of the orthogonal array, this paper gives a reasonable estimate of covariance matrix Σ of random error term ε.

2.2. Estimation of variance. This subsection gives a method to estimate the covariance matrix Σ of random errors ε in the formula of (2.2) by using orthogonal array. For the convenience of description, we consider the case $p = 3$, i.e, there are three independent variables x_1, x_2, x_3 in the model, and other situations can be promoted similarly.

For example, orthogonal array $L_9(3^4)$, which is generated with the help of the knowledge of combinatorial mathematics and probability[9].

$$L_9(3^4) = \begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 \\ 1 & 2 & 3 & 2 & 3 & 1 & 3 & 1 & 2 \\ 1 & 2 & 3 & 3 & 1 & 2 & 2 & 3 & 1 \end{pmatrix}^{'}. \tag{2.3}$$

The detailed process of heteroscedastic estimation for the case $p = 3$ is as follows:

(1) For each set of observation $(x_{i1}, x_{i2}, x_{i3}), i = 1, 2, \cdots, n$, taking each independent variable as a factor, we make the following treatment for it: using (x_{i1}, x_{i2}, x_{i3}) as central value and $(\frac{x_{i1}}{\Delta}, \frac{x_{i2}}{\Delta}, \frac{x_{i3}}{\Delta})$ as tolerance, we can get three levels of each factor:

$$(x_{i1}, x_{i2}, x_{i3}) \to \begin{pmatrix} x_{i1} - \frac{x_{i1}}{\Delta} & x_{i2} - \frac{x_{i2}}{\Delta} & x_{i3} - \frac{x_{i3}}{\Delta} \\ x_{i1} & x_{i2} & x_{i3} \\ x_{i1} + \frac{x_{i1}}{\Delta} & x_{i2} + \frac{x_{i2}}{\Delta} & x_{i3} + \frac{x_{i3}}{\Delta} \end{pmatrix}$$

Where $\frac{1}{\Delta}$ is usually called tolerance and its value often depends on magnitude of data.

(2) Regarding the data produced in the first step as three levels of each factor, we can obtain following data with the help of orthogonal array $L_9(3^4)$:

$$\begin{pmatrix} x_{i1} - \frac{x_{i1}}{\Delta} & x_{i2} - \frac{x_{i2}}{\Delta} & x_{i3} - \frac{x_{i3}}{\Delta} \\ x_{i1} - \frac{x_{i1}}{\Delta} & x_{i2} & x_{i3} \\ x_{i1} - \frac{x_{i1}}{\Delta} & x_{i2} + \frac{x_{i2}}{\Delta} & x_{i3} + \frac{x_{i3}}{\Delta} \\ x_{i1} & x_{i2} - \frac{x_{i2}}{\Delta} & x_{i3} \\ x_{i1} & x_{i2} & x_{i3} + \frac{x_{i3}}{\Delta} \\ x_{i1} & x_{i2} + \frac{x_{i2}}{\Delta} & x_{i3} - \frac{x_{i3}}{\Delta} \\ x_{i1} + \frac{x_{i1}}{\Delta} & x_{i2} - \frac{x_{i2}}{\Delta} & x_{i3} + \frac{x_{i3}}{\Delta} \\ x_{i1} + \frac{x_{i1}}{\Delta} & x_{i2} & x_{i3} - \frac{x_{i3}}{\Delta} \\ x_{i1} + \frac{x_{i1}}{\Delta} & x_{i2} + \frac{x_{i2}}{\Delta} & x_{i3} \end{pmatrix} \doteq \begin{pmatrix} x_{i11} & x_{i21} & x_{i31} \\ x_{i12} & x_{i22} & x_{i32} \\ x_{i13} & x_{i23} & x_{i33} \\ x_{i14} & x_{i24} & x_{i34} \\ x_{i15} & x_{i25} & x_{i35} \\ x_{i16} & x_{i26} & x_{i36} \\ x_{i17} & x_{i27} & x_{i37} \\ x_{i18} & x_{i28} & x_{i38} \\ x_{i19} & x_{i29} & x_{i39} \end{pmatrix}$$

(3) For each set of observation $y_i, i = 1, 2, \cdots, n$, we can take 9 independent random numbers $y_{ik}, (k = 1, 2, \cdots, 9)$ from normal distribution $N(y_i, \theta^2)$ or uniform distribution $U[y_i - h, y_i + h]$, where θ^2 and h often take a relatively small value to satisfy the need that produced data y_{ik} have little deviation.

(4) According to the source of data in previous step, we know that 9 random numbers are produced from one distribution independently, i.e. they have same variance. Further, from the regression model we can see the variance of random error term and the variance of dependent variable are the same. So we can consider that the regression of y_{ik} and x_{ijk}(fix $i, j = 1, 2, 3, k = 1, 2, \cdots, 9$) is homoscedastic. So, using the OLS for this

regression is reasonable. For each i, according to the regression of y_{ik} and x_{ijk},

$$\begin{cases} y_{ik} = \gamma_0 + \gamma_1 x_{i1k} + \gamma_2 x_{i2k} + \gamma_3 x_{i3k} + \varepsilon_{ik}, \\ E(\varepsilon_{ik}) = 0, D(\varepsilon_{ik}) = \sigma_{ik}^2, k = 1, 2, \cdots, 9. \end{cases} \tag{2.4}$$

we can obtain residual squares:

$$\begin{cases} e_{ik}^2 = (y_{ik} - \hat{y_{ik}})^2, \hat{y_{ik}} = \hat{\gamma_0} + \hat{\gamma_1} x_{i1k} + \hat{\gamma_2} x_{i2k} + \hat{\gamma_1} x_{i3k}, \\ i = 1, 2, \cdots, n; \ k = 1, 2, \cdots, 9. \end{cases} \tag{2.5}$$

(5) Note the variance of ε_i in the model (2.1) as σ_i^2. According to the calculation formula of variance $\sigma_i^2 = E(\varepsilon_i^2) - [E(\varepsilon_i)]^2$ and the basic assumptions for ε_i, i.e. $E(\varepsilon_i) = 0$, there is a conclusion that $\sigma_i^2 = E(\varepsilon_i^2)$. So this paper uses $\sum_{k=1}^{9} e_{ik}^2/9$ to estimate $E(\varepsilon_i^2)$, i.e. $\hat{\sigma_i^2} = \sum_{k=1}^{9} e_{ik}^2/9$, $i = 1, 2, \cdots, n$, Finally we get the covariance matrix of ε as $\hat{\Sigma}$=diag$(\hat{\sigma_1^2}, \hat{\sigma_2^2}, \cdots, \hat{\sigma_n^2})$.

3. Example

Simulation Study: In this part, we want to confirm our results by simulation experiments. Here we consider a simple heteroscedastic variance problem where the variance is the square of first variable x_1 corresponding to it.

Let us consider a simple three variable linear model:

$$\begin{cases} y_i = 0.2 + 2x_{i1} + 3x_{i2} + 4x_{i3} + \varepsilon_i, \\ \varepsilon_i \sim N(0, x_{i1}^2), i = 1, 2, \cdots, n. \end{cases} \tag{3.1}$$

Above all, in our simulation study, all the values of independent variables are being taken equally from the uniform distribution $U[0, 1]$. From the model we know ε_i are generated from $N(0, x_{i1}^2)$. Further, y_i is easily obtained.

Then, using the simulation data, we take advantage of the proposed method in subsection 2.2 to estimate the variance and get the simulation equation with the aid of formula (2.2). Further, we obtain the absolute value of prediction error. We run this simulation experiment under the following situation: normal distribution $N(y_i, \theta^2)$ or uniform distribution $U[y_i - h, y_i + h]$, three different sample sizes $n = 30, 60, 90$, the number of experiment $m = 100, 1000$, $\frac{1}{\Delta} = 0.01$ or 0.001 and θ^2=0.01 or 0.001. The absolute value of prediction error of this simulation experiment is arranged in table 1(using SAS macro).

From the table 1, we can find that the absolute value of prediction error has little differences by the proposed method (method 1) in this paper, which is unrelated to the the choice of distribution and parameters. Therefore the choice of distribution and parameters has little effect on the proposed method (method 1) in this paper and the proposed method is stable.

To confirm the performance of our method, we adopt the method (method 2) in paper [10] and the method (method 3) in paper [6] and also get the absolute value of prediction error with the help of weighted least square estimation (WLSE). See Table 1. The process of method 2 can be described as follows: Sort the explanatory variables x_1 from small to large and other variables y_i, x_2, x_3 maintain the original correspondence. Divide the x_1 into k groups and j-th group contains n_j numbers. Let the mean of numbers in the j-th group as x'_{1j} and use x'_{1j} in place of the original data in j-th group. So the data becomes $(x'_{1j}, x_{i2}, x_{i3}, y_i), j = 1, 2, \cdots, k$. $i = 1, 2, \cdots, n$. We divide the sample variance of the i-th group s_i^2 on the both sides of the classic linear regression model and use OLS to estimate the parameter. Meanwhile, the estimator proposed in paper [6], called HC4, is as formula (3.2).

Table 1: The absolute value of prediction error in simulation

		m=100				m=1000			
normal distribution (by method 1)		$\frac{1}{\Delta}=0.01$ θ^2=0.01	$\frac{1}{\Delta}=0.01$ θ^2=0.001	$\frac{1}{\Delta}=0.001$ θ^2=0.01	$\frac{1}{\Delta}=0.001$ θ^2=0.001	$\frac{1}{\Delta}=0.01$ θ^2=0.01	$\frac{1}{\Delta}=0.01$ θ^2=0.001	$\frac{1}{\Delta}=0.001$ θ^2=0.01	$\frac{1}{\Delta}=0.001$ θ^2=0.001
	n=30	0.4094	0.4092	0.4160	0.4133	0.4050	0.4038	0.4032	0.4046
	n=60	0.4152	0.4122	0.4104	0.4165	0.4108	0.4101	0.4108	0.4110
	n=90	0.4072	0.4096	0.4128	0.4106	0.4105	0.4099	0.4108	0.4088
uniform distribution (by method 1)		$\frac{1}{\Delta}=0.01$ h=0.02	$\frac{1}{\Delta}=0.01$ h=0.04	$\frac{1}{\Delta}=0.001$ h=0.02	$\frac{1}{\Delta}=0.001$ h=0.04	$\frac{1}{\Delta}=0.01$ h=0.02	$\frac{1}{\Delta}=0.01$ h=0.04	$\frac{1}{\Delta}=0.001$ h=0.02	$\frac{1}{\Delta}=0.001$ h=0.04
	n=30	0.4089	0.4064	0.4121	0.4110	0.4002	0.3998	0.4024	0.4012
	n=60	0.4077	0.4109	0.4115	0.4084	0.4071	0.4094	0.4079	0.4085
	n=90	0.4097	0.4089	0.4086	0.4102	0.4081	0.4072	0.4080	0.4083

		m=100	m=1000
by method 2	n=30	0.6934	0.6711
	n=60	0.5743	0.5759
	n=90	0.5196	0.5222

		m=100	m=1000
by method 3	n=30	0.3974	0.3893
	n=60	0.40387	0.40249
	n=90	0.40650	0.40470

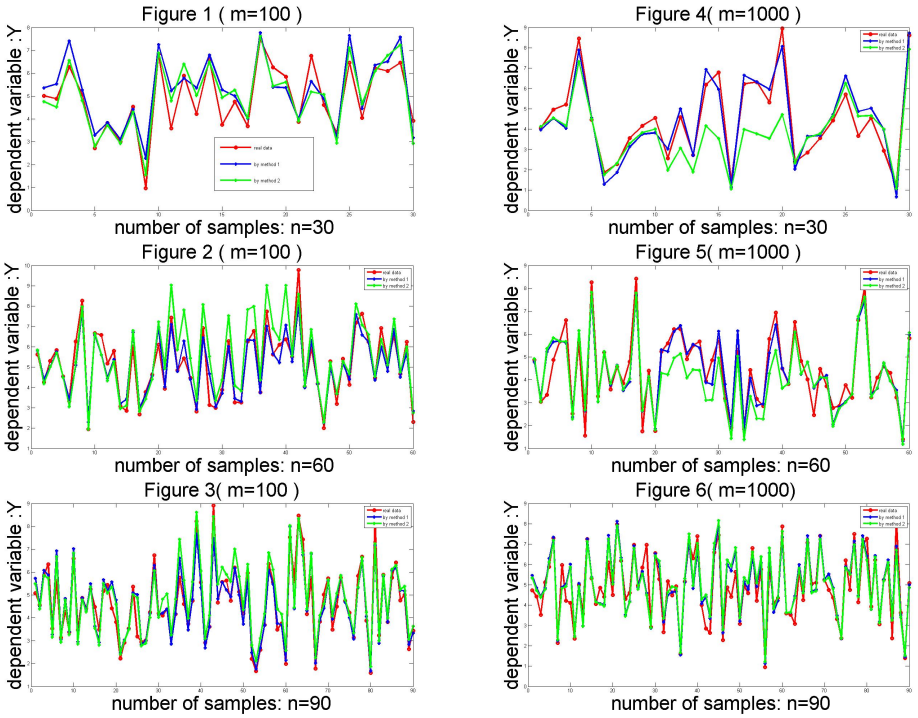

Figure 1-6: Simulation study

$$\hat{\Omega} = diag\{\hat{u_1}^2/(1-h_1)^{\delta_1}, \cdots, \hat{u_n}^2/(1-h_n)^{\delta_n}\} \tag{3.2}$$

Where $\delta_i = min\{4, \frac{nh_i}{\sum_{j=1}^{n} h_i}\}$, h_i is the ith diagonal element of the "hat matrix" $H = X(X'X)^{-1}X'$ and $\hat{u_i}^2$ is i-th diagonal element of the diagonal matrix formed out of the vector of squared least-squares residuals.

From the table 1, we can see, on the one hand, the results using method 1 are smaller than them using method 2 on the whole. On the other hand, with the increase of the number of samples, the error by method 1 is changing little compared with those by method 2, which illustrates its stability and also shows that the newly proposed method in this paper is fitted with the data including 90 samples. Also, we can see that the results in table 1 by method 1 are almost the same with them by method 3. But the most important point we should not neglect is that the method 1(proposed method in this paper) does not rely on the distribution of data. However, according to formula of estimator proposed in paper [6], we can find that it depends on the normal distribution. So, the method we proposed in this paper has a wider range in practise.

Meanwhile, we use Figure 1-6(using MATLAB) to demonstrate the advantage of the method proposed (method 1) in this paper. In Figure 1-6, the horizontal axis represents the number of sample and the longitudinal axis notes the predicted value of the dependent variable obtained by different method. Red, blue and green lines respectively denote the value of independent variable using different methods. Red notes real values of dependent variable, blue indicates the predicted value of dependent variable using the proposed method (method 1, using $N(y_i, \theta^2)$, $\frac{1}{\Delta} = 0.01$, $\theta^2 = 0.01$) in this paper and green represents the values of dependent variable using the method proposed in article [10](method 2). As is shown in Figure 1-3(fix m), with the increasing of n, the value of dependent variable Y is closer to the real value using method 1. On the other hand, from the Figure 1 and 4, we can see that the effect of method 1(blue line) is better than method 2(green line) with the increasing of m.

Real Example: This example uses the proposed method to estimate the heteroscedasticity of data in example 2.6.2 in literature [7] and gives the regression equation in the presence of heteroscedasticity.

Let y, x_1, x_2, x_3 expresses total GDP and its components in the three industry respectively, namely primary industry, secondary industry and tertiary industry. We take the data from 31 provinces (autonomous regions and municipalities) of China in 2009 for example.

According to the way proposed in 2.2($\mathrm{N}(y_i, \theta^2)$, $\frac{1}{\Delta} = 0.01$ and $\theta^2 = 0.01$), calculate the variance of random term. By the formula (2.2), we get the regression equation and the prediction of dependent variable $\hat{y}$. See Table 2 (using SAS macro).

Table 2: data about real example

No.	y	x_1	x_2	x_3	$\hat{y}$	No.	y	x_1	x_2	x_3	$\hat{y}$
1	12153.03	118.29	2855.55	9179.19	12153.03	17	12961.10	1795.90	6038.08	5127.12	12961.10
2	7521.85	128.85	3987.84	3405.16	7521.85	18	13059.69	1969.69	5687.19	5402.81	13059.69
3	17235.48	2207.34	8959.83	6068.31	17235.48	19	39482.56	2010.27	19419.70	18052.59	39482.56
4	7358.31	477.59	3993.80	2886.92	7358.31	20	7759.16	1458.49	3381.54	2919.13	7759.16
5	9740.25	929.60	5114.00	3696.65	9740.25	21	1654.21	462.19	443.43	748.59	1654.21
6	15212.49	1414.90	7906.34	5891.25	15212.49	22	6530.01	606.80	3448.77	2474.44	6530.01
7	7278.75	980.57	3541.92	2756.26	7278.75	23	14151.28	2240.61	6711.87	5198.80	14151.28
8	8587.00	1154.33	4060.72	3371.95	8587.00	24	3912.68	550.27	1476.62	1885.79	3912.68
9	15046.45	113.82	6001.78	8930.85	15046.45	25	6169.75	1067.60	2582.53	2519.62	6169.753
10	34457.30	2261.86	18566.37	13629.07	34457.30	26	441.36	63.88	136.63	240.85	441.36
11	22990.35	1163.08	11908.49	9918.78	22990.35	27	8169.80	789.64	4236.42	3143.74	8169.80
12	10062.82	1495.45	4905.22	3662.15	10062.82	28	3387.56	497.05	1527.24	1363.27	3387.56
13	12236.53	1182.74	6005.30	5048.49	12236.53	29	1081.27	107.40	575.33	398.54	1081.27
14	7655.18	1098.66	3919.45	2637.07	7655.18	30	1353.31	127.25	662.32	563.74	1353.31
15	33896.65	3226.64	18901.83	11768.18	33896.65	31	4277.05	759.74	1929.59	1587.72	4277.05
16	19480.46	2769.05	11010.50	5700.91	19480.46						

Obtain the relation between y and x_1, x_2, x_3 by using σ_i^2 and formula: $y = x_1 + x_2 + x_3$. Compare and analyze the above regression equation with results of article from the following aspects:

(1) According to the meaning of independent variable and dependent variable, we can find the equation above is closer to the actual situation than the literature [7]. This can also be confirmed from the differences between the actual value of dependent variable and its prediction in Table 1.

(2) The literature [7] requires the specific distribution type for y_i, however the reported method in this paper does not rely on the limitation. As a consequence, the proposed method in this paper has a wider range in use.

4. Conclusions

When the covariance matrix of the random error term in the heteroscedastic regression model is unknown, this paper proposes a nonparametric method for the estimation of heteroscedasticity by orthogonal arrays. Most of all, this method does not rely on the distribution of data. Based on the fact that orthogonal arrays have good statistical properties, from the regression equation and the results in the simulation we can find that the proposed method is better than some other methods, which presents the validity and the stability of the proposed method in the paper. In most of the cases, people only focus on the test on the existence of heteroscedasticity and estimation of the heteroscedasticity, however, few people study the degree of impact of heteroscedasticity and variable that causes heteroscedasticity. These two aspects can be discussed further.

Acknowledgements

The authors are highly grateful to the anonymous reviewers for their helpful comments and suggestions for improving the paper. This research is partially supported by a grant of Project of National Natural Science Foundation of Nature(71031006), Youth Science and Technology Fund of Shanxi Province(201202105-1), National Natural Science Foundation of China(41101440).

References

[1] Gong X. The treatment on heteroscedastic data in the regression model, *East China Normal University*, 2002.

[2] He Q. A linear model heteroscedasticity local polynomial regression, *Systems Engineering Theory Methodology Applications*, **12(2)**, 153-156, 2003.

[3] Fan J. and Li R. Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**, 1348-1360, 2001.

[4] Hall P. and Carroll R. Variance function estimation in regression: the effect of estimating the mean, *Journal of the Royal Statistical Society Series B*, **51(1)**, 3-14, 1989.

[5] Carroll R., Wu C. and Ruppert D. The effect of estimating weights in weighted least squares, *Journal of the American Statistical Association*, **83**, 1045-1054, 1988.

[6] Cribari-Neto F. Asymptotic inference under heteroscedasticity of unknown form. Computational Statistics and Data Analysis, 45, 215-233, 2004.

[7] Huang L. Statistical inference based on the log-normal distribution model of heteroscedasticity, *Kunming University of Science and technology*, China, 2011.

[8] Zhang X. and Hao H. A new method to estimate variance in heteroscedastic model, *Journal of North University of China(Natural Science Edition)*, **34(5)**, 481-484, 2013.

[9] Mao S. Experimental Design. *China Statistics Press*, 2004.

[10] Zhang H. Testing for heteroscedasticity and two-stage estimation based on packet, *Quantitative and Technical Economics Research*, **1**, 129-137, 2006.