

Linear regression models for heteroscedastic and non-normal data

Raksmei Thin^a, Klairung Samart^{a,b,*}, Naratip Jansakul^a

^a Department of Mathematics and Statistics, Faculty of Science, Prince of Songkla University, Songkhla 90110 Thailand

^b Statistics and Applications Research Unit, Faculty of Science, Prince of Songkla University, Songkhla 90110 Thailand

*Corresponding author, e-mail: klairung.s@psu.ac.th

Received 10 Nov 2019

Accepted 27 Apr 2020

ABSTRACT: In common practices, heteroscedasticity and non-normality are frequently encountered when fitting linear regression models. Several methods have been proposed to handle these problems. In this research, we applied four different estimation methods: ordinary least squares (OLS), transform both sides (TBS), power of the mean function (POM) and exponential variance function (VEXP), dealing with three different forms of the non-constant variances under four symmetric distributions. In order to study the performance of the four methods in estimating the studied model parameters, a simulation study with sample sizes of 20, 50 and 100 was conducted. Relative bias, mean squared error (MSE) and coverage probability of the nominal 95% confidence interval for regression parameters were accessed. The simulation results and application to real life data suggest that each estimation method performed differently on different variance structures and different distributions whereas the sample size did not give much effect on each estimation method. In overall, the TBS method performed best in terms of smallest bias and MSE, especially under extreme heteroscedasticity. On the other hand, the OLS method was very accurate in maintaining the nominal coverage probabilities although it had relatively poor performance in terms of bias.

KEYWORDS: heteroscedasticity, symmetric distribution, variances function, simulation

MSC2010: 62J05

INTRODUCTION

In statistical analysis, linear regression has been widely applied for modeling the relationship between the response variable and one or more predictors. In order to use this approach sufficiently, four fundamental assumptions: linearity, independence, homoscedasticity, and normality must be verified.

In practice, however, the violations of these properties may adversely affect performance characteristics of methods of inference. For the data with the heteroscedastic response, it has been noted that the point estimates of the model parameters are inefficient and hence may yield unreliable inference for the model parameters [1]. To deal with these problems, several possible methods have been proposed. One common method that has been used in several research [2–5] is known as data transformation using logarithm transform both sides (TBS). This method had been used in modeling blueberry fruits and flower buds, red raspberry growth and yield, and Pharmacokinetics data [5–7]. Although, the transformation has been applied in many sit-

uations, one might be uncertain of this method because it might be difficult to make inference about the original scale [2]. Meanwhile, two kinds of so-called residual variance functions: power of the mean function (POM) and exponential function (VEXP) were also used to reduce heteroscedasticity [1, 5, 6, 8, 9]. Its parameters are estimated by generalized least square method (GLS).

To the best of our knowledge, the previous studies were only focused on heteroscedasticity where normality assumption was assumed. Therefore, this leads to the aim of our study to compare and discuss four different estimation methods: OLS, TBS, POM, and VEXP methods to deal with three different forms of non-constant variances under four symmetric distributions namely normal distribution, Laplace distribution, location scale of t distribution, and logistic distribution.

ESTIMATION METHODS

In statistical methodology, regression analysis is a study to examine the relationship between two or

more quantitative variables. Focusing on the relationship of two variables, the model can be stated as follows [4]:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad (1)$$

where Y_i is the value of the response variable in the i -th trial; β_0 and β_1 are parameters; x_i is the value of the predictor variable in the i -th trial which is a known constant; ϵ_i is a random error term which is normally distributed with mean $E[\epsilon_i] = 0$ and constant variance $\text{Var}(\epsilon_i) = \sigma^2$; ϵ_i and ϵ_j are uncorrelated for all $i \neq j$, and $E[Y_i] = \mu_i = \beta_0 + \beta_1 x_i$ is the linear mean function.

Ordinary least squares estimation (OLS)

The method of least squares is a standard approach to obtain the parameter estimates that minimize the sum of the squares of the residuals. That is under model (1), the OLS estimator $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$ of $\beta = (\beta_0, \beta_1)^T$ is obtained by minimizing the sum of squared error:

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2. \quad (2)$$

For a sample of n observations (x_i, y_i) , the OLS estimates are

$$\hat{\beta}_0 = \frac{1}{n} \left(\sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i \right) = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

The transform-both-sides estimation (TBS)

Under model (1), transform both sides model has been used to some context in practices in order to transform distribution, mean and variance [3, 5, 10–12]. The TBS allows the parameters to be estimated efficiently using OLS estimation method. In practice, the logarithmic transformation on response and explanatory variables is commonly used and the corresponding model (1) then become

$$\log(Y_i) = \beta_0 + \beta_1 \log(x_i) + \epsilon_i. \quad (3)$$

Variance functions estimation

Another way to prescribe models with the variance is by modifying the constant error variance assumption [5]. A commonly applicable assumption about

the model error variance is the power of the mean variance function model (POM) [1, 9].

$$\text{Var}(Y_i) = \sigma^2 \hat{y}_i^\theta, \quad (4)$$

where σ and θ are the unknown parameters of the variance function model. This variance function indicates that the observed variance depends on the predicted value, \hat{y}_i . In general, $\theta = 0$ refers to homoscedasticity or constant variance model; $\theta = 1$ represents the variance proportional to mean or the data come from a Poisson distribution; and $\theta = 2$ corresponds to constant coefficient of variation (CV) model [9].

Another extended variance function model is an exponential variance function (VEXP) defined by

$$\text{Var}(Y_i) = \sigma^2 \exp(2\theta \hat{y}_i). \quad (5)$$

In practice, one may have prior knowledge about the value of θ in certain cases. Therefore, the parameter β may be estimated by the weighted least squares method with weights $1/\hat{y}_i^\theta$ or $1/\exp(2\theta \hat{y}_i)$ for models (4) and (5), respectively [1]. Nevertheless, [6] stated that one must be clear about the true value of θ , thus the right weight scheme can be incorporated in fitting linear models.

SIMULATION STUDY

This section contains results from a simulation study to assess and compare the performance of the three estimation methods discussed in the previous section. The assessment of the estimates was based on relative bias, mean squared error (MSE) and coverage probability at the nominal 95% confidence interval.

In the simulation study, a random sample of size n (20, 50 and 100) was generated and values of X were independently drawn from the uniform distribution over $[0, 1]$ with corresponding values of Y given by

$$Y_i = 3 + x_i + e_i, \quad (6)$$

where e_i were independently drawn from four symmetric distributions: standard normal distribution, Laplace distribution, location scale of t -distribution, and logistic distribution. All distributions were set with the mean of 0 and three different variance structures in order to form heteroscedasticity.

The three variance structures consist of power of the predictor (\sqrt{x}), exponential function of the predictor ($\exp(x)$), and a constant plus fitted values ($0.5 + \hat{y}_i$). The simulation process was replicated $N = 1000$ times. The parameters estimates, biases,

MSEs, and coverage probabilities were then calculated. All the simulations were achieved by using R software [13].

The simulation results for each sample size of 20, 50, and 100, and variance structure are summarized in terms of relative biases, MSEs, as well as the 95% empirical coverage probabilities and the average lengths of the confidence intervals (in parentheses). Tables 1–3 show the results when $\text{Var}(\epsilon_i) = \text{sqrt}(x_i)$, $\text{Var}(\epsilon_i) = \exp(x)$, and $\text{Var}(\epsilon_i) = 0.5 + \hat{y}$, respectively.

The results set out in Table 1 show that when the sample size is small with normal distribution, all of the estimation methods produce reasonably accurate estimates. However, under non-normal distributions namely Laplace, location scale of t and logistic distributions, TBS tends to have smaller bias and MSE than the other methods. On the other hand, the OLS estimation obviously overestimates the slope parameter, especially under Laplace and logistic distributions. In terms of coverage probabilities, the OLS and VEXP methods seem to be better than the others whereas the POM and TBS give the shortest average length for β_0 and β_1 , respectively.

Likewise, with large sample sizes of 50 and 100, the results are similar to those in small sample size. However, it is quite obvious in all distributions that POM has the smallest bias and MSE for β_0 and TBS has the smallest bias and MSE for β_1 .

The results displayed in Table 2, where the variance is in exponential form, show that when the sample size is small, in all distributions, VEXP provides the smallest bias and MSE for β_0 whereas TBS provides the smallest bias and MSE for β_1 . However, the 95% CIs based on OLS are quite accurate in maintaining the nominal confidence level although it gains the accuracy with wider intervals.

When the sample size increases to 50 and 100, the results displayed in Table 2 reveal that VEXP still provides the smallest bias and MSE for β_0 whereas TBS provides the smallest bias and MSE for β_1 . In terms of coverage probabilities, the OLS and VEXP method gain the accuracy with wider intervals in all cases.

The results presented in Table 3, where the variance is a function of the fitted values, reveal that the OLS, POM, and VEXP methods provide severe bias and MSE in all distributions. TBS, although considered to be the best one, still gives small bias estimates. Moreover, TBS also performs well in terms of coverage probabilities and the average lengths of the confidence intervals. When the sample size increases as shown in Table 3, the

results are in agreement with those in small sample size. However, as expected, the severity of bias and MSE of the OLS, POM, and VEXP estimates seem to decrease.

APPLICATION TO REAL LIFE DATA

In this section, we apply all the studied methods to real dataset, Bike Sharing Dataset. The data were extracted from [14]. To form the relationship between the count of rental bikes (cnt) and count of registered users (registered) collected during the period from 1 Jan 2011 to 31 Dec 2012 in Washington DC, USA, the regression analysis was used. The results are shown in Figs. 1 and 2. The models validation of the errors assumptions by using Shapiro-Wilk test for normality and Levene's test for homoscedasticity [15] was conducted. The results, although not reported here, have shown that this dataset is heteroscedastic with non-normal distribution.

Fig. 1 shows the comparison of the fitted models with prediction intervals. The OLS and VEXP models obviously overestimate the total number count with small numbers of registered users, although, OLS acquires the accuracy of wider prediction intervals. The TBS and POM methods, however, are considered to be the best ones among all other models which most of the observations are within the prediction intervals.

Fig. 2 shows the standardized residuals of the fitted model. The OLS and VEXP methods present the funnel effect with greater error variance corresponding to larger predicted values, which indicate that OLS and VEXP methods do not confirm the assumption of homogeneity of variance. The POM method, although is slightly better, still presents a little of funnel effect whereas the TBS method does not show the funnel effect, and thus the assumption of homogeneity was verified.

CONCLUSION

In this article, we reviewed and compared four estimation methods under three different forms of non-constant variances and four symmetric distributions. The assessment of the estimation methods is based on their relative bias, MSE, coverage probability, and average length. The results given by the simulations study indicate that each estimation method performs differently on different variance structures and different distributions, whereas the sample size does not give much effect on each estimation method.

Table 1 Relative biases (Rel. bias), MSEs and coverage probabilities (average length) under $\text{Var}(\epsilon_i) = \text{sqrt}(x)$.

Distribution	Estimator	n	β_0		β_1		Coverage probability	
			Rel. bias	MSE	Rel. bias	MSE	β_0	β_1
Normal	OLS	20	0.0	0.06	0.0	0.33	0.99(1.38)	0.94(2.37)
		50	0.2	0.02	-1.9	0.13	0.99(0.82)	0.95(1.41)
		100	-0.1	0.01	1.0	0.07	0.99(0.57)	0.94(0.98)
	TBS	20	-0.1	0.08	-0.2	0.03	0.92(0.99)	0.99(0.78)
		50	-0.2	0.03	-0.3	0.01	0.92(0.58)	0.99(0.43)
		100	0.1	0.01	0.1	0.00	0.92(0.40)	0.99(0.29)
	POM	20	0.0	0.03	0.0	0.23	0.88(0.59)	0.94(0.59)
		50	0.0	0.01	-0.7	0.07	0.89(0.50)	0.95(1.28)
		100	-0.1	0.00	0.6	0.03	0.91(0.19)	0.94(0.69)
	VEXP	20	0.0	0.04	0.1	0.25	0.97(0.87)	0.95(0.87)
		50	0.1	0.01	-1.0	0.09	0.97(0.50)	0.97(1.28)
		100	-0.1	0.01	0.5	0.04	0.98(0.35)	0.96(0.89)
Laplace	OLS	20	-0.4	0.11	3.1	0.63	0.99(1.88)	0.95(3.27)
		50	0.0	0.04	-0.7	0.24	1.00(1.14)	0.96(1.98)
		100	0.2	0.02	-2.1	0.11	0.99(0.79)	0.97(1.37)
	TBS	20	0.3	0.14	0.7	0.04	0.92(1.36)	0.99(1.07)
		50	-0.2	0.06	-0.2	0.01	0.91(0.59)	0.99(0.60)
		100	-0.3	0.02	-0.6	0.01	0.93(0.56)	0.99(0.41)
	POM	20	-0.1	0.06	0.8	0.42	0.86(0.81)	0.94(2.49)
		50	0.0	0.01	-1.1	0.14	0.88(0.42)	0.94(1.3)
		100	0.1	0.00	-1.7	0.06	0.93(0.26)	0.97(0.96)
	VEXP	20	-0.4	0.08	2.2	0.48	0.97(1.21)	0.96(2.91)
		50	0.1	0.02	-1.7	0.18	0.98(0.71)	0.97(1.77)
		100	0.1	0.01	-1.7	0.07	0.98(0.49)	0.98(1.24)
Location scale of t	OLS	20	-0.1	0.07	0.6	0.37	0.99(1.42)	0.94(2.46)
		50	0.0	0.02	0.9	0.14	0.99(0.83)	0.94(1.43)
		100	0.1	0.01	0.3	0.06	0.99(1.02)	0.96(1.77)
	TBS	20	0.1	0.08	0.6	0.03	0.93(1.03)	0.98(0.81)
		50	0.2	0.03	0.3	0.01	0.92(0.59)	0.99(0.44)
		100	0.0	0.01	-0.1	0.00	0.92(0.73)	0.99(0.53)
	POM	20	-0.3	0.04	1.9	0.25	0.86(0.81)	0.94(2.49)
		50	-0.1	0.01	1.3	0.07	0.89(0.30)	0.94(1.03)
		100	0.0	0.00	0.2	0.03	0.92(0.19)	0.94(0.70)
	VEXP	20	-0.3	0.04	2.4	0.28	0.97(1.21)	0.96(2.91)
		50	-0.1	0.01	1.5	0.09	0.98(0.51)	0.97(1.29)
		100	0.0	0.00	-0.1	0.04	0.99(0.35)	0.98(0.90)
Logistic	OLS	20	-0.5	0.19	3.6	1.09	0.99(2.43)	0.95(4.23)
		50	-0.3	0.06	0.9	0.38	0.99(1.47)	0.95(2.54)
		100	-0.1	0.03	0.3	0.20	0.99(0.57)	0.94(0.99)
	TBS	20	0.3	0.24	0.7	0.08	0.93(1.78)	0.99(1.40)
		50	0.0	0.09	0.4	0.02	0.91(1.05)	0.99(0.77)
		100	0.0	0.04	0.1	0.01	0.92(0.41)	0.99(0.29)
	POM	20	-0.4	0.10	3.3	0.75	0.86(1.05)	0.94(3.25)
		50	-0.2	0.03	0.3	0.23	0.89(0.53)	0.94(1.83)
		100	-0.1	0.01	0.3	0.10	0.92(0.34)	0.95(1.24)
	VEXP	20	-0.3	0.13	1.1	0.87	0.97(1.54)	0.95(3.79)
		50	-0.2	0.04	0.2	0.28	0.99(0.89)	0.96(2.28)
		100	-0.1	0.02	0.7	0.13	0.98(0.63)	0.96(1.60)

Table 2 Relative biases (Rel. bias), MSEs and coverage probabilities (average length) under $\text{Var}(\epsilon_i) = \exp(x)$.

Distribution	Estimator	n	β_0		β_1		Coverage probability	
			Rel. bias	MSE	Rel. bias	MSE	β_0	β_1
Normal	OLS	20	1.0	0.49	-8.3	2.53	0.98(3.45)	0.93(5.99)
		50	0.2	0.17	0.3	0.97	0.99(2.06)	0.92(3.57)
		100	-0.3	0.08	2.4	0.44	0.99(1.42)	0.93(2.47)
	TBS	20	-1.2	0.50	-2.5	0.20	0.91(2.50)	0.97(1.95)
		50	0.3	0.18	0.3	0.06	0.92(1.48)	0.97(1.08)
		100	0.3	0.09	0.7	0.03	0.91(1.01)	0.98(0.73)
	POM	20	1.2	0.45	-9.7	2.33	0.80(2.00)	0.88(4.90)
		50	-0.2	0.17	2.2	0.85	0.85(1.25)	0.88(2.94)
		100	-0.2	0.08	2.1	0.41	0.86(0.90)	0.89(2.05)
	VEXP	20	0.9	0.42	-7.5	2.23	0.92(6.05)	0.92(5.49)
		50	0.0	0.14	1.8	0.77	0.93(1.43)	0.94(3.31)
		100	-0.3	0.07	2.4	0.37	0.94(0.99)	0.95(2.30)
Laplace	OLS	20	-1.0	0.86	6.9	4.64	0.98(4.72)	0.94(8.20)
		50	-0.7	0.31	2.9	1.69	0.98(2.88)	0.94(4.98)
		100	-0.2	0.17	0.6	0.88	0.99(2.01)	0.93(3.48)
	TBS	20	0.6	0.95	1.6	0.38	0.93(3.45)	0.97(2.72)
		50	0.1	0.35	1.1	0.12	0.91(2.06)	0.97(1.52)
		100	-0.1	0.17	0.1	0.05	0.91(1.43)	0.97(1.03)
	POM	20	-1.0	0.78	7.6	4.09	0.82(2.66)	0.91(6.67)
		50	-0.6	0.29	2.1	1.47	0.87(1.68)	0.89(4.04)
		100	-0.2	0.15	0.7	0.75	0.86(1.23)	0.89(2.85)
	VEXP	20	-0.8	0.73	4.3	4.00	0.92(3.26)	0.93(7.40)
		50	-0.6	0.27	1.7	1.40	0.94(1.97)	0.94(4.55)
		100	-0.3	0.14	1.5	0.69	0.93(1.40)	0.94(3.22)
Location scale of t	OLS	20	-0.4	0.51	0.4	2.54	0.98(3.60)	0.94(6.24)
		50	0.5	0.15	-2.4	0.80	0.99(2.10)	0.96(3.64)
		100	0.1	0.44	-0.8	0.08	0.99(1.44)	0.94(2.50)
	TBS	20	-0.1	0.52	0.9	0.22	0.92(2.62)	0.97(2.07)
		50	0.0	0.17	-0.3	0.05	0.92(1.50)	0.99(1.11)
		100	-0.1	0.09	-0.3	0.03	0.91(1.02)	0.98(0.75)
	POM	20	0.0	0.47	-2.1	2.32	0.80(2.09)	0.89(5.14)
		50	0.3	0.15	-1.3	0.74	0.87(1.28)	0.92(2.99)
		100	-0.1	0.08	0.4	0.40	0.89(0.93)	0.89(2.08)
	VEXP	20	0.0	0.43	-2.8	2.23	0.92(2.51)	0.92(5.72)
		50	0.2	0.12	-0.4	0.66	0.95(1.47)	0.96(3.38)
		100	0.0	0.06	0.1	0.35	0.95(1.02)	0.94(2.33)
Logistic	OLS	20	0.0	1.46	-0.8	7.48	0.98(6.22)	0.95(10.80)
		50	0.0	0.53	-2.3	2.73	0.99(3.70)	0.95(6.42)
		100	0.0	0.28	0.5	1.45	0.99(2.60)	0.94(4.51)
	TBS	20	-0.5	1.55	-0.8	0.60	0.93(4.50)	0.98(3.53)
		50	-0.6	0.59	-0.8	0.18	0.91(2.65)	0.98(1.94)
		100	0.0	0.26	-0.2	0.08	0.93(1.85)	0.98(1.34)
	POM	20	-0.7	1.38	2.6	6.77	0.82(3.58)	0.90(8.79)
		50	0.6	0.51	-5.6	2.46	0.85(2.22)	0.90(5.25)
		100	0.1	0.28	-0.1	1.28	0.86(1.63)	0.91(3.73)
	VEXP	20	0.1	1.27	-2.5	6.58	0.92(4.33)	0.94(9.80)
		50	0.4	0.45	-5.4	2.34	0.95(2.58)	0.95(5.93)
		100	0.0	0.22	0.8	1.18	0.94(1.81)	0.95(4.19)

Table 3 Relative biases (Rel. bias), MSEs and coverage probabilities (average length) under $\text{Var}(\epsilon_i) = 0.5 + \hat{y}$.

Distribution	Estimator	n	β_0		β_1		Coverage probability	
			Rel. bias	MSE	Rel. bias	MSE	β_0	β_1
Normal	OLS	20	-0.7	3.26	5.0	10.74	0.96(7.81)	0.95(13.57)
		50	-1.9	1.16	11.1	4.15	0.98(4.62)	0.95(8.04)
		100	-1.3	0.59	4.3	1.98	0.96(3.20)	0.95(5.55)
	TBS	20	0.9	2.04	2.6	1.17	0.94(5.61)	0.95(4.40)
		50	1.2	0.70	3.8	0.34	0.95(3.32)	0.97(2.41)
		100	-0.2	0.35	0.9	0.16	0.94(2.28)	0.96(1.65)
	POM	20	-2.6	3.73	15.4	12.39	0.87(6.40)	0.91(12.43)
		50	-2.2	1.19	12.9	4.21	0.94(4.10)	0.94(7.64)
		100	-1.3	0.61	4.4	2.01	0.93(2.88)	0.93(5.32)
	VEXP	20	-1.9	3.61	14.1	12.37	0.91(6.99)	0.93(13.08)
		50	-2.3	1.16	14.4	4.20	0.96(4.27)	0.95(7.94)
		100	-1.4	0.58	5.0	1.97	0.95(2.96)	0.95(5.50)
Laplace	OLS	20	-0.9	5.93	2.9	21.08	0.97(10.80)	0.95(18.91)
		50	-0.9	2.24	4.7	7.79	0.96(6.48)	0.95(11.22)
		100	1.3	1.26	-1.7	4.37	0.95(4.49)	0.94(7.80)
	TBS	20	-0.2	3.99	0.5	2.13	0.95(7.89)	0.96(6.08)
		50	0.5	1.40	1.9	0.65	0.95(4.61)	0.96(3.40)
		100	0.7	0.74	-1.0	0.33	0.93(3.20)	0.96(2.30)
	POM	20	0.2	6.46	-4.6	22.78	0.87(8.46)	0.92(16.97)
		50	-0.9	2.20	4.3	7.59	0.92(5.53)	0.93(10.44)
		100	1.6	1.22	-3.5	4.27	0.92(3.95)	0.89(7.38)
	VEXP	20	0.3	6.28	-7.4	22.32	0.93(9.50)	0.93(17.89)
		50	-0.7	2.16	3.3	7.51	0.94(5.91)	0.95(10.97)
		100	1.4	1.22	-2.9	4.22	0.93(4.13)	0.94(7.69)
Location scale of t	OLS	20	2.8	3.37	-10.3	11.40	0.96(8.14)	0.96(14.18)
		50	1.2	1.27	-6.2	4.20	0.96(4.68)	0.96(8.12)
		100	-0.8	0.60	5.8	2.10	0.96(3.22)	0.95(5.59)
	TBS	20	0.7	2.10	-1.8	1.17	0.94(5.87)	0.96(4.62)
		50	0.2	0.73	-0.2	0.36	0.95(3.35)	0.96(2.46)
		100	0.9	0.36	2.3	0.16	0.96(2.30)	0.96(1.65)
	POM	20	2.4	3.83	-6.2	12.82	0.87(6.64)	0.92(13.01)
		50	1.1	1.33	-6.0	4.31	0.92(4.15)	0.89(7.71)
		100	-0.9	0.62	6.3	2.12	0.92(2.90)	0.52(2.68)
	VEXP	20	2.8	3.75	-8.6	12.95	0.93(7.26)	0.93(13.67)
		50	1.1	1.29	-5.8	4.22	0.94(4.33)	0.95(8.00)
		100	-0.9	0.60	6.6	2.08	0.94(2.98)	0.95(5.54)
Logistic	OLS	20	-3.9	10.43	12.5	35.68	0.96(13.78)	0.94(24.11)
		50	0.4	3.75	-15.6	13.08	0.97(8.33)	0.95(14.49)
		100	-3.1	1.96	19.4	6.52	0.95(5.77)	0.95(10.01)
	TBS	20	-1.4	6.80	1.6	3.84	0.93(10.02)	0.95(7.78)
		50	-4.0	2.40	-5.3	1.08	0.93(5.96)	0.96(4.37)
		100	-3.4	2.02	20.8	6.64	0.92(5.19)	0.94(9.58)
	POM	20	-4.7	11.58	16.1	38.70	0.87(11.34)	0.91(22.13)
		50	1.0	3.87	-19.6	13.44	0.92(7.22)	0.94(13.62)
		100	-3.4	2.02	20.8	6.64	0.92(5.19)	0.94(9.58)
	VEXP	20	-3.9	11.16	12.1	38.01	0.92(12.38)	0.93(23.16)
		50	0.5	3.76	-16.9	13.23	0.95(7.62)	0.95(14.24)
		100	-3.2	1.95	20.1	6.49	0.94(5.37)	0.95(9.91)

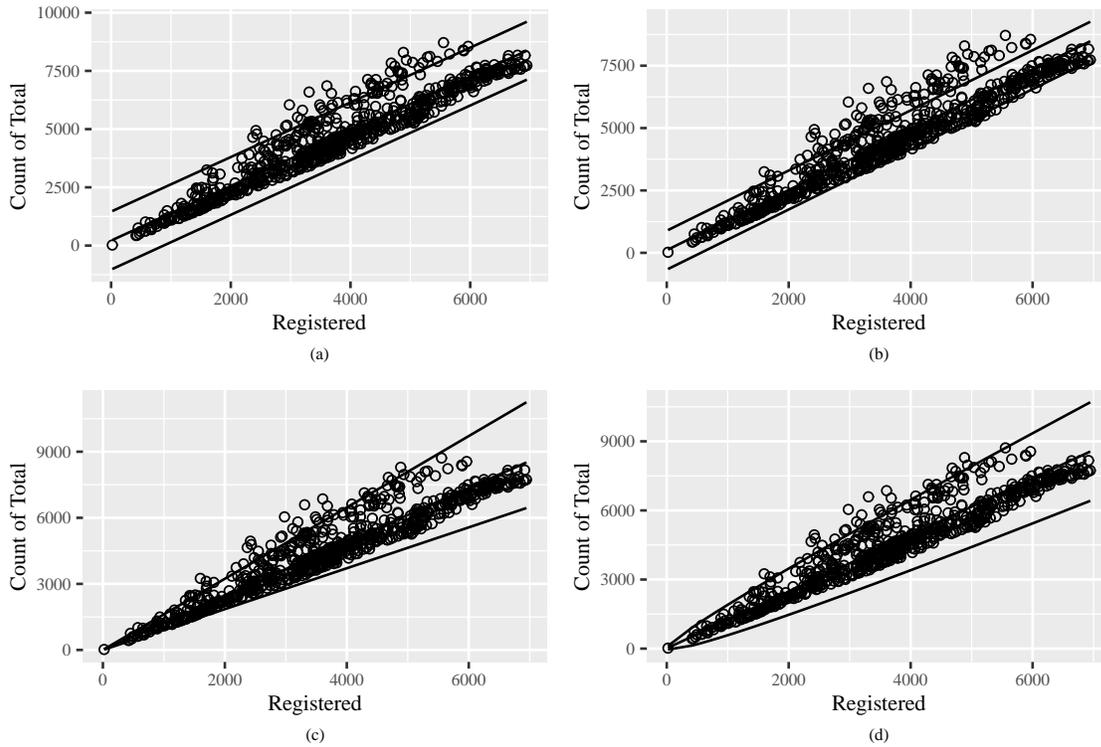


Fig. 1 Prediction intervals of the model fitted to bike rental count daily and registered users with (a) OLS, (b) VEXP, (c) TBS and (d) POM.

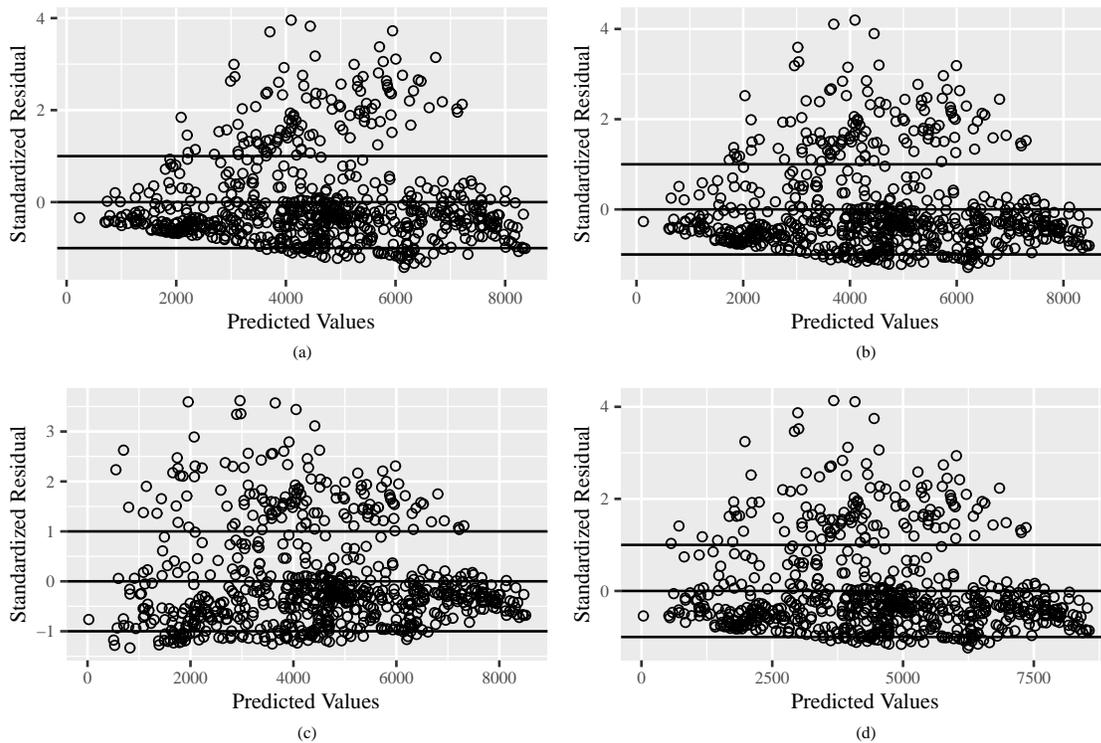


Fig. 2 Standardized residuals of the models with (a) OLS, (b) VEXP, (c) TBS and (d) POM.

As expected, under the power of the predictor variance structure with normal distribution, there is not much difference between estimation methods. Under non-normal distribution, however, the POM and TBS methods perform best in terms of bias and MSE, whereas the OLS method obviously overestimates the slope parameter. When the variance is in the form of an exponential function of the predictor, VEXP and TBS are the best methods to estimate the intercept and slope parameters, respectively. Moreover, the TBS method also performs obviously well under extreme heteroscedasticity such as variance structure of fitted values function whereas the other methods have quite poor performances under this variance structure. In addition, it is obvious that under non-normal distribution, the MSEs are much higher than those under normal distribution with the same values of biases, especially under extreme heteroscedasticity.

In real life data, all discussed estimation methods are in agreement with the performance in simulation study. The TBS and POM methods perform best in terms of model estimates, prediction intervals, and standardized residuals.

In overall, the TBS method seems to perform best in terms of smallest bias and MSE. On the other hand, the OLS method is very accurate in maintaining the nominal coverage probabilities although it has relatively poor performance in terms of bias.

Acknowledgements: We thank all the reviewer for comment that greatly improved the manuscript and Fanaee-T et al for sharing their data set. We also thank the Higher Education Research Promotion and the Thailand's Education Hub for Southern Region of ASEAN Countries Project of the Higher Education Commission, contract no. TEH-AC 099/2016 for funding support.

REFERENCES

- Sidik K, Jonkman JN (2016) A comparison of the variance estimation methods for heteroscedastic nonlinear models. *Stat Med* **35**, 4856–4874.
- Giltinan DM, Carroll RJ, Ruppert D (1986) Some new estimation methods for weighted regression when there are possible outliers. *Technometrics* **28**, 219–230.
- Carroll RJ, Ruppert D (1988) *Transformation and Weighting in Regression*, 1st edn, CRC Press, New York.
- Kutner MH, Nachtsheim CJ, Neter J (2005) *Applied Linear Statistical Models*, 5th edn, McGraw-Hill/Irwin, New York.
- Ávila J, Salvo S, Muñoz C (2013) Comparison of linear regression models considering heteroscedasticity of fruits and flower buds of highbush blueberry cultivated in Chile. *Sci Hort* **51**, 57–62.
- Giltinan DM, Ruppert D (1989) Fitting heteroscedastic regression models to individual pharmacokinetic data using standard statistical software. *J Pharmacokinet Biopharm* **17**, 601–614.
- Clark KE, Hartley SE, Brennan RM, Jennings SN, McMenemy LS, McNicol JW, Mitchell C, Johnson SN (2012) Effects of cultivar and egg density on a colonizing vine weevil (*Otiiorhynchus sulcatus*) population and its impacts on red raspberry growth and yield. *Crop Prot* **32**, 76–82.
- Hao X, Yujun S, Xinjie W, Jin W, Yao F (2015) Linear mixed-effects models to describe individual tree crown width for China-fir in Fujian province, southeast China. *PLoS One* **10**, e0122257.
- Pinheiro J, Bates D (2010) *Mixed-effects Models in S and S-PLUS*, Springer, New York.
- Carroll RJ, Ruppert D (1984) Power transformations when fitting theoretical models to data. *J Am Stat Assoc* **79**, 321–328.
- Miller DM (1984) Reducing transformation bias in curve fitting. *Am Stat* **38**, 124–126.
- Newman MC (1993) Regression analysis of log-transformed data: Statistical bias and its correction. *Environ Toxicol Chem* **12**, 1129–1133.
- R Core Team (2017) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Fanaee-T H, Gama J (2014) Event labeling combining ensemble detectors and background knowledge. *Prog Artif Intell* **2**, 113–127.
- Salvo S, Muñoz C, Ávila J, Bustos J, Ramírez-Valdivia M, Silva C, Vivallo G (2012) An estimate of potential blueberry yield using regression models that relate the number of fruits to the number of flower buds and to climatic variables. *Sci Hort* **133**, 56–63.