Simulation Study on The Estimation Methods for a Joined Point in Tobit-Piecewise Regression

Titirut Thipbharos*

*Faculty of Applied Science, Dhurakij Pundit University.

Abstract

The objective of this study is to introduce the estimation of a joined point in Tobit-piecewise regression model (Mekbunditkul, 2010). There are two interested methods, in this paper, introduced to estimate unknown joined point in Tobit-piecewise regression in the case of observed data containing outliers. Those are based on Quandt's method which is one of the maximum likelihood method and on Levenberg-Marquardt method which is one of nonlinear least square based. Simulation study is provided to compare the potential applicability of each estimator which is considered in terms of the average sum of squares of residuals (ASSR) (Mekbunditkul, 2010). It is found that Tobit-piecewise regression model when the unknown joined point is estimated by LS method yields non-significant smaller ASSR than by ML method in every situation according to the simulation study.

Keywords: Tobit-piecewise regression, LS, MLE, Outliers

Introduction

In the present, an important problem usually found in regression analysis is that observed data consist of outliers. Outliers are data points not typically located close to the usual data. Rousseeuw and Leroy (1987: 3-59) indicated that outliers in $y-direction$ and $x-direction$ affects the heteroscedasticity problem. The outliers in this research are considered in the sense of regression outliers. They are the observed data that are distinct from the linear relationship representing most of the data and they can draw a regression line away from the usual data. However, they exclude unusual incidents. To cope with the occurring problem of outliers, some may fix the model by deleting (weighted by zero) the outliers but this method can be dangerous as it can give the user a false sense of precision in estimation and prediction. This research will apply Tobit-piecewise (abbreviation of Tobit-piecewise) regression model. Mekbunditkul (2010) constructed the Tobit-piecewise regression model by the combination of the Tobit and piecewise regression models. Moreover, there was found that a two-limit Tobit model (Tobin, 1958, Rosett, 1975 and Jöreskog, 2002), limited by some desired variables can reduce the effect of outliers in some situatons. In addition, according to the piecewise regression model (Quandt, 1958: 874, Hudson, 1966: 1097-1129, Goldfeld, Kelejian and Quandt, 1971, Suits, Mason and Chan 1978: 132-133), for instance, fits one data set with two regression regimes when a single regression is inadequate so that the structural change is taken into account as they should be. The structural change in the meaning of regression analysis is a change in one or more of the parameters in a regression model (Poon, et.al. 2008).

The different benefit of two ideas considered in Mekbunditkul's research can be concluded as following: First, Tobit regression is a tool used for investigate the linear relationship when the dependent variable in a regression model is limited. This concept is taken into account for this study in sense that the putting limited value at some desired variable can reduce effect value of outliers in y- and xy-directions. However, the existing of other types of outliers has been not manipulated. Second, piecewise regression is a regression analysis properly applied when the structural change in regression occurs. Hence, in this regression analysis, outliers in x- and xy-directions are taken into account however piecewise is rather not suitable for data consist of outliers in y-directions. Moreover, according to the evident in simulation results, we found that: Tobit-piecewise regression model can more reduce the effect value of outliers and reduce the effect of the structural change than piecewise, Tobit and LS. However, there was not studied the estimation of joined point. In this paper, therefore, this point has been studied.

Tobit-Piecewise Regression Model

The combination of the two-limits Tobit regression model and piecewise regression model to be the Tobit-piecewise model (Mekbunditkul, 2010) is shown in the model (1):

$$Y_i = \begin{cases} L_0 & ; \quad \alpha_1 + \beta_1 x_i + \beta_2 x_i^* + \varepsilon_i \le L_0 \\ \alpha_1 + \beta_1 x_i + \beta_2 x_i^* + \varepsilon & ; \quad L_0 < \alpha_1 + \beta_1 x_i + \beta_2 x_i^* + \varepsilon_i < U_0 \\ U_0 & ; \quad \alpha_1 + \beta_1 x_i + \beta_2 x_i^* + \varepsilon_i \ge U_0, \end{cases} \qquad (1)$$

where regressors are x_i and x_i^* , $x_i^* = (x_i - \delta)D_i$, δ is an unknown joined point of two regression lines, and ε_i's are i.i.d. $N(0, \sigma_i^2)$. Note $\sigma_i^2 = \begin{cases} \sigma_a^2 & \text{if } x_i \le \delta \\ \sigma_b^2 & \text{if } x_i > \delta \end{cases}$. The locally lower and upper limits are $L_0 = \begin{cases} L_a & ; x_i \le \delta \\ L_b & ; x_i > \delta \end{cases}$, and $U_0 = \begin{cases} U_a & ; x_i \le \delta \\ U_b & ; x_i > \delta \end{cases}$.

The probability density function (p.d.f.) of Y is

$$f_Y(y) = \begin{cases} \Phi\left(\dfrac{L_{0i} - \underset{\sim}{x}_{1i}\underset{\sim}{\theta}}{\sigma_i}\right) & ; \quad y_i = L_{0i}, \ i=1,\dots,n_1 \\ \dfrac{1}{\sigma_j}\phi\left(\dfrac{y_j - \underset{\sim}{x}_{2j}\underset{\sim}{\theta}}{\sigma_j}\right) & ; \quad L_{0j} < y_j < U_{0j}, \ j=1,\dots,n_2 \\ 1 - \Phi\left(\dfrac{U_{0k} - \underset{\sim}{x}_{3k}\underset{\sim}{\theta}}{\sigma_k}\right) & ; \quad y_k = U_{0k}, \ k=1,\dots,n_3. \end{cases} \qquad (2)$$

Note $n = \sum_{j=1}^{3} n_j$. The Tobit-piecewise estimator of $\underset{\sim}{\theta}$ can be can be achieved by the ML method when the log-likelihood function of $\underset{\sim}{\theta} = \left(\alpha_1, \beta_1, \beta_2; L_0, U_0, \delta_0, \sigma\right)'$ given $\underset{\sim}{Y}$ for some fixed values of L_0, U_0, $\delta = \delta_0$, and σ^2 known can be written as

$$\ln L(\underset{\sim}{\theta};\underset{\sim}{Y})=\sum_{i=1}^{n_1}\left\{\ln\Phi\left(\frac{L_{0i}-\alpha_1-\beta_1x_i-\beta_2x_i^*}{\sigma_i}\right)\right\}+\sum_{j=1}^{n_2}\left\{\ln\left(\frac{1}{\sigma_j}\phi\left(\frac{y_j-\alpha_1-\beta_1x_j-\beta_2x_j^*}{\sigma_j}\right)\right)\right\}$$
$$+\sum_{k=1}^{n_3}\left\{\ln\left(1-\Phi\left(\frac{U_{0k}-\alpha_1-\beta_1x_k-\beta_2x_k^*}{\sigma_k}\right)\right)\right\}$$

Let $\lambda_j=\dfrac{y_j-\alpha_1-\beta_1x_j-\beta_2x_j^*}{\sigma_j}$; $L_{0j}<y_j<U_{0j}$, $\lambda_i=\dfrac{L_{0i}-\alpha_1-\beta_1x_i-\beta_2x_i^*}{\sigma_i}$;

$y_i=L_{0i}$ and $\lambda_k=\dfrac{U_{0k}-\alpha_1-\beta_1x_k-\beta_2x_k^*}{\sigma_k}$; $y_k=U_{0k}$.

We can get the Tobit-piecewise estimator by solving the solution of normal equations (3a) to (3c)

$$\frac{\partial\ln L(\underset{\sim}{\theta};\underset{\sim}{y})}{\partial\alpha_1}=\sum_{i=1}^{n_1}\left\{-\frac{\phi(\hat{\lambda}_i)}{\sigma_i\Phi(\hat{\lambda}_i)}\right\}+\sum_{j=1}^{n_2}\left\{\frac{\hat{\lambda}_j}{\sigma_j}\right\}+\sum_{k=1}^{n_3}\left\{\frac{\phi(\hat{\lambda}_k)}{\sigma_k\left(1-\Phi(\hat{\lambda}_k)\right)}\right\}=0 \qquad (3a)$$

$$\frac{\partial\ln L(\underset{\sim}{\theta};\underset{\sim}{y})}{\partial\beta_1}=\sum_{i=1}^{n_1}\left\{-\frac{x_i\phi(\hat{\lambda}_i)}{\sigma_i\Phi(\hat{\lambda}_i)}\right\}+\sum_{j=1}^{n_2}\left\{\frac{x_j\hat{\lambda}_j}{\sigma_j}\right\}+\sum_{k=1}^{n_3}\left\{\frac{x_k\phi(\hat{\lambda}_k)}{\sigma_k\left(1-\Phi(\hat{\lambda}_k)\right)}\right\}=0 \qquad (3b)$$

$$\frac{\partial\ln L(\underset{\sim}{\theta};\underset{\sim}{y})}{\partial\beta_2}=\sum_{i=1}^{n_1}\left\{-\frac{x_i^*\phi(\hat{\lambda}_i)}{\sigma_i\Phi(\hat{\lambda}_i)}\right\}+\sum_{j=1}^{n_2}\left\{\frac{x_j^*\hat{\lambda}_j}{\sigma_j}\right\}+\sum_{k=1}^{n_3}\left\{\frac{x_k^*\phi(\hat{\lambda}_k)}{\sigma_k\left(1-\Phi(\hat{\lambda}_k)\right)}\right\}=0, \qquad (3c)$$

where $\hat{\lambda}$ is an estimator of λ. Now we let $\underset{\sim}{Y}_1=\left(Y_{11},...,Y_{1n_1}\right)'$ where $Y_{1i}=L_{0i}$, $\underset{\sim}{Y}_2=\left(Y_{21},...,Y_{2n_2}\right)'$ where $L_{0j}<Y_{2j}<U_{0j}$, $\underset{\sim}{Y}_3=\left(Y_{31},...,Y_{3n_3}\right)'$ where $Y_{3k}=U_{0k}$. The regressor matrices are X_1, X_2 and X_3 corresponding to $\underset{\sim}{Y}_1$, $\underset{\sim}{Y}_2$ and $\underset{\sim}{Y}_3$. And Σ_1 , Σ_2 , Σ_3 are denoted the covariance matrices of $\underset{\sim}{Y}_1$, $\underset{\sim}{Y}_2$ and $\underset{\sim}{Y}_3$, respectively. We can obtain the vector of Tobit-piecewise estimator by

$$\hat{\underset{\sim}{\theta}}_{TP}=\left(X_2'\Sigma_2^{-1}X_2\right)^{-1}\left(X_2'\Sigma_2^{-1}\underset{\sim}{Y}_2\right)-\left(X_2'\Sigma_2^{-1}X_2\right)^{-1}\left[X_1'\Sigma_1^{-1/2}\left\{\underset{\sim}{H}_1(\hat{\lambda})\right\}\right]$$
$$+\left(X_2'\Sigma_2^{-1}X_2\right)^{-1}\left[X_3'\Sigma_3^{-1/2}\left\{\underset{\sim}{H}_3(\hat{\lambda})\right\}\right], \qquad (4)$$

where

$$\underset{\sim}{H}_1(\hat{\lambda})=\left(\frac{\phi\left(\dfrac{L_{01}-\underset{\sim}{x}_{11}\hat{\underset{\sim}{\theta}}_{TP}}{\sigma_{11}}\right)}{\Phi\left(\dfrac{L_{01}-\underset{\sim}{x}_{11}\hat{\underset{\sim}{\theta}}_{TP}}{\sigma_{11}}\right)}\ \cdots\ \frac{\phi\left(\dfrac{L_{0n_1}-\underset{\sim}{x}_{1n_1}\hat{\underset{\sim}{\theta}}_{TP}}{\sigma_{1n_1}}\right)}{\Phi\left(\dfrac{L_{0n_1}-\underset{\sim}{x}_{1n_1}\hat{\underset{\sim}{\theta}}_{TP}}{\sigma_{1n_1}}\right)}\right)'_{1\times n_1},$$

$$\underset{\sim}{H}_3(\hat{\lambda}) = \left(\frac{\phi(\hat{\lambda}_1)}{1-\Phi(\hat{\lambda}_1)} \quad \cdots \quad \frac{\phi(\hat{\lambda}_{n_3})}{1-\Phi(\hat{\lambda}_{n_3})} \right)'_{1\times n_3}$$

, Φ and ϕ are the cumulative density function (c.d.f) and the probability density function (p.d.f) of the standard normal distribution, respectively.

An Unknown Joined Point Estimation

In 1958, Quandt firstly introduced the piecewise regression model and he estimated the parameters for a regression system obeying two separate regimes when the change-point was assumed be unknown parameter need to be estimated. In order to find the value of the change-point, represented by t in Quandt's paper, which maximizes the logarithm of the maximum likelihood,

$$L(t) = -T\log\sqrt{2\pi} - t\log\hat{\sigma}_1 - (T-t)\log\hat{\sigma}_2 - \frac{T}{2},$$

for a given value of T and is a function of t alone. The differentiation of L(t) with respect to t setting equal to zero is inappropriate because t is not continuously variable. Quandt, therefore, suggested a procedure for calculating the value of t by selecting t which gives the maximum likelihood function as follows: Order the observations according to time period and divide the data into a left hand group and a right hand group. Estimate separate regression lines for the two groups, and then move the point of division between the two groups by one unit at a time to the right and one unit at a time to the left. Calculate for each of these new divisions separate regression lines for the left hand group and the right hand group. Move the dividing line again and proceed in an analog fashion. For each division, an expression of L(t) that reaches the maximum likelihood can be found and evaluated (Mekbunditkul, 2010: 14). Nevertheless, Quandt discussed the estimation problem in which the two regression lines are not required to join. D.E. Robison (quoted in Hudson, 1966) found maximum likelihood estimates for the joined points of two polynomial regressions. In addition, Hudson (1966) introduced a parameter estimate based on the LS method and the model was assumed to be joined at $\hat{\delta} = x_i$, for some i. So that we have utilized the combination of Quandt and Hudson's way to estimate the unknown joined point by the following steps:

1. Determining the initial value of t.
2. Putting $\hat{\delta} = x_t$ in the Tobit-piecewise model as shown in equation (1) and in the piecewise model (Mekbunditkul, 2010: 19).

3. Estimating remaining parameters in the Tobit-piecewise model and piecewise model by ML based.

4. Substituting all parameters in step 3 back to each log-likelihood function of Tobit-piecewise and piecewise models and calculating each log-likelihood value.

5. Repeat for t=2,...,n-1

6. Choosing the t value such that maximizes each log-likelihood function.

Another way to solve the curve-fitting problems is a standard technique used to solve nonlinear least squares problems such as the gradient descent method, Gauss-Newton method and Levenberg-Marquardt method, etc. Nevertheless, in this research, we have used the method of Levenberg-Marquardt (Marquardt, 1963). This is the method compromising between the Gauss-Newton and steepest descent methods. The Levenberg-Marquardt updating formula is as followed:

$$\Delta = \left(Z'Z + \lambda \text{diag}\left(Z'Z\right)\right)^{-} Z'\underset{\sim}{e}. \qquad (5)$$

The compromise is as $\lambda \to 0$, the direction approaches Gauss-Newton and as $\lambda \to \infty$,the direction approaches steepest descent.

Marquardt's method is equivalent to performing a series of ridge regressions and is useful when the parameter estimates are highly correlated or the objective function is not well approximated by a quadratic.

Results of Simulations Studies

The performance of Tobit-piecewise regression model is investigated in term of the average sum of squares of residuals (ASSR) (Mekbunditkul, 2010) investigated by simulation studies. SAS program was used to simulate all of situations in this study. There are two situations is considered such as the y-direction, and xy-direction. Nevertheless, other two situations are not taken into account. Since, in case of, no outlier existing it was found that the ASSR1 of Tobit to be equal to the LS regression model while the ASSR for piecewise and Tobit-piecewise are the same. However, both the Tobit and LS results are significantly different from the piecewise and Tobit-piecewise methods. The data fitted by the piecewise and Tobit-piecewise regression models yield the value of ASSR that are smaller than the Tobit and LS methods by about relative efficiency(RE) equal to 0.35 (Mekbunditkul, 2010). These mean that the piecewise and Tobit-piecewise regressions are suitable than LS and Tobit models. In addition, the numerical examples and simulation results in case of the x-direction outliers existing as studied in Mekbunditkul's dissertation are evidenced that Tobit and LS are identical meanwhile the piecewise and Tobit-piecewise are the same.

1. Outliers in the y-direction

For y-direction outliers, the average values of ASSR for 1,000 samples, with certain percentage of outliers and different estimates of the regression coefficients, i.e. LS, Tobit, piecewise and Tobit-piecewise, are presented in Tables 1 and 2 and Figures 1 and 2. The value of ASSR of Tobit-piecewise and of piecewise regression models which each joined point estimated by Levenberg-Marquardt method, nonlinear LS based, is shown in Table 1 corresponding to Figure 1. In addition, the value of ASSR for joined point in Tobit-piecewise and piecewise estimated by Quandt's method, ML based, are presented in Table 2 corresponding to Figure 2.

Table 1 ASSR of four different regression models for Levenberg-Marquardt method in cases of y-direction outliers

Sample Size	% of Y-Outliers	ASSR			
		LS	Tobit	piecewise	Tobit-piecewise
10	5	-	-	-	-
	10	7,691	5,657	6,685	3,253
	15	-	-	-	-
	20	12,041	6,005	10,953	4,454
20	5	4,519	2,280	4,116	1,463
	10	7,841	5,233	7,237	3,027
	15	10,286	5,546	9,655	3,825
	20	11,954	5,903	10,874	4,247
40	5	157	157	56	56
	10	4,569	2,251	4,243	1,599
	15	7,968	5,147	7,641	3,132
	20	10,436	5,525	10,097	3,651
60	5	162	162	54	54
	10	4,588	2,184	4,280	1,521
	15	7,998	4,869	7,620	3,108
	20	10,490	5,463	10,149	3,200
100	5	165	165	55	56
	10	4,594	2,116	4,259	1,440
	15	8,009	4,873	7,426	2,849
	20	10,550	5,435	9,779	2,993

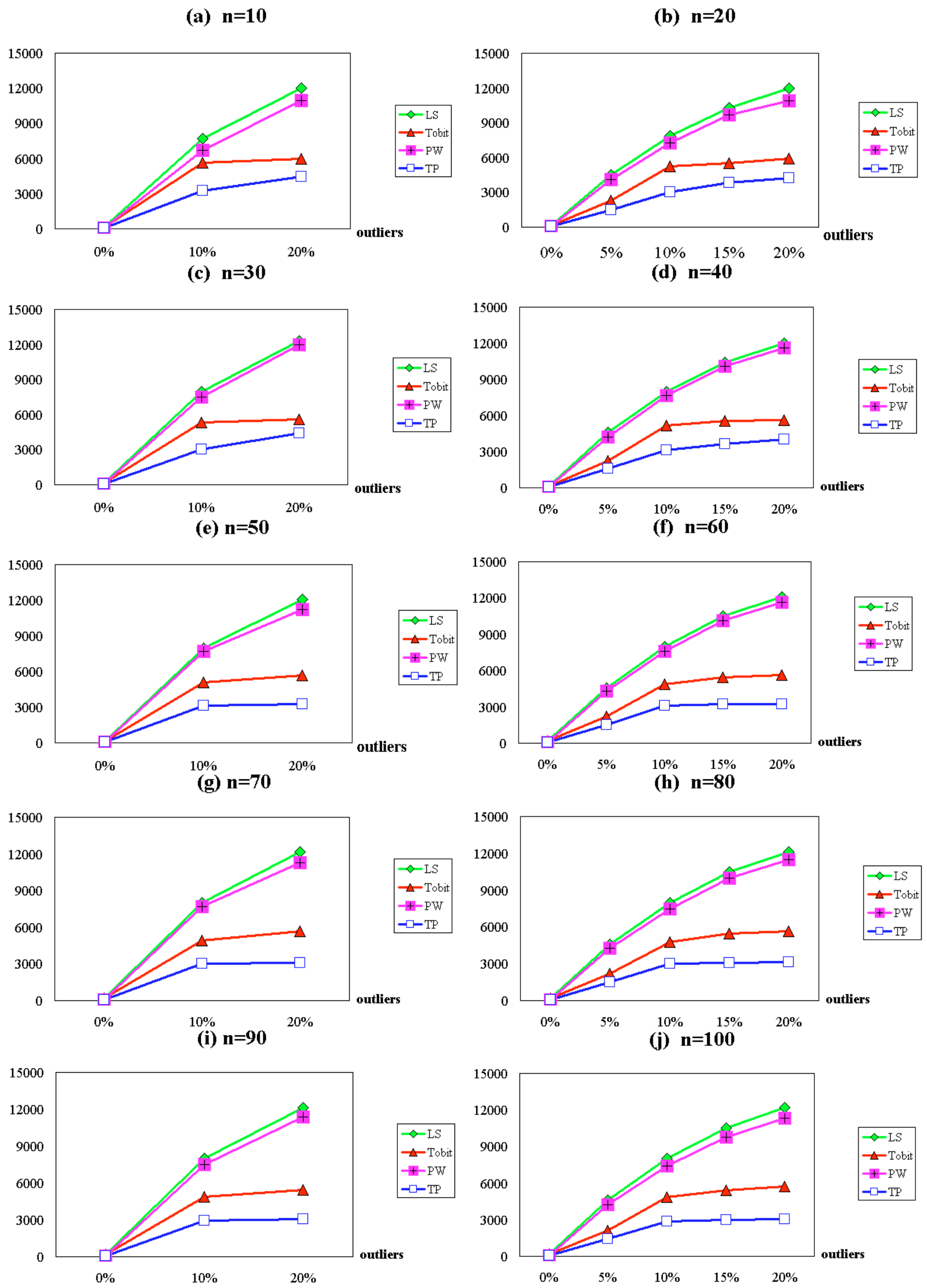

Figure 1 ASSR of four different regression models for Levenberg-Marquardt method varied by percentage of outliers when n=10, 20,..., 100 where y-direction outliers exist

From Table 1 and Figure 1, for all percentages of outliers we found that the significantly smallest ASSR is of Tobit-piecewise regression model, followed by of Tobit, piecewise and LS regression models, in that order.

When considering the information in Table 1, we can see that when the percentage of outliers increases, the Tobit-piecewise regression model with unknown joined point estimated by Levenberg-Marquardt method is preferable to LS model. Furthermore, it was found that both of Tobit and piecewise regression models are more fit than LS model for all values of percentage of outliers. Thus, in this particular case, it can be concluded that the Tobit-piecewise regression model yields the best results followed by the Tobit, piecewise and LS regression models, in that order. Moreover, in the case where outliers exist in the y-direction not only Tobit-piecewise but also the Tobit regression model are preferable to both piecewise and LS. These results correspond to the finding in Mekbunditkul's research.

In Figure 1, it was found that the value of ASSR of four different regression models increases when the percentage of outliers increases for all sample sizes considered.

Next, the results of simulation studies for the y-direction outliers, when unknown joined points in Tobit-piecewise and in piecewise regression models are estimated by Quandt's method, ML based, are exhibited in Table 2 and Figure 2.

From the Table 2, there is the evident that Tobit-piecewise regression with the unknown joined point estimated by Quandt's method yields the smallest ASSR among all of different estimations, followed by Tobit, piecewise with the unknown joined point estimated by Quandt's method and LS, in that order. In addition, from the Figure 2, it is found that the value of ASSR for all type estimations increases when the percentage of outliers increases for all sample sizes considered.

In particular case, the comparison of ML based and LS based by the ASSR is considered. There is found that Tobit-piecewise regression with the unknown joined point estimated by LS based yields non-significant smaller ASSR than by ML based for all values of percentage of outliers considered.

Table 2 ASSR of four different regression models for Quandt's method in cases of y-direction outliers

Sample Size	% of Y-Outliers	ASSR LS	ASSR Tobit	ASSR piecewise	ASSR Tobit-piecewise
10	5	-	-	-	-
	10	7,691	5,657	6,690	3,256
	15	-	-	-	-
	20	12,041	6,005	10,965	4,458
20	5	4,519	2,280	4,120	1,465
	10	7,841	5,233	7,239	3,029
	15	10,286	5,546	9,659	3,827
	20	11,954	5,903	10,875	4,247
40	5	4,569	2,251	4,243	1,599
	10	7,968	5,147	7,643	3,133
	15	10,436	5,525	10,101	3,652
	20	12,009	5,633	11,618	4,011
60	5	4,588	2,184	4,280	1,521
	10	7,998	4,869	7,621	3,109
	15	10,490	5,463	10,152	3,201
	20	12,067	5,636	11,642	3,240
100	5	4,594	2,116	4,263	1,441
	10	8,009	4,873	7,429	2,850
	15	10,550	5,435	9,788	2,995
	20	12,202	5,738	11,350	3,053

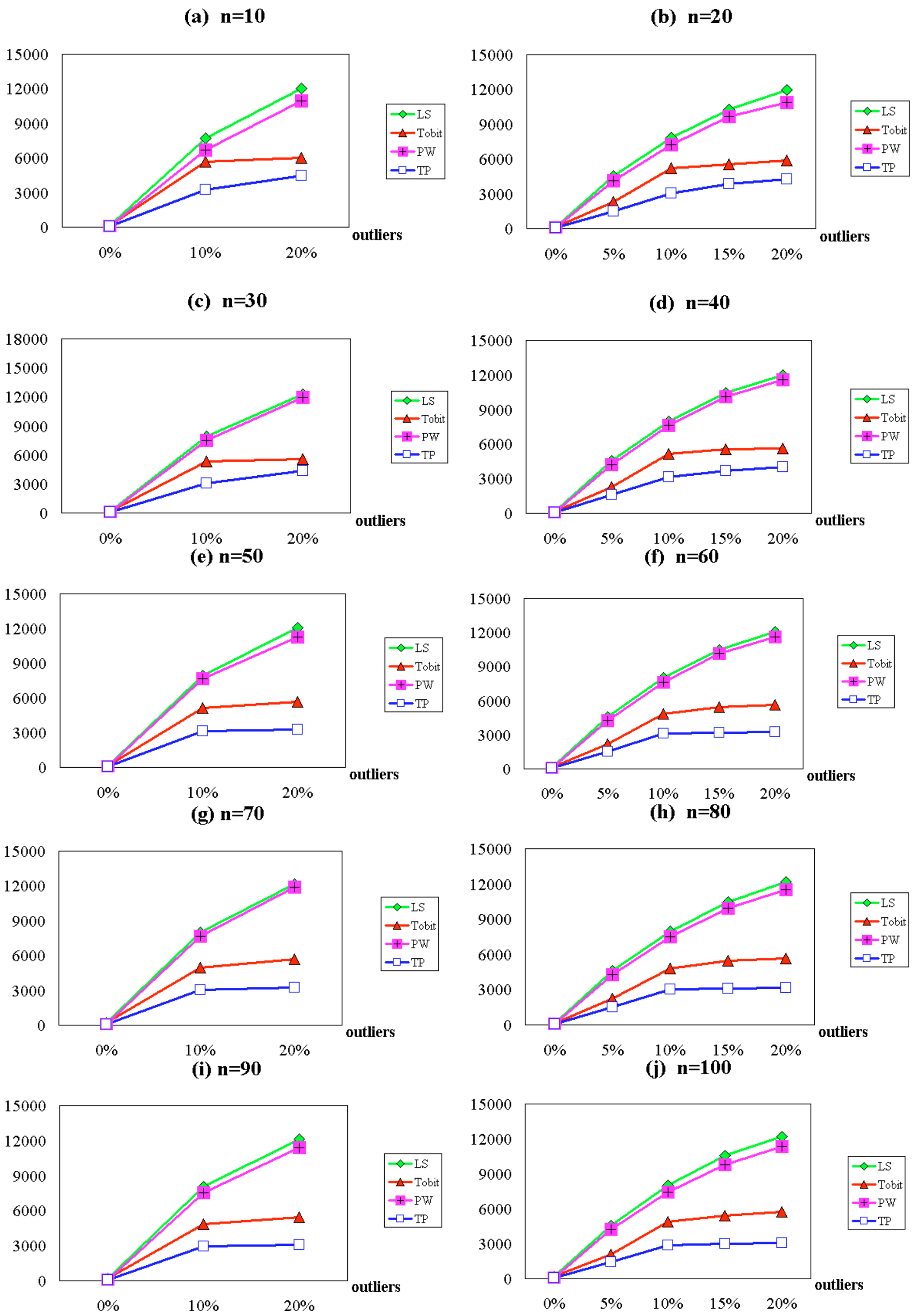

Figure 2 ASSR of four different Regression Models for Quandt's method varied by percentage of Outliers when n=10, 20,..., 100 where y-direction Outliers exist

2. Outliers in the xy-direction

For samples with xy-direction outliers, Tables 3 and 4 give the average values of ASSR from 1,000 generated samples of various sizes and various percentages of outliers considered. The corresponding graphs of ASSR of each different estimation method against percentage of outliers are shown in Figures 3 and 4.

The different of Tables 3 and 4 is that ASSR value being on the Table 3 is from Tobit-piecewise and piecewise with each joined point estimated by Levenberg-Marquardt method. Meanwhile, the ASSR as shown in Table 4 obtains from those fitting lines with each joined point estimated by Quandt's method.

Table 3 ASSR of four different regression models for Levenberg-Marquardt method in cases of xy-direction outliers

Sample Size	% of Y-Outliers	ASSR			
		LS	Tobit	piecewise	Tobit-piecewise
10	5	-	-	-	-
	10	4,157	1,416	2,027	896
	15	-	-	-	-
	20	5,070	4,837	2,591	1,107
20	5	3,069	581	1,573	440
	10	4,270	1,337	2,044	947
	15	5,204	4,769	2,551	1,015
	20	5,286	4,733	2,550	1,385
40	5	3,077	495	1,561	389
	10	4,636	1,044	2,267	384
	15	5,245	2,221	2,412	612
	20	5,497	4,442	2,499	1,112
60	5	3,154	551	1,676	256
	10	4,473	741	1,962	520
	15	5,102	1,559	2,153	665
	20	5,354	3,900	2,226	862
100	5	3,118	493	1,576	204
	10	4,552	838	1,998	376
	15	5,196	1,615	2,133	542
	20	5,459	3,641	2,219	685

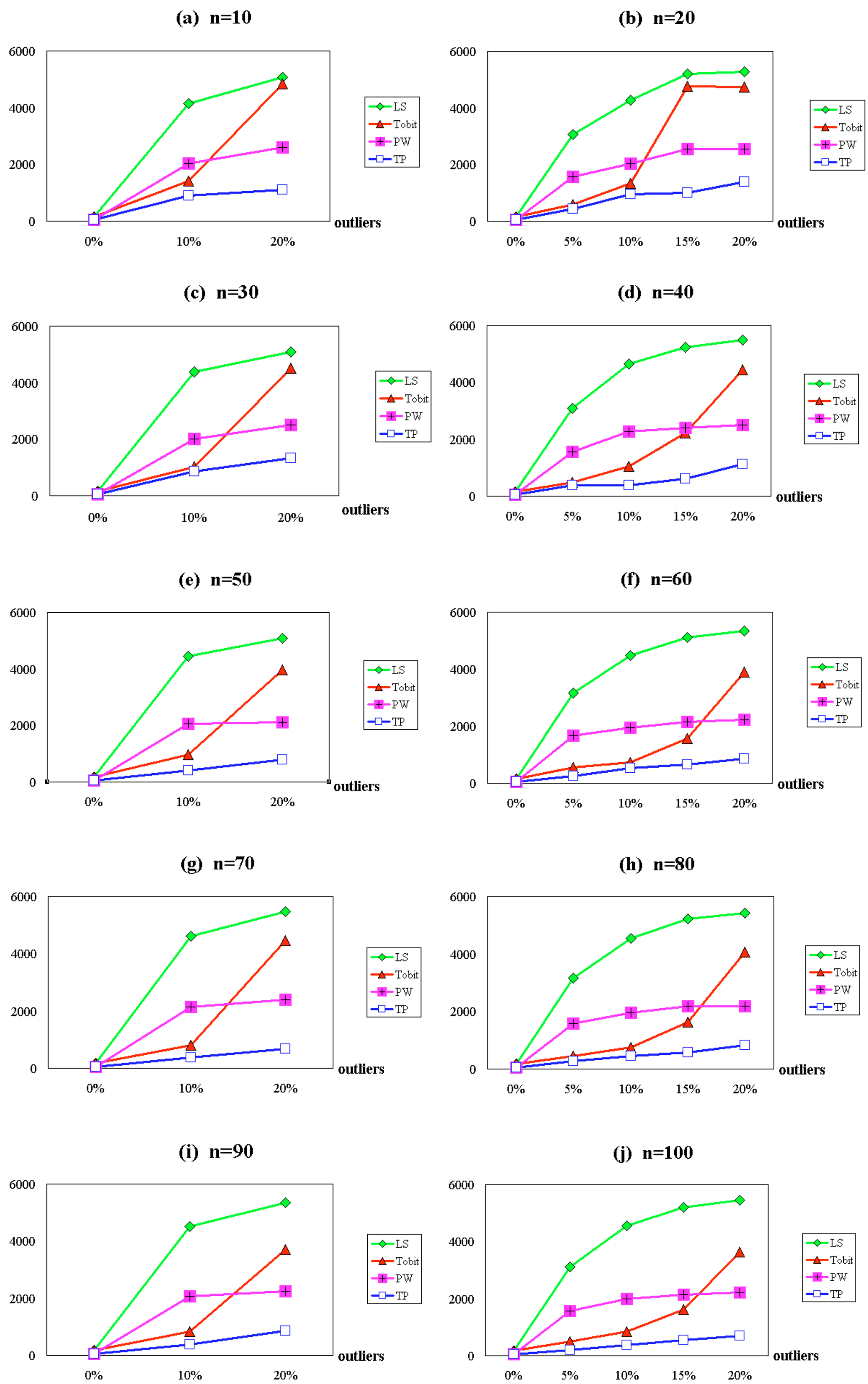

Figure 3 ASSR of four different regression models for Levenberg-Marquardt method varied by percentage of outliers when n=10, 20,..., 100 where xy-direction outliers exist

From Tables 3 and 4, where the percentages of outliers are 5%, 10% and 15%, it was found that the significantly smallest ASSR is of the Tobit-piecewise regression model, followed by that of Tobit, piecewise and LS regression models, in that order. Meanwhile, at 20% outliers, the smallest ASSR is from the Tobit-piecewise regression model, followed sequentially by that of piecewise, Tobit, and LS. These results indicate that, in the case where xy-direction outliers exist, the potential applicability Tobit regression decreases when the percentage of outliers increases. It can be seen that the piecewise regression model slightly changes when the percentage of outliers increases, as evident in Figures 3 and 4. Moreover, in the case of outliers existing in the xy-direction between 5% to 15%, it was found that not only the Tobit-piecewise regression model but also Tobit is preferable to both piecewise and LS. Meanwhile at 20% outliers, the piecewise regression is preferable to both Tobit and LS since, when the percentage of outliers is high, it indicates that data are more likely to be from two structures. These results are the same of Mekbunditkul's dissertation which each joined point in both of Tobit-piecewise and piecewise regression models are assumed to be known.

When looking at Figures 3 and 4, it was found that the ASSR of the four different regression models increases when the percentage of outliers increases, for all sample sizes. Furthermore, it was found that when the percentage of outliers increases, the ASSR of Tobit-piecewise and piecewise slightly increases.

Table 4 ASSR of four different regression models for Quandt's method in cases of xy-direction outliers

Sample Size	% of Y-Outliers	ASSR LS	Tobit	piecewise	Tobit-piecewise
10	5	-	-	-	-
	10	4,157	1,416	2,031	897
	15	-	-	-	-
	20	5,070	4,837	2,594	1,108
20	5	3,069	581	1,577	441
	10	4,270	1,337	2,053	951
	15	5,204	4,769	2,569	1,022
	20	5,286	4,733	2,573	1,398
40	5	3,077	495	1,561	389
	10	4,636	1,044	2,267	384
	15	5,245	2,221	2,413	612
	20	5,497	4,442	2,499	1,112
60	5	3,154	551	1,677	256
	10	4,473	741	1,963	521
	15	5,102	1,559	2,154	666
	20	5,354	3,900	2,228	862
100	5	3,118	493	1,577	204
	10	4,552	838	1,999	376
	15	5,196	1,615	2,135	542
	20	5,459	3,641	2,220	685

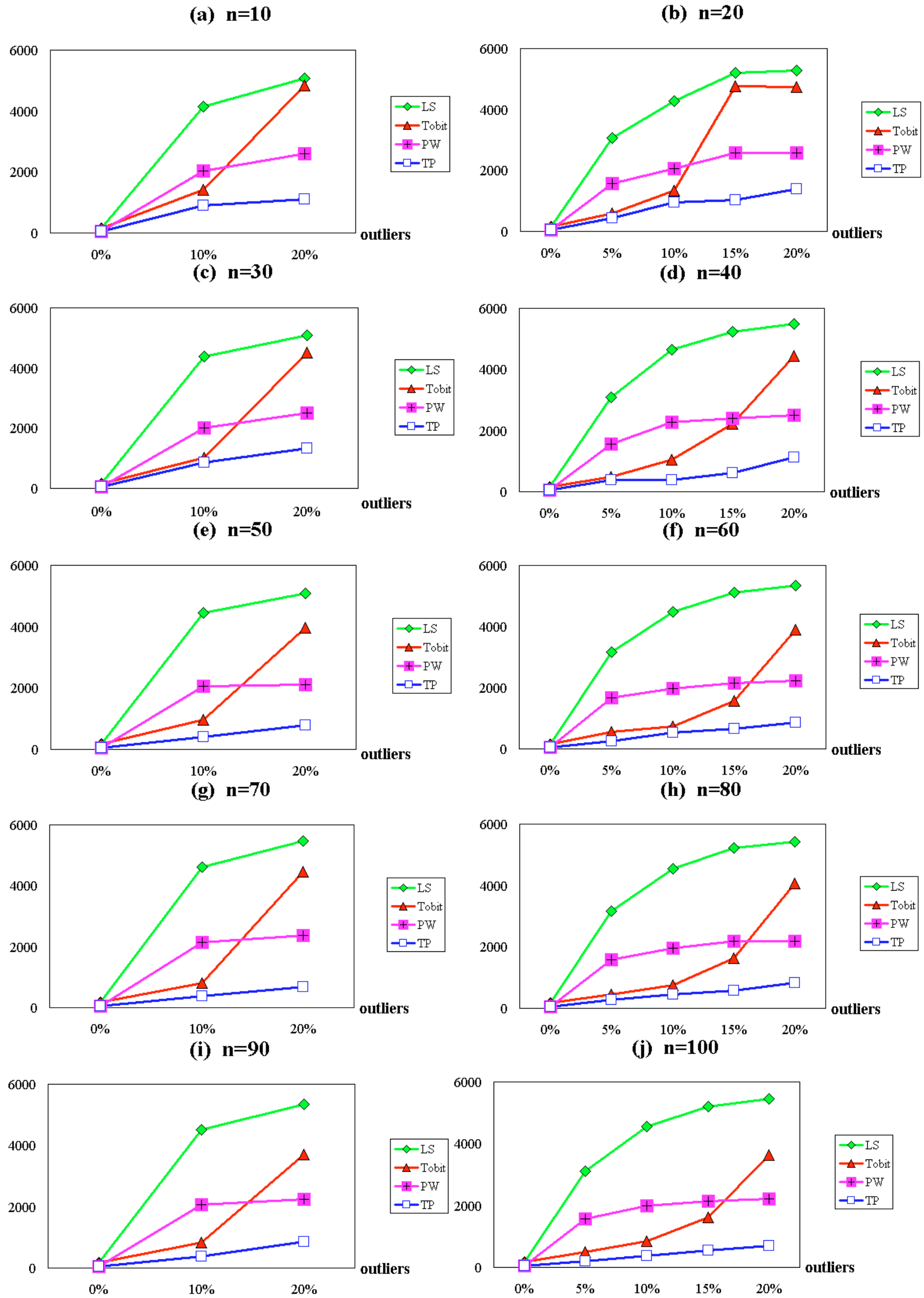

Figure 4 ASSR of four different regression models for Quandt's method varied by percentage of outliers when n=10, 20,..., 100 where xy-direction outliers exist

Conclusion

The important point of research finding is that the ASSR of Tobit-piecewise model with joined point estimated by Quandt's method does not significantly differ from the Levenberg-Marquardt method. The comparison of four different estimators such as LS, Tobit, piecewise and Tobit-piecewise, is described below.

Tobit-piecewise regression model, first introduced in Mekbunditkul's research, have been used to cope with the important problem in regression analysis such that data contain outliers. The Tobit-piecewise estimator is one of MLE although it is a biased estimator of a regression coefficient it is a consistent estimator and keeps all properties of MLE. The potential applicability of estimator was studied in sense of ASSR's and RE's, and was shown in the simulation results. From simulation studies, there were found that Tobit-piecewise regression model is the best. Nevertheless, the joined point in Tobit-piecewise regression model was not found in literatures. So that this paper, estimations of the unknown joined point in Tobit-piecewise are introduced. These are based on ML method introduced by Quandt and on nonlinear LS method suggested by Marquardt, called the Levenberg-Marquardt. When considering the simulation results shown in this paper, in case where y-direction outliers for all percentages of outliers considered and where xy-direction outliers for small percentage, by about 5% - 15%, of outliers considered, there is found that Tobit-piecewise regression model yields the smallest ASSR and sequentially by that of Tobit, piecewise and LS. Meanwhile, in situation of data containing xy-direction outliers with large percentage of outliers, by about 20%, Tobit-piecewise model gives the smallest ASSR and followed by piecewise, Tobit and LS, respectively.

Recommendation

There are two recommendations for further study arising from this study. First, estimators of the desired limited value in the TP regression model has not been included in the scope of this study but it could be attained in the future because they rather affect the fit of the regression line and the verification of ASSR. Second, a measure of error other than ASSR could be developed to investigate the potential applicability of the TP regression estimator. ASSR, as shown in Mekbunditkul's (2010) research, might not be suitable for the reason that ASSR is rather affected by limit values.

References

Goldfeld, S.M., Kelejian, H.H. and Quandt, R.E. (1971). Least Squares and Maximum Likelihood Estimation of Switching Regressions. *Econometric Research Program.* 130 (November).

Hudson, D.J. (1966). Fitting Segmented Curves Whose Join Points Have to be Estimated. *Journal of the American Statistical Association, 61*, 1097-1129.

Jöreskog, K.G. (2002). *Censored Variables and Censored Regression.* (Online). Available at http://www.ssicentral.com/lisrel/techdocs/censor.pdf

Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics, 11*, 431–441.

Mekbunditkul, T. (2010). An Alternative Estimator of Regression Coefficient with Outliers (*Doctoral dissertation)*. National Institute of Development Administration.

Poon,K.S. et. al. (2008). Analysis of structural change-point in multivariate regression models. *An application on IT productivity impact.*

Quandt, R.E. (1958). The estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association,* 53, 873-880.

Rosett, R.N. and Nelson, F.D. (1975). Estimation of the two-limit Probit regression model. Econometrica, 43, 141–146.

Rousseeuw, P.J. and Leroy, A. M. (1987). *Robust Regression and Outlier detection.* New York : John Wiley & Sons.

Suits, D., Marson, A. and Chan, L. (1978). Spline Functions Fitted by Standard Regression Methods. *Review of Economics and Statistics.* 132 – 139.

Tobin, J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica.* 26, 24-36.