RESEARCH REPORT

การประมาณค่าจุดเปลี่ยนในตัวแบบถดถอย โทบิท-พีซไวส์

AN ESTIMATION OF A JOINED POINT IN TOBIT-PIECEWISE REGRESSION MODEL

Titirut Thipbharos

THIS RESEARCH REPORT WAS GENERATED BY
DHURAKIJ PUNDIT UNIVERSITY
YEAR 2011

Title : An Estimation of a Joined Point in Tobit-Piecewise Regression Model
Researcher : Titirut Thipbharos, Ph.D. **Institution** : Dhurakij Pundit University
Year of Publication : 2011 **Publisher** : Dhurakij Pundit University
Sources : Dhurakij Pundit University **No. of page** : 82 pages

ABSTRACT

The objective of this study is to introduce estimation method of a joined point in Tobit-piecewise regression model which is first constructed in Mekbunditkul (2010). The factored likelihood method applied on TP model makes possible parameter estimation such as TP estimator which retains good properties of a MLE, e.g. consistency and best asymptotically normal (B.A.N). There are two interested foundations, in this research, introduced to estimate the unknown joined point in Tobit-piecewise model in the case that observed data contain outliers. Those estimators are obtained under the maximum likelihood method, for example Quandt's method and under the nonlinear least square method, namely Levenberg-Marquardt method. The numerical analysis and simulation study are provided to compare the potential applicability of each estimator considered in terms of the average sum of squares of residual (ASSR) and relative efficiency (RE). It is found that Tobit-piecewise regression model with the joined point estimated by nonlinear LS method yields non-significant smaller ASSR than by ML method in every situation. In addition, for this particular case, Tobit-piecewise estimator is the best among all four different estimators.

ACKNOWLEDGEMENTS

I am greatly appreciative of Dhurakij Pundit University for giving me this opportunity and the benevolent financial support for the study. Thanks are also dedicated to people who concerned me a scholarship such as the Dean of Arts and Sciences Faculty and all of committees and staffs of Research Center of Dhurakij Pundit University.

I would like to express my gratitude to Associate Professor Dr. Sorachai Bhisalbutra, who guided my research towards being a valuable study with beneficial applications. I am sure that I have forgotten some names of contributors to this study (at least all my friends). I would like to express my deep thanks to all persons, if they have been mentioned above or not, for their energy and help, spirit, patience, and co-operation.

Finally, I am grateful to my family, namely my father, mother, sister and brothers, who have always provided encouragement to me and of course, a very special thanks to my husband, Piyavit Thipbharos, and son for the moral support and motivation to finish this research.

Titirut Thipbharos, Ph.D.

November 2011

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figures **Page**

CHAPTER 1

INTRODUCTION

1.1 Background

A statistical analysis used for analyzing and indicating the relationship between a dependent/response and independent/predictor variables is called regression analysis. It is now presentable a change on one of the variables in correspondence with a change in the other. In addition, the regression model can estimate or predict the value of response variable when knowing the value of predictor variables. A linear or straight line relationship can be written as follow:

$$\underset{\sim}{Y} = X\underset{\sim}{\theta} + \underset{\sim}{\varepsilon}, \tag{1.1}$$

where $\underset{\sim}{Y}$ is an $n \times 1$ vector of dependent/response variable, X is an $n \times (k+1)$ matrix of independent/predictor variables, $\underset{\sim}{\theta}$ is a (k+1)x1 vector of unknown parameters, $\underset{\sim}{\varepsilon}$ is an $n \times 1$ vector of errors where it element, ε_i, is independent identically distributed as normal with zero mean and constant variance. It is also assumed that matrix X is of full rank, i.e., rank(X) is p = k+1 and it is less than the sample size n. If $\underset{\sim}{\varepsilon}$ are $\mathrm{NID}\left(\underset{\sim}{0}, \sigma^2 I_n\right)$, then the least square (LS) estimator of $\underset{\sim}{\theta}$ is the same as the maximum likelihood (ML) estimator and it is $\hat{\underset{\sim}{\theta}} = \left(X'X\right)^{-1}\left(X'\underset{\sim}{Y}\right)$. In addition, under Gauss-Markov assumptions, LS method yields the best linear unbiased estimator (BLUE) of $\underset{\sim}{\theta}$.

Whenever the assumptions are violated, e.g. non-normality, heteroscedasticity (variances of the errors are not constant), and non-linearity, the LS estimator would not be preferable. Another important problem is that the observed data contain outliers. These outliers may have a large effect on the estimated value, Heteroscedasticity problem in regression analysis might be caused by outliers in $y-direction$ and/or $x-direction$ (Rousseeuw and Leroy, 1987: 3-59). Hence, a better way to estimate the parameters $\underset{\sim}{\theta}$ in the regression model (1.1) are needed.

1.2 Statement of the Problem

The least square method is a mathematical way to make the “magnitude” of random errors as small as possible. Magnitude of errors are measured as the square of n terms of $e_i = y_i - \underset{\sim}{x}_i \hat{\underset{\sim}{\theta}}$ and add them up. Then the $\hat{\underset{\sim}{\theta}}$ minimizing the resulting sum of squared errors is called LS estimator. LS approach, a traditional approach to regression analysis, attains the BLUE estimator when Guass-Markov assumptions hold. In practice, it often found that the normal distribution might not be in line with the assumption caused by some observations being quite far from the bulk of the data. These observations are called outliers. These are data points not typically located close to the usual data (Montgomery and Peck, 1982:70). 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. Nevertheless, they exclude unusual incidents. It is often found in practice that outliers can have a large distorting influence on LS estimate, for example as shown in figure 1.1.

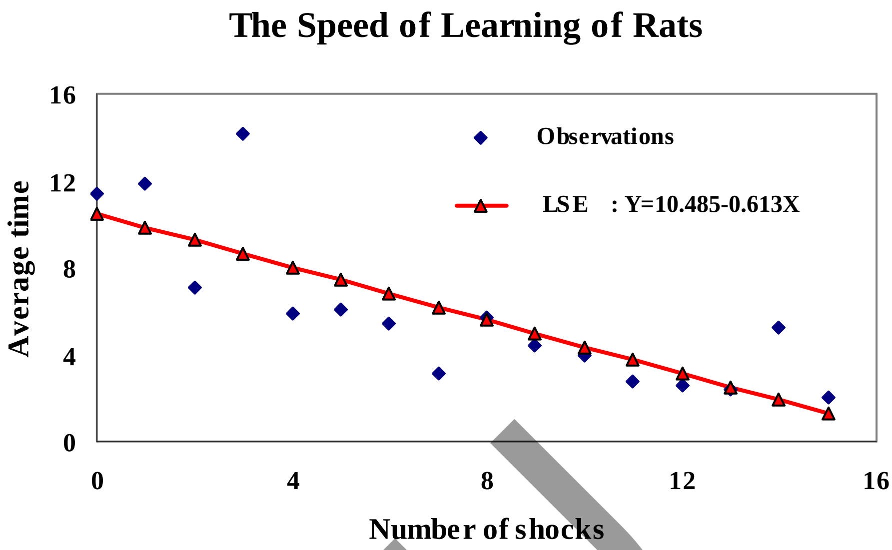


Figure 1.1 LSE fitting with Shock data

Source of Data: Maronna, Martin and Yohai, 2006: 87-88.

To cope with the existent 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. Another two ideas with the different benefit first considered in Mekbunditkul's earlier research (Mekbunditkul, 2010) can be concluded as following: First, Tobit regression is a tool used to investigate the linear relationship when the dependent variable in a regression model is limited. This concept is taken into account for this study in the sense that putting limited value at some desired variable can reduce effect value of outliers in *y-* and *xy-directions*. However, the existence of other types of outliers has not been manipulated. Second, piecewise regression is a regression analysis properly applied when 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 consisting of outliers in *y-directions*.

Another approach, termed TP (abbreviated from Tobit-piecewise) regression, employs a fitting criterion to unusual data that is not as contained as LS. An alternative approach, in addition, is used to reduce or down-weigh value of outliers. This research applied Tobit and piecewise regression to TP regression model (Mekbunditkul, 2010). She constructed the TP regression model by the combination of the Tobit and piecewise regression models. Moreover, there was first found the evidence that the Tobit model (Tobin, 1958, Rosett, 1975 and Jöreskog, 2002), limited by some desired variables could reduce the effect of outliers in some situations in the study of Mekbunditkul. Nevertheless, censoring the data set with one value of either lower or upper limit might not be suitable, thus we need to filter outliers with more than one limiting value. It is dependent on the structure of the whole data. 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, one data set is fit with two regression regimes when a single regression is inadequate; the structural change is then taken into account as they should be. This 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).

Moreover, according to the evidence in simulation results of Mekbunditkul's dissertation, we found that: TP regression model can reduce the effect value of outliers and can utilize the data with structural change more effectively than piecewise, Tobit and LS. Nevertheless, there was not any study regarding an estimation of joined point in TP regression model. Therefore, in this research, the matter is studied. Considering figure 1.1, three obvious outlier data affect the LS regression drawn away from the bulk of the data. This means that the LS regression might not be preferable for the particular case. Whilst these data are analyzed by both Tobit and piecewise regression models, they yield better results than LS regression. Moreover, we can see that TP regression model yields the best among all four different regression models. Therefore, instead of using LS, Tobit or piecewise, we use TP regression model to fit the data consisting of outliers in the sense that TP regression model can reduce the effect of outliers better than Tobit and piecewise.

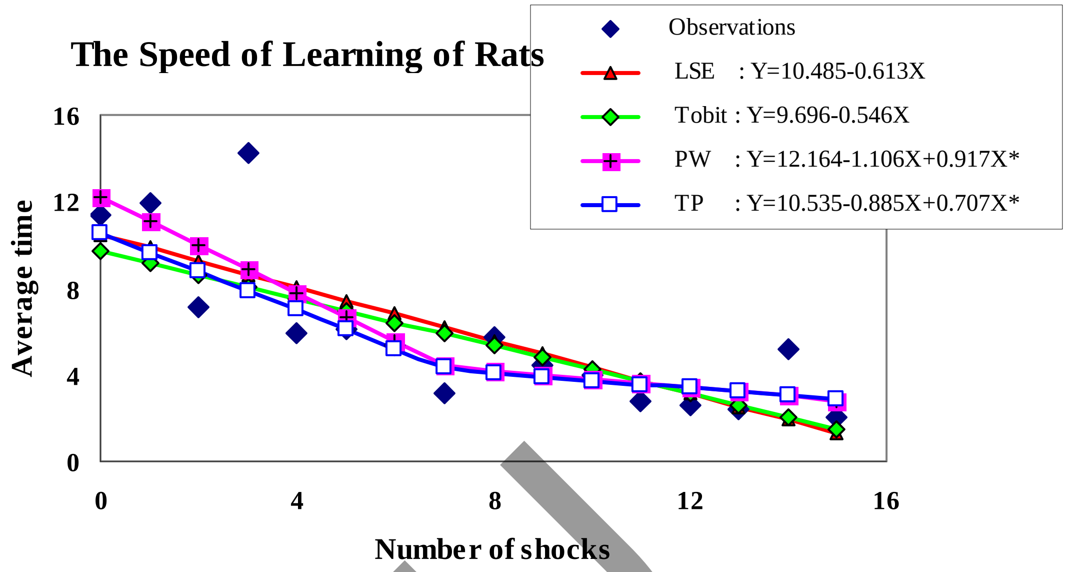


Figure 1.2 Four different regression models fitting with Shock data

Source of Data: Maronna, Martin and Yohai, 2006: 87-88.

Figure 1.2, in this particular case, shows that TP regression model seemingly results better than any other. Therefore, TP regression model is an alternative robust method for those situations where outliers exist. Such "belief" is strongly supported by the data set in the figure above. Nevertheless, there has been little to none literature that describes the estimation method for the joined point in TP regression model. As a result, this point is now first being focused throughout the research.

1.3 Objectives of the Study

To summarize, objectives of the study are as followed:

1) To estimate the joined point in TP regression model.

2) To apply the TP regression model with the real data, for example, socio-economic survey data. This research on the joined point in TP regression model can be estimated by two approaches such as ML based, Quandt's method, and LS based, Levenberg-Marquardt method,

3) To compare two estimation methods of joined point such as ML based and nonlinear LS based. The comparison will be done by both simulation and empirical data analysis.

1.4 Scope of the Study

The TP regression model achieved from the combination of two principal ideas, namely 1) Tobit regression, and 2) piecewise regression. Two estimation methods for the joined point in TP regression model are compared by means of the average sum of square residual (ASSR).

This study is restricted by the following statements:

1. The regression model (1.1)
2. $\underset{\sim}{\varepsilon}$ is normally distributed with the zero mean vector and covariance matrix Σ, where Σ is a diagonal matrix which is positive definite

3. The estimators of regression coefficients in TP regression model are derived by using maximum likelihood estimation (Mekbunditkul, 2010)

4. Two estimation methods for the joined point in TP regression model can be compared the performance by ASSR of regression model. Performances of four different estimators are compared by using the Monte-Carlo simulation. Moreover, these method will also applied to the survey data from the socio-economic survey in Thailand. The four different estimators are LS, Tobit, piecewise and TP.

5. ASSR of regression model is defined as follow.

$$\mathrm{ASSR}_j = \left(d_j + \sqrt{\frac{1}{2} d_j^2 + 2 f_j} \right)^2, \qquad j = a, b,$$

where $d_j = \dfrac{\underset{\sim}{1}'\left(\underset{\sim}{\hat{Y}}_{j1} - \underset{\sim}{L}_{j1}\right)}{n_{j2}}$, and $f_j = \dfrac{\left(\underset{\sim}{\hat{Y}}_{j1} - \underset{\sim}{L}_{j1}\right)'\left(\underset{\sim}{\hat{Y}}_{j1} - \underset{\sim}{L}_{j1}\right)}{n_{j2}} + \dfrac{\left(\underset{\sim}{Y}_{j2} - \underset{\sim}{\hat{Y}}_{j2}\right)'\left(\underset{\sim}{Y}_{j2} - \underset{\sim}{\hat{Y}}_{j2}\right)}{n_{j2}}$

(Mekbunditkul, 2010).

CHAPTER 2

THEORY

2.1 Outliers

According to Barnett, V. et al.(1994) (access by http://en.wikipedia.org), an outlier in the sense of statistics is an observation that is numerically distant from the rest of the data. They can occur by chance in any distribution, but they are often indicative either of measurement error or that the population has a heavy-tailed distribution. In the former case one wishes to discard them or use statistics that are robust to outliers, while in the latter case they indicate that the distribution has high kurtosis and that one should be very cautious in using tool or intuitions that assume a normal distribution. Outliers, being the most extreme observations, will include the sample maximum or sample minimum, or both, depending on whether they are extremely high or low. However, the sample maximum and minimum need not be outliers if they are not unusually far from other observations. This definition is similar to others such as Montgomery, et al. (1982), Moore, et al.(1999), Monhor, D. et al.(2005) and Walfish, S. (2005), etc.

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. Nevertheless, they exclude unusual incidents of outliers. Moreover, types of regression outliers (Rousseeuw and Zomeren, 1990: 633) studied in this research are as below:

1) *y-direction* outliers

These are points that are outliers only because they are extreme *y-coordinates*. The extent to which such outliers will affect the parameter is estimated depending on both their *x-coordinate* and the general configuration of other points. Thus, those points could also be a regression outliers or residual outliers.

2) ***x-direction* outliers**

These are points that deviate only with regard to the *x-coordinates*. Such points can cause some regression estimates to perform poorly. The *x-direction* outliers could also be regression or residual outliers.

3) ***xy-direction* outliers**

These are points outlying in both *x*- and *y*- *coordinates*. It may be a regression outliers or residual outliers (Ryan, 1997: 350).

Classical estimate such as the sample mean, variance, covariance and correlation, or the LS fit of a regression model, can be very adversely influenced by outliers, even by a single one, and often fail to provide good fits to the bulk of the data. An alternative approach such as robust approach has been introduced to cope with outliers' problem in order to provide a good fit for the bulk of the data containing outliers, as well as when the data are free of them. Nevertheless, robust approach is quite complicate.

2.2 LS Regression

The second topic to be studied is the regression analysis which is well concluded by Ampanthong (2009: 10-12). Accordingly, regression analysis is a statistical tool for modeling and analyzing several variables which underlie vital assumptions. To estimate the unknown parameters in a regression model is among the most significant objectives of regression analysis. This process is also called "fitting the data to the model". According to Gauss-Markov theorem, the maximum likelihood estimate of $\underset{\sim}{\theta} = (\alpha, \beta_1, \beta_2, ..., \beta_k)$ turns out to be the BLUE of $\underset{\sim}{\theta}$. One usually tries to estimate the unknown parameters in a regression model from a data set by the LS method to obtain $\hat{\underset{\sim}{\theta}} = (X'X)^{-1}(X'\underset{\sim}{Y})$. When the method is applied to acquire the estimates $\hat{\theta}_j$ of θ_j, for j=1, 2,… , k, those so found are called LS estimates of the regression coefficients. The fitted regression equation concluded from the data set is $\hat{y}_i = \hat{\alpha} + x_{i1}\hat{\beta}_1 + x_{i2}\hat{\beta}_2 + ... + x_{ik}\hat{\beta}_k$ for i=1, 2, …, n.

This equation is regarded as the estimate of the regression model $\underset{\sim}{Y} = \underset{\sim}{x}_i \underset{\sim}{\theta}$, where $\underset{\sim}{x}_i = \left(x_{i1}, x_{i2}, \ldots, x_{ik} \right)$, i=1, 2, …, n. The residual r_i is defined as the difference between the observed value y_i and the fitted value $\hat{y}_i$, i.e. $r_i = y_i - \hat{y}_i$. The method of obtaining the LS of $\hat{\theta}_j$, for each j=1, 2,…, p is the most popular estimate method. The LS estimator that minimizes the sum of squared residuals, i.e. $\hat{\beta}_1,\ \hat{\beta}_2,\ \ldots,\ \hat{\beta}_k$ will have $\underset{\hat{\underset{\sim}{\theta}}}{\mathrm{Min}} \sum_{i=1}^{n} r_i^2 = \underset{\hat{\underset{\sim}{\theta}}}{\mathrm{Min}} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2$. Nevertheless, LS estimator is not suitable when the distribution of their residuals is not normal. In some case, when data have a heavy tail in any direction due to the presence of outliers, LS method of estimate might not be preferable. While in other cases, data may have only a small fraction of outliers, but LS estimate is still not suitable for further analysis. A small fraction of outliers may have a large effect on the LS estimator.

2.3 Two-Limit Tobit Model

The two-limit Tobit model (Tobin, 1958) is the simplest model for censored data. Here, we stimulate the discussion using an example based on Tobin's application of the model. Let a dependent variable be the monthly expenditure on luxury goods of each household and let an independent variable or explanatory variable be such as the monthly income for the corresponding the household's monthly expenditure. The parameters vector $\underset{\sim}{\theta}$, which contains the set of population regression parameters related to the variables, need to be estimated. In this example, link variable or Y_i^* might be the capacity of households to spend their income on luxury goods, but this is only realized as actual expenditure, Y_i, if that expenditure exceeds zero. Thus, even if many observations might have value to be zero on the Y_i, they can be considered as having changing values on the link variable Y_i^*. The two-limit Tobit model (Tobin, 1958: 26 , Rosett, 1975: 141 and Jöreskog, 2002: 13) can be written as the dependent variable Y_i satisfies

$$Y_i = \begin{cases} L & ; \quad Y_i^* \le L \\ Y_i^* & ; \; L < Y_i^* < U \\ U & ; \quad Y_i^* \ge U, \end{cases} \tag{2.1}$$

where Y_i^*, for $i = 1,2,...,n$, is the link function generated by the linear regression model

$$Y_i^* = \alpha + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i ,$$

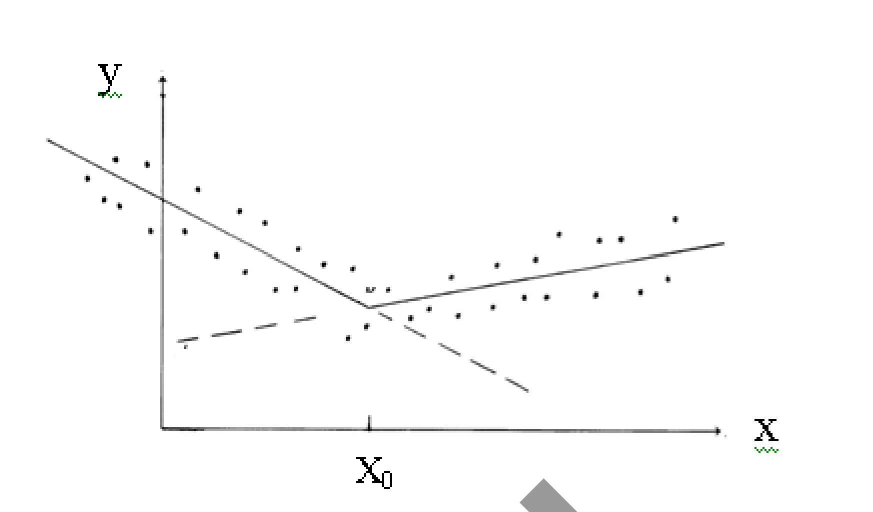
where $x_1, x_2, ..., x_k$ are regressors, and ε_i 's are the error terms having independent normal distributions with zero mean and constant variance $\left(\varepsilon_i \sim \text{i.i.d.}N\left(0,\sigma^2\right)\right)$ and are independent of $\underset{\sim}{x}_i$. L and U in the model (2.1) are an lower and upper limits, respectively.

The probability density function (p.d.f.) of Y for given values of each L and U is determined by $f_Y(Y_i) = \Phi\left(\frac{L - \underset{\sim}{x}_i \underset{\sim}{\theta}}{\sigma_i}\right)$ if $Y_i = L_i$, $f_Y(Y_i) = \frac{1}{\sigma_i}\phi\left(\frac{Y_i - \underset{\sim}{x}_i \underset{\sim}{\theta}}{\sigma_i}\right)$ if $Y_i = Y_i^*$, and $f_Y(Y_i) = 1 - \Phi\left(\frac{U_i - \underset{\sim}{x}_i \underset{\sim}{\theta}}{\sigma_i}\right)$ if $Y_i = U_i$. Where Φ and ϕ are the cumulative distribution function (c.d.f.) and the p.d.f. of a standard normal distribution, respectively. From the p.d.f. of Y, we then get the log-likelihood function and by the ML fashion, the Tobit estimator was constructed. Actually, LS fashion might be inappropriate in the case that the dependent variable is limited by some desired variable. As the mention of Greene (1981: 505-513) who described that the LS estimator of parameter vector in Tobit model is the bias and also the asymptotic bias of the regression coefficients, this means that the LS based for the limited dependent variable case is inconsistent.

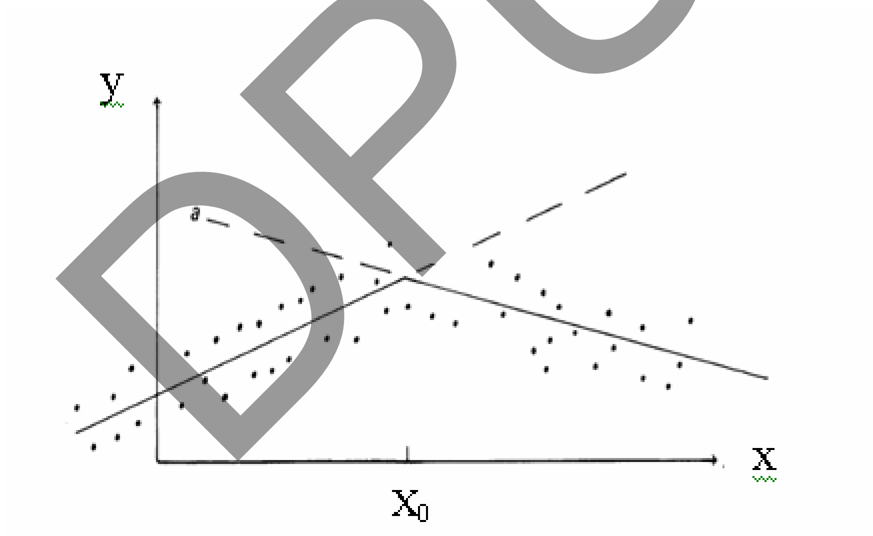
2.4 Piecewise Regression Model

In this research, the structural change in the regression model is taken into account, thus the piecewise regression model (Quandt, 1958: 874, Hudson, 1966: 1097-1129, Goldfeld, Kelejian and Quandt, 1971, Suits, Mason and Chan 1978: 132-133) is considered. Quandt first introduced that economic variables may sometimes be fitted by linear relations with the property that the parameters of the relationship are subject to discontinuous changes. For example, consider the consumption function $Y = \alpha X + \beta$. Aggregate consumption depends upon the level of aggregated income. In addition, it may be hypothesized that consumption depends non-linearly on other factors such as the state of expectations concerning the future of the economy, the volume of installment buying, the level of the interest rate, etc. These other variables may have the effect of altering the parameters of the consumption function in the following fashion: when the critical outside variable i satisfies $i < i^*$ then $Y = \alpha_1 X + \beta_1$ and when $i \geq i^*$ then $Y = \alpha_2 X + \beta_2$, where i^* is the critical level of the outside variable in question. In general, one may not be able to identify the critical outside variable and one may not be able to state at what time the system $Y = \alpha X + \beta$ changes from one regime to the other. In the paper of Quandt, there was indicated an estimating procedure for the switching point under the conditions when it is known that the time period under consideration contains a single switching. Parameters in the piecewise regression model were estimated by the ML method.

Subsequently, Hudson (1966: 1097-1129) discussed a similar estimate problem in which the two regression regimes are required to be intersected, see example on figure 2.1. Parameters estimate based on the LS method and the models were assumed to be joined at the value x_0.



(a) Model (2.14a).



(b) Model (2.14b).

Figure 2.1 Two possible types of the Piecewise Regression Model
Source: Tishler and Zang, 1981: 117.

Two regression regimes joined at point υ (Hudson, 1966) can be represented by

$$Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_i & ; \ x_i \le \upsilon \\ \alpha_2 + \beta_2 x_i + \varepsilon_i & ; \ x_i > \upsilon \end{cases} \tag{2.2}$$

where Y_i is a dependent variable, x_i is a corresponding independent variable, and error terms, ε_i's are assumed to be normally and independently distributed with mean zero and variance σ^2 and are independent of the independent variable. In addition, the model (2.2) is subject to $\alpha_1 + \beta_1\upsilon = \alpha_2 + \beta_2\upsilon$ and can be written as

$$Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_i & ; \ x_i \le \upsilon \\ \alpha_1 + \beta_1 x_i + (\beta_2 - \beta_1)(x_i - x_0) + \varepsilon_i & ; \ x_i > \upsilon. \end{cases} \tag{2.3}$$

Suits et al. (1978: 132-133) extended model (2.3) so that it can be written in the multiple regression model with a dummy variable, D_i, consisting of two independent variables as follows;

$$Y_i = \alpha_1 + \beta_1 x_i + \beta_2 x_i^* + \varepsilon_i, \tag{2.4}$$

where $x_i^* = (x_i - \upsilon)D_i$ and $D_i = \begin{cases} 0 & ; x_i \le \upsilon \\ 1 & ; x_i > \upsilon \end{cases}$.

2.5 Methods for Finding the Minimum of the Sum of Squares

This research is related to solve nonlinear least square problem. Therefore, there are three methods introduced to find the minimum of sum of squares for that problem. Nonlinear least square problems occur for instance in nonlinear regression namely piecewise and TP regression as considered in this study.

2.5.1 Gauss-Newton Method

The Gauss-Newton approximation (Seber and Wild, 1988: 25) is described as the followings. Suppose $\underset{\sim}{\theta}^{(a)}$ is an approximation to the LS estimator $\hat{\underset{\sim}{\theta}}$ of a nonlinear model. By the Taylor's expansion, we will get

$$f\left(X;\underset{\sim}{\theta}\right) \approx f\left(X;\underset{\sim}{\theta}^{(a)}\right) + F_{.}^{(a)}\left(\underset{\sim}{\theta} - \underset{\sim}{\theta}^{(a)}\right), \tag{2.5}$$

where $F_{.}^{(a)}$ is $F_{.}\left(X;\underset{\sim}{\theta}^{(a)}\right)$. It can be applied to the residual vector, $r\left(X;\underset{\sim}{\theta}\right)$, as

$$\begin{aligned} r\left(X;\underset{\sim}{\theta}\right) &= \underset{\sim}{Y} - f\left(X;\underset{\sim}{\theta}\right) \\ &= r\left(X;\underset{\sim}{\theta}^{(a)}\right) - F_{.}^{(a)}\left(\underset{\sim}{\theta} - \underset{\sim}{\theta}^{(a)}\right). \end{aligned}$$

From the equation $S\left(\underset{\sim}{\theta}\right) = \sum_{i=1}^{n}\left(Y_i - f\left(\underset{\sim}{x}_i;\underset{\sim}{\theta}\right)\right)^2$, we will get

$$\begin{aligned} S\left(\underset{\sim}{\theta}\right) \approx\; & r'\left(X;\underset{\sim}{\theta}^{(a)}\right) r\left(X;\underset{\sim}{\theta}^{(a)}\right) - 2r'\left(X;\underset{\sim}{\theta}^{(a)}\right) F_{.}^{(a)}\left(\underset{\sim}{\theta} - \underset{\sim}{\theta}^{(a)}\right) \\ & + \left(\underset{\sim}{\theta} - \underset{\sim}{\theta}^{(a)}\right)' F_{.}^{(a)\prime} F_{.}^{(a)}\left(\underset{\sim}{\theta} - \underset{\sim}{\theta}^{(a)}\right). \end{aligned} \tag{2.6}$$

Thus, we can conclude that the right hand side of the approximation (2.6) is minimized with respect to $\underset{\sim}{\theta}$ when

$$\begin{aligned} \underset{\sim}{\theta} - \underset{\sim}{\theta}^{(a)} &= \left(F_{.}^{(a)\prime} F_{.}^{(a)}\right)^{-1} F_{.}^{(a)\prime}\, r\left(X;\underset{\sim}{\theta}^{(a)}\right) \\ &= \underset{\sim}{\delta}^{(a)}. \end{aligned} \tag{2.7}$$

This equation gives the approximation of $\underset{\sim}{\theta}^{(a)}$ then we can say that the next approximation is followed by

$$\underset{\sim}{\theta}^{(a+1)} = \underset{\sim}{\theta}^{(a)} + \underset{\sim}{\delta}^{(a)}. \quad (2.8)$$

The equations (2.7) and (2.8) determine the updating result and the $\hat{\underset{\sim}{\theta}}$ can be attained by the equation (2.8). In addition, Seber and Wild (1988) mentioned that the Gauss-Newton algorithm is convergent.

2.5.2 Steepest Descent Method

The steepest descent method is also known as the gradient descent. It is based on the gradient of $\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon}$. Seber and Wild (1988: 594) described the theory of this method as the followings. The steepest descent method is one of iterative processes where an initial guess $\underset{\sim}{\theta}^{(1)}$ is furnished, from which the algorithm sequentially moves in $\Re^p$ of points $\underset{\sim}{\theta}^{(2)}$, $\underset{\sim}{\theta}^{(3)}$,… which are aimed to converge to a local minimum $\hat{\underset{\sim}{\theta}}$. Practically useful is this algorithm method to make sure that $h(\underset{\sim}{\theta})$, a real-valued function of p parameters vector $\underset{\sim}{\theta} = (\theta_1, \theta_2, ..., \theta_p)'$, is reduced at each iteration so that $h\left(\underset{\sim}{\theta}^{(a+1)}\right) < h\left(\underset{\sim}{\theta}^{(a)}\right)$. In this research the function $h(\underset{\sim}{\theta})$ is defined as $S(\underset{\sim}{\theta}) = \sum_{i=1}^{n} \left(Y_i - f(\underset{\sim}{x}_i; \underset{\sim}{\theta})\right)^2$.

The approximation of the (a+1)th iterative is the same as (2.8) nevertheless the updating $\underset{\sim}{\delta}^{(a)}$, the computation of the ath step, is differently defined as

$$\underset{\sim}{\delta}^{(a)} = \rho^{(a)} \underset{\sim}{d}^{(a)}, \quad (2.9)$$

where the vector $\underset{\sim}{d}^{(a)}$ is called the direction of the step and $\rho^{(a)}$ is the step length. Frequently $\rho^{(a)}$ is selected to approximately minimize $S(\underset{\sim}{\theta})$ along the line $\underset{\sim}{\theta}^{(a)}+\rho\underset{\sim}{\delta}^{(a)}$, this process known as a line search. If the $\rho^{(a)}$ is the exact minimum at each iteration, then the algorithm is said to have exact line searches; otherwise, it employs approximate line searches. Whenever there exists the convergent of $S(\underset{\sim}{\theta})$ to the local minimum, this means that the exact line searches are not manipulated and $S(\underset{\sim}{\theta})$ is sufficiently reduced at each iteration (Gill, 1991: 100). A descent direction $\underset{\sim}{d}^{(a)}$ can be determined by

$$\underset{\sim}{g}^{(a)\prime}\underset{\sim}{d}=\left.\frac{\partial S\left(\underset{\sim}{\theta}^{(a)}+\rho\underset{\sim}{d}\right)}{\partial\rho}\right|_{\rho=0}<0 \qquad (2.10)$$

By the Taylor expansion, we get

$$S\left(\underset{\sim}{\theta}^{(a)}+\rho\underset{\sim}{d}\right)\approx S\left(\underset{\sim}{\theta}^{(a)}\right)+\rho\underset{\sim}{g}^{(a)\prime}\underset{\sim}{d}+O\left(\rho^{2}\right). \qquad (2.11)$$

Thus, when the approximation (2.11) attains, the decrease in function $S\left(\underset{\sim}{\theta}^{(a)}\right)$ can be obtained by the small enough step ρ in the direction $\underset{\sim}{d}$. The important theorem is stated in order to know "how descent directions can be calculated?" as followed.

Theorem 1 A direction $\underset{\sim}{d}$ is a descent direction at parameters vector $\underset{\sim}{\theta}$ if and only if there exists a positive definite matrix R such that

$$\underset{\sim}{d}=-R\underset{\sim}{g}.$$

Proof . It is available on Nonlinear Regression (Seber and Wild, 1988: 595).

Therefore from equations (2.8) and (2.9), we can state that

$$\underset{\sim}{\theta}^{(a+1)}=\underset{\sim}{\theta}^{(a)}-\rho^{(a)}R^{(a)}\underset{\sim}{g}^{(a)}. \qquad (2.12)$$

Where the choice of R is $R = I_p$ the descent direction becomes to $\underset{\sim}{d} = -\underset{\sim}{g}^{(a)}$ and this is called the steepest descent direction. However, the direction of steepest descent depends entirely upon the scaling of $\underset{\sim}{\theta}$.

2.5.3 **Levenberg-Marquardt Method** (Seber and Wild, 1988)

This method is a compromise between the Gauss-Newton and steepest descent methods. As $\underset{\sim}{d} \to 0$, the direction approaches Gauss-Newton. As $\underset{\sim}{d} \to \infty$, the direction approaches steepest descent. Levenberg-Marquardt 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.

Let a model be fitted into the data, there is likely the function $f\left(\underset{\sim}{x}_i;\underset{\sim}{\theta}^*\right)$ as expressed in the model $Y_i = f\left(\underset{\sim}{x}_i;\underset{\sim}{\theta}^*\right) + \varepsilon_i$, The problem is to compute the estimates of parameters which will minimize $S(\underset{\sim}{\theta}) = \left\|\underset{\sim}{Y} - f\left(\underset{\sim}{x}_i;\underset{\sim}{\theta}\right)\right\|^2$. By utilizing ideas of Levenberg together with Marquardt (1963), thus the Levenberg-Marquardt algorithm adaptively varies the parameter updates between the gradient descent and Gauss-Newton update (Seber and Wild, 1988)

$$\underset{\sim}{\delta}^{(a)} = -\left(\dot{F}^{(a)\prime}\,\dot{F}^{(a)} + \eta^{(a)}D^{(a)}\right)^{-1}\dot{F}^{(a)\prime}\, r\left(X;\underset{\sim}{\theta}^{(a)}\right), \qquad (2.13)$$

where $D^{(a)}$ is a diagonal matrix with positive diagonal element frequently defined to be the same as $\dot{F}^{(a)\prime}\,\dot{F}^{(a)}$ and $\eta^{(a)}$ is the a^{th} step direction. When $D^{(a)}$ is I_p and $\eta^{(a)} \to 0$, the direction approaches Guass-Newton. Whereas, $\eta^{(a)} \to \infty$, the direction approaches steepest descent. In the case that $\eta^{(a)} > 0$ then $\dot{F}^{(a)\prime}\,\dot{F}^{(a)} + \eta^{(a)}D^{(a)}$ is positive definite, as $D^{(a)}$ is positive definite. Thus the updating function $\underset{\sim}{\delta}^{(a)}$ as in (2.13) and by the Theorem 1, is a descent direction. And in the case that

$\eta^{(a)} \to \infty$ then $\underset{\sim}{\delta}^{(a)}$ tend to zero. Therefore if we select very large $\eta^{(a)}$ then we can so fast reduce the function $S(\underset{\sim}{\theta}) = \left\| \underset{\sim}{Y} - f\left(\underset{\sim}{x}_i; \underset{\sim}{\theta}\right)\right\|^2$. Nevertheless, for many iterations $\eta^{(a)}$ which are too large, the algorithm adapt with little progress.

Marquardt's finding indicates that the average angle between Gauss-Newton and steepest descent directions is about 90 degree. A choice of initial value of the direction $\eta^{(0)}$ for this research is $\eta^{(0)} = 10^{-3}$ used to start and compute the updating vector $\underset{\sim}{\delta}^{(a)}$. If $S\left(\underset{\sim}{\theta}^{(a)} + \underset{\sim}{\delta}^{(a)}\right) < S\left(\underset{\sim}{\theta}^{(a)}\right)$, then η becomes $\frac{\eta}{10}$ for the next iteration. Otherwise $S\left(\underset{\sim}{\theta}^{(a)} + \underset{\sim}{\delta}^{(a)}\right) > S\left(\underset{\sim}{\theta}^{(a)}\right)$, then η is 10η for the next iteration.kk

2.6 TP Regression Model

There was first interested the combination of two principal ideas, i.e. Tobit and piecewise regression, where each has a different benefit as described before. And there were not any literatures which applied these two ideas to cope with the outliers problem until Mekbunditkul (2010) first introduced the TP (abbreviated from Tobit-piecewise) regression as the derivation of the TP regression model and the log-likelihood function of $\underset{\sim}{\theta}$ described below:

According to the two-limit Tobit model (2.1) and by assuming that there exists the structural change in regression parameter, the piecewise multiple linear regression (Quandt, 1958: 874, Hudson, 1966: 1097-1129, Goldfeld and Quandt, 1971, Suits et al. 1978: 132-133) is utilized. The link function as mentioned in the model (2.1), Y_i^* can be broken into two regression regimes as

$$Y_i^* = \begin{cases} \alpha_1 + \beta_{11}x_{i1} + \beta_{12}x_{i2} + \ldots + \beta_{1k}x_{ik} + \varepsilon_i & ;\text{ if } \upsilon_i \le \upsilon, \\ \alpha_2 + \beta_{21}x_{i1} + \beta_{22}x_{i2} + \ldots + \beta_{2k}x_{ik} + \varepsilon_i & ;\text{ if } \upsilon_i > \upsilon, \end{cases} \tag{2.14}$$

where Y_i^* is a dependent variable, x_{ij} represents the ith observation of the jth independent variable, for $j=1, \ldots, k$ and $i=1, \ldots, n$. In addition, $\upsilon_i = \underset{\sim}{x}_i \underset{\sim}{\vartheta}$ (Quandt, 1972: 307), where $\underset{\sim}{x}_i$ is the row vector with k variables of the i^{th} observation and $\underset{\sim}{\vartheta}$ is a k-dim vector of unknown parameters. The errors ε_i 's are $N\left(0,\sigma_i^2\right)$. Suppose $\underset{\sim}{x}_0=\left(x_{01},\ldots,x_{0k}\right)$ is a vector of regressors at a joined point, i.e., $\underset{\sim}{x}_0\underset{\sim}{\vartheta}=\upsilon$, then, from model (2.5), $\alpha_1+\sum_{j=1}^{k}\beta_{1j}x_{0j}=\alpha_2+\sum_{j=1}^{k}\beta_{2j}x_{0j}$, i.e., $\alpha_2=\alpha_1-\sum_{j=1}^{k}\left(\beta_{2j}-\beta_{1j}\right)x_{0j}$.

For $\upsilon_i>\upsilon$,

$$Y_i^*=\alpha_1+\sum_{j=1}^{k}\beta_{1j}x_{ij}+\sum_{j=1}^{k}\left(\beta_{2j}-\beta_{1j}\right)x_{ij}-\sum_{j=1}^{k}\left(\beta_{2j}-\beta_{1j}\right)x_{0j}+\varepsilon_i .$$

By using a dummy variable, $D_i=\begin{cases}1 & ;\upsilon_i>\upsilon,\\ 0 & ;\upsilon_i\le\upsilon,\end{cases}$, the model (2.5) can be written in a single equation as (Mekbunditkul, 2010)

$$Y_i^*=\alpha_1+\sum_{j=1}^{k}\beta_{1j}x_{ij}+\sum_{j=1}^{k}\beta_{2j}^*x_{ij}D_i+\beta_3^*D_i+\varepsilon_i , \qquad (2.15)$$

where $\beta_2^*=\beta_{2j}-\beta_{1j}$ and $\beta_3^*=-\sum_{j=1}^{k}\left(\beta_{2j}-\beta_{1j}\right)x_{0j}$.

Thus the TP regression model can be written as

$$Y_i=\begin{cases}L_i & ; \quad Y_i^*\le L_i\\ Y_i^* & ; \quad L_i<Y_i^*<U_i\\ U_i & ; \quad Y_i^*\ge U_i ,\end{cases}$$

where $Y_i^*=\alpha_1+\sum_{j=1}^{k}\beta_{1j}x_{ij}+\sum_{j=1}^{k}\beta_{2j}^*x_{ij}D_i+\beta_3^*D_i+\varepsilon_i$. In addition, this model can be written in the matrix form as

$$\underset{\sim}{Y}=X\underset{\sim}{\theta}+\underset{\sim}{\varepsilon}, \qquad (2.16)$$

where

$$\underset{\sim}{Y} = \left[\underset{\sim}{L}_1' \;\vdots\; \underset{\sim}{Y}_2' \;\vdots\; \underset{\sim}{U}_3'\right]_{n\times 1}',$$

$$= \left[L_{11} \quad L_{21} \quad \cdots \quad L_{n_1 1} \;\vdots\; Y_{12} \quad Y_{22} \quad \cdots \quad Y_{n_2 2} \;\vdots\; U_{13} \quad U_{23} \quad \cdots \quad U_{n_3 3}\right]',$$

$\underset{\sim}{Y}_2 = \underset{\sim}{Y}_2^*$ and $\underset{\sim}{\theta} = \left(\alpha_1, \beta_{11}, \ldots, \beta_{1k}, \beta_{21}^*, \ldots, \beta_{2k}^*, \beta_3^*\right)'$ and X is defined as in the equation (2.9). Moreover, the limits L and U are defined by

$$L_{im} = \begin{cases} L_a & ; \upsilon_{im} \le \upsilon, \\ L_b & ; \upsilon_{im} > \upsilon, \end{cases} \quad \text{and} \quad U_{im} = \begin{cases} U_a & ; \upsilon_{im} \le \upsilon, \\ U_b & ; \upsilon_{im} > \upsilon. \end{cases} \tag{2.17}$$

The vector $\underset{\sim}{Y}^*$ can be described as below: Without loss of generality (WLOG), all of the observed data Y_i^*'s as well as $x_{i1}, \ldots, x_{ik}$ to which Y_i^* corresponds, for $i=1, \ldots, n$ are rearranged. Hence, observation vector $\underset{\sim}{Y}^*$ consists of three parts. One of them is observation with values smaller than the lower limit L. The second comprises of values that lie between the limit (L,U) and the third indicates values greater than the upper limit U. To be specific, suppose that the observation in each part are n_1, n_2 and n_3, respectively. Thus,

$$\underset{\sim}{Y}^* = \left[\underset{\sim}{Y}_1^{*\prime} \;\vdots\; \underset{\sim}{Y}_2^{*\prime} \;\vdots\; \underset{\sim}{Y}_3^{*\prime}\right]_{n\times 1}',$$

$$= \left[Y_{11}^* \quad Y_{21}^* \quad \cdots \quad Y_{n_1 1}^* \;\vdots\; Y_{12}^* \quad Y_{22}^* \quad \cdots \quad Y_{n_2 2}^* \;\vdots\; Y_{13}^* \quad Y_{23}^* \quad \cdots \quad Y_{n_3 3}^*\right]'.$$

The variance-covariance matrix of $\underset{\sim}{Y}^*$ is assumed to be

$$\Sigma = \left[\begin{array}{c:c:c} \Sigma_1 & 0 & 0 \\ \hdashline 0 & \Sigma_2 & 0 \\ \hdashline 0 & 0 & \Sigma_3 \end{array}\right], \qquad \text{where} \qquad \Sigma_m = \operatorname{diag}\left(\sigma_{1m} \ldots \sigma_{n_m m}\right) \quad \text{and}$$

$$\sigma_{im}^2 = \begin{cases} \sigma_a^2 & \text{if } \upsilon_{im} \le \upsilon, \\ \sigma_b^2 & \text{if } \upsilon_{im} > \upsilon \end{cases}, \; m = 1, 2, 3; \; i = 1, \ldots, n_m .$$

Therefore, the inverse matrix of Σ is $\Sigma^{-1} = \left[\begin{array}{c:c:c} \Sigma_1^{-1} & 0 & 0 \\ \hdashline 0 & \Sigma_2^{-1} & 0 \\ \hdashline 0 & 0 & \Sigma_3^{-1} \end{array}\right]$.

Since Σ_m are diagonal matrices, so $\Sigma_m^{-1} = \operatorname{diag}\left(\frac{1}{\sigma_{1m}}, \ldots, \frac{1}{\sigma_{n_m m}}\right)$, where m = 1, 2, 3.

In addition, it is assumed that $\varepsilon_{im} \sim N\left(0, \sigma_{im}^2\right)$.

The matrix of X independent variables corresponding to

$\underset{\sim}{Y} = \left[\underset{\sim}{L}_1' \,\vdots\, \underset{\sim}{Y}_2' \,\vdots\, \underset{\sim}{U}_3'\right]'_{n\times 1}$, is

$$X = \left[\begin{array}{c} X_1 \\ \hdashline X_2 \\ \hdashline X_3 \end{array}\right] = \left[\begin{array}{cccccccc} 1 & x_{111} & \cdots & x_{1k1} & x_{111}^* & \cdots & x_{1k1}^* & x_{11}' \\ 1 & x_{211} & \cdots & x_{2k1} & x_{211}^* & \cdots & x_{2k1}^* & x_{21}' \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n_1 11} & \cdots & x_{n_1 k1} & x_{n_1 11}^* & \cdots & x_{n_1 k1}^* & x_{n_1 1}' \\ \hdashline 1 & x_{112} & \cdots & x_{1k2} & x_{112}^* & \cdots & x_{1k2}^* & x_{11}' \\ 1 & x_{212} & \cdots & x_{2k2} & x_{212}^* & \cdots & x_{2k2}^* & x_{22}' \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n_2 12} & \cdots & x_{n_2 k2} & x_{n_2 12}^* & \cdots & x_{n_2 k2}^* & x_{n_2 2}' \\ \hdashline 1 & x_{113} & \cdots & x_{1k3} & x_{113}^* & \cdots & x_{1k3}^* & x_{13}' \\ 1 & x_{213} & \cdots & x_{2k3} & x_{213}^* & \cdots & x_{2k3}^* & x_{23}' \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n_3 13} & \cdots & x_{n_3 k3} & x_{n_3 13}^* & \cdots & x_{n_3 k3}^* & x_{n_3 3}' \end{array}\right]_{n\times(2k+2)} \tag{2.9}$$

where $x^{*}_{ijm} = x_{ijm}D_{im}$, $x'_{im} = D_{im}$ and $D_{im} = \begin{cases} 1 & ;\upsilon_{im} > \upsilon, \\ 0 & ;\upsilon_{im} \le \upsilon, \end{cases}$ where m=1, 2, 3; j=1, …, k; i=1, …, n_m, and k is the number of regressor variables. Note that 2k+2 is less than n.

2.7 TP Estimator

The log-likelihood function of $\underset{\sim}{\theta}$ in TP regression model was derived by Mekbunditkul (2011) via the p.d.f. of Y_{im}, the im^{th} element of vector $\underset{\sim}{Y}$. The p.d.f. of ε_{im} is assumed to be normal with zero mean and σ^2_{im} variance and it is independent from each other. When given the values of L_{im} and U_{im}, for $i = 1, \ldots, n_m$ and

m = 1, 2 and 3, the p.d.f. of $\underset{\sim}{Y}$ was derived into three parts as the followings.

Part 1. For $Y_{i1} = L_{i1}$, where $i = 1, \ldots, n_1$:

$$\begin{aligned} f_{Y_{i1}}(L_{i1}) &= P\left(Y_{i1} = L_{i1}\right), \\ &= P\left(Y^{*}_{i1} \le L_{i1}\right), \\ &= P\left(\underset{\sim}{x}_{i1}\underset{\sim}{\theta} + \varepsilon_{i1} \le L_{i1}\right), \\ &= P\left(\frac{\varepsilon_{i1}}{\sigma_{i1}} \le \frac{L_{i1} - \underset{\sim}{x}_{i1}\underset{\sim}{\theta}}{\sigma_{i1}}\right), \\ &= \Phi\left(\frac{L_{i1} - \underset{\sim}{x}_{i1}\underset{\sim}{\theta}}{\sigma_{i1}}\right). \end{aligned}$$

Part 2. For $L_{i2} < Y_{i2} < U_{i2}$, where $i = 1, \ldots, n_2$:

$$P\left(L_{i2} < Y_{i2} < U_{i2}\right) = P\left(L_{i2} < \underset{\sim}{x}_{i2}\underset{\sim}{\theta} + \varepsilon_{i2} \le y^{*}_{i2}\right),$$

$$= P\left(\frac{L_{i2} - \underset{\sim}{x}_{i2}\underset{\sim}{\theta}}{\sigma_{i2}} < \frac{\varepsilon_{i2}}{\sigma_{i2}} \le \frac{y_{i2}^{*} - \underset{\sim}{x}_{i2}\underset{\sim}{\theta}}{\sigma_{i2}}\right),$$

$$= \Phi\left(\frac{Y_{i2}^{*} - \underset{\sim}{x}_{i2}\underset{\sim}{\theta}}{\sigma_{i2}}\right) - \Phi\left(\frac{L_{i2} - \underset{\sim}{x}_{i2}\underset{\sim}{\theta}}{\sigma_{i2}}\right).$$

Hence, $f_{Y_{i2}}\left(y_{i2}\right) = \frac{1}{\sigma_{i2}}\phi\left(\frac{y_{i2}^{*} - \underset{\sim}{x}_{i2}\underset{\sim}{\theta}}{\sigma_{i2}}\right) = \frac{1}{\sigma_{i2}}\phi\left(\frac{y_{i2} - \underset{\sim}{x}_{i2}\underset{\sim}{\theta}}{\sigma_{i2}}\right)$, where $y_{i2} \in \left(L_{i2}, U_{i2}\right)$.

Part 3. For $Y_{i3} = U_{i3}$, where $i = 1, \ ..., \ n_3$:

$$P\left(Y_{i3} = U_{i3}\right) = P\left(\underset{\sim}{x}_{i3}\underset{\sim}{\theta} + \varepsilon_{i3} \ge U_{i3}\right),$$

$$= 1 - P\left(\frac{\varepsilon_{i3}}{\sigma_{i3}} < \frac{U_{i3} - \underset{\sim}{x}_{i3}\underset{\sim}{\theta}}{\sigma_{i3}}\right),$$

$$= 1 - \Phi\left(\frac{U_{i3} - \underset{\sim}{x}_{i3}\underset{\sim}{\theta}}{\sigma_{i3}}\right).$$

Functions Φ and ϕ are the c.d.f. and p.d.f. of a standard normal distribution, respectively.

Some notations were indicated to be used in the next part as

$I_L = \left\{ i \mid Y_{i1} = L_{i1}, \ i=1, \ ..., \ n_1 \right\}$,

$I_Y = \left\{ i \mid L_{i2} < Y_{i2} < U_{i2}, \ i = 1, \ ..., \ n_2 \right\}$, and

$I_U = \left\{ i \mid Y_{i3} = U_{i3}, \ i = 1, \ ..., \ n_3 \right\}$.

From the independent property of each element in the vector $\underset{\sim}{Y}$, the p.d.f. of $\underset{\sim}{Y}$ can be expressed as

$$f_{\underset{\sim}{Y}}\left(\underset{\sim}{y}\right)=\prod_{i\in I_L}\Phi\left(\frac{L_{i1}-\underset{\sim}{x}_{i1}\underset{\sim}{\theta}}{\sigma_{i1}}\right)\cdot\left[\frac{\exp\left\{-\frac{1}{2}\left(\underset{\sim}{Y}_2-X_2\underset{\sim}{\theta}\right)'\Sigma_2^{-1}\left(\underset{\sim}{Y}_2-X_2\underset{\sim}{\theta}\right)\right\}}{(2\pi)^{n_2/2}\left|\Sigma_2\right|^{n_2/2}}\right]$$

$$\cdot\prod_{i\in I_U}1-\Phi\left(\frac{U_{i3}-\underset{\sim}{x}_{i3}\underset{\sim}{\theta}}{\sigma_{i3}}\right). \tag{2.19}$$

Thus, the log-likelihood function is thus given by

$$\ln L\left(\underset{\sim}{\theta};\underset{\sim}{Y}\right)=\sum_{i\in I_L}\ln\Phi\left(\lambda_{i1}^{-}\right)-\frac{n_2}{2}\ln 2\pi-\frac{1}{2}\ln\left|\Sigma_2\right|-\frac{1}{2}\left(\underset{\sim}{Y}_2-X_2\underset{\sim}{\theta}\right)'\Sigma_2^{-1}\left(\underset{\sim}{Y}_2-X_2\underset{\sim}{\theta}\right)$$

$$+\sum_{i\in I_U}\ln\left[1-\Phi\left(\lambda_{i3}\right)\right]. \tag{2.20}$$

where $\lambda_{i1}^{-}=\dfrac{L_{i1}-\underset{\sim}{x}_{i1}\underset{\sim}{\theta}}{\sigma_{i1}}$ and $\lambda_{i3}=\dfrac{U_{i3}-\underset{\sim}{x}_{i3}\underset{\sim}{\theta}}{\sigma_{i3}}$.

The ML estimators of $\underset{\sim}{\theta}$ can be obtained straightforwardly from the log-likelihood equation (2.20), which consists of three parts, as

$$\frac{\partial\ln L\left(\underset{\sim}{\theta};\underset{\sim}{y}\right)}{\partial\underset{\sim}{\theta}}=-\sum_{i\in I_L}\left(\frac{\phi\left(\hat{\lambda}_{i1}^{-}\right)}{\Phi\left(\hat{\lambda}_{i1}^{-}\right)}\right)\frac{\underset{\sim}{x}_{i1}'}{\sigma_{i1}}-\left(X_2'\Sigma_2^{-1}X_2\right)\hat{\underset{\sim}{\theta}}_{TP}+\left(X_2'\Sigma_2^{-1}\underset{\sim}{Y}_2\right)$$

$$+\sum_{i\in I_U}\left(\frac{\phi\left(\hat{\lambda}_{i3}\right)}{1-\Phi\left(\hat{\lambda}_{i3}\right)}\right)\frac{\underset{\sim}{x}_{i3}'}{\sigma_{i3}}=\underset{\sim}{0}_{(2k+1)\times 1}.$$

From Mekbunditkul (2010), $\hat{\underset{\sim}{\theta}}_{TP}$ was verified as in the form of

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

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

where $$H_1\left(\hat{\bar{\lambda}}\right)=\left(h\left(\hat{\bar{\lambda}}_{11}\right) \ \cdots \ h\left(\hat{\bar{\lambda}}_{n_1 1}\right)\right)'=\left(\frac{\phi\left(\hat{\bar{\lambda}}_{11}\right)}{\Phi\left(\hat{\bar{\lambda}}_{11}\right)} \ \cdots \ \frac{\phi\left(\hat{\bar{\lambda}}_{n_1 1}\right)}{\Phi\left(\hat{\bar{\lambda}}_{n_1 1}\right)}\right)', \tag{2.21}$$

and $$\underset{\sim}{H}_3\left(\hat{\lambda}\right)=\left(h\left(\hat{\lambda}_{13}\right) \ \cdots \ h\left(\hat{\lambda}_{n_3 3}\right)\right)'=\left(\frac{\phi\left(\hat{\lambda}_{13}\right)}{1-\Phi\left(\hat{\lambda}_{13}\right)} \ \cdots \ \frac{\phi\left(\hat{\lambda}_{n_3 3}\right)}{1-\Phi\left(\hat{\lambda}_{n_3 3}\right)}\right)', \tag{2.22}$$

where $\hat{\bar{\lambda}}_{n_1 1}$ and $\hat{\lambda}_{i3}$ are estimators of $\bar{\lambda}_{n_1 1}$ and λ_{i3}, respectively.

There exist three parts of TP estimator which the first part is the LS estimator based on n_2 observations where values are not at the limits. The other two parts concern n_1 and n_3 observations whose values are truncated respectively by the lower and upper desired limits.

2.8 Properties of TP Estimator

In Mekbunditkul's dissertation, there were verified some properties of TP estimator in terms of its bias and mean square error (MSE). Two situations were defined: (1) $U \to \infty$ and L is finite, (2) $L \to -\infty$ and U is finite. Some vectors/matrices such as $\underset{\sim}{Y}$, X, $\underset{\sim}{L}$ and $\underset{\sim}{U}$ and Σ were defined as in section (2.4) and the following statements were indicated to refer throughout this section

C1: Assuming that $U \to \infty$ and L is finite, the response variable $\underset{\sim}{Y}$ in TP regression model is as $\underset{\sim}{Y} = \begin{bmatrix} \underset{\sim}{L}_1 \\ --- \\ \underset{\sim}{Y}_2 \end{bmatrix}_{n\times 1}$, where $n = n_1 + n_2$.

C2: Assuming that $L \rightarrow -\infty$ and U is finite, the response variable $\underset{\sim}{Y}$ in TP regression model is as $\underset{\sim}{Y} = \begin{bmatrix} \underset{\sim}{Y}_2 \\ --- \\ \underset{\sim}{U}_3 \end{bmatrix}_{n\times 1}$, where $n = n_2 + n_3$.

The TP estimator, $\hat{\underset{\sim}{\theta}}_{TP}$ corresponding respectively to each of C1 and C2 is

$$\hat{\underset{\sim}{\theta}}_{TP} = \left(X_2'\Sigma_2^{-1}X_2\right)^{-1}\left(X_2'\Sigma_2^{-1}\underset{\sim}{Y}_2\right) - \left(X_2'\Sigma_2^{-1}X_2\right)^{-1}\left[X_1'\Sigma_1^{-1/2}\left\{\underset{\sim}{H}_1\left(\hat{\lambda}\right)\right\}\right], \tag{2.23}$$

$$\hat{\underset{\sim}{\theta}}_{TP} = \left(X_2'\Sigma_2^{-1}X_2\right)^{-1}\left(X_2'\Sigma_2^{-1}\underset{\sim}{Y}_2\right) + \left(X_2'\Sigma_2^{-1}X_2\right)^{-1}\left\{X_3'\Sigma_3^{-1/2}\left\{\underset{\sim}{H}_3\left(\hat{\lambda}\right)\right\}\right\}, \tag{2.24}$$

where $\underset{\sim}{H}_1\left(\hat{\lambda}\right)$ and $\underset{\sim}{H}_3\left(\hat{\lambda}\right)$ are in the forms of equations (2.23) and (2.24).

Theorem 2. Assume C1**,** $\hat{\underset{\sim}{\theta}}_{TP}$, as defined in equation (2.23), is biased where the bounds of the bias are (Mekbunditkul, 2010)

$$\text{Bias}_{\ell} = -A_1\left(I + A_1\right)^{-1}\underset{\sim}{\theta} - \left\{I - A_1\left(I + A_1\right)^{-1}\right\}\left(X_2'\Sigma_2^{-1}X_2\right)^{-1}X_1'\Sigma_1^{-1/2}\left(\underset{\sim}{1} - \Sigma_1^{-1/2}\underset{\sim}{L}_1\right),$$

$$\text{Bias}_{u} = \left\{I - A_1\left(I + A_1\right)^{-1}\right\}\underset{\sim}{\theta} - \left\{I - A_1\left(I + A_1\right)^{-1}\right\}\left(X_2'\Sigma_2^{-1}X_2\right)^{-1} \cdot\left(X_2'\Sigma_2^{-1}\underset{\sim}{L}_2 - X_1'\Sigma_1^{-1}\underset{\sim}{L}_1 + X_2'\Sigma_2^{-1/2}\underset{\sim}{1}\right).$$

Proof. It is available in Mekbunditkul's dissertation.

Theorem 3. Assume C1 holds. The asymptotic variance-covariance matrix of $\hat{\underset{\sim}{\theta}}_{TP}$, as defined in (2.23), is (Mekbunditkul, 2010)

$\left(X'WX\right)^{-1} = \left\{\left(X_1'W_1X_1\right) + \left(X_2'W_2X_2\right)\right\}^{-1}$, where $W_1 = G_1\left(\lambda^{-}\right)\Sigma_1^{-1}$ and $W_2 = \Sigma_2^{-1}$.

Proof. It is available in Mekbunditkul's dissertation.

Theorem 4. If C2 holds, the estimator $\hat{\underset{\sim}{\theta}}_{TP}$, as defined in equation (2.24), is biased where the bounds of the bias are (Mekbunditkul, 2010)

$$\text{Bias}_{\ell} = \left\{ I - A_2 (I + A_2)^{-1} \right\} \underset{\sim}{\theta} - \left\{ I - A_2 (I + A_2)^{-1} \right\}$$
$$\cdot \left\{ \left(X_2' \Sigma_2^{-1} X_2 \right)^{-1} \left(X_3' \Sigma_3^{-1} \underset{\sim}{U}_3 - X_2' \Sigma_2^{-1} \underset{\sim}{U}_2 - X_2' \Sigma_2^{-1} \underset{\sim}{1} \right) \right\},$$

$$\text{Bias}_{u} = \left\{ I - A_2 (I + A_2)^{-1} \right\} \left[\left(X_2' \Sigma_2^{-1} X_2 \right)^{-1} \left\{ X_3' \Sigma_3^{-1/2} \left(\underset{\sim}{U}_3 + \underset{\sim}{1} \right) \right\} \right] - A_2 (I + A_2)^{-1} \underset{\sim}{\theta}.$$

Proof. It is available in Mekbunditkul's dissertation.

Theorem 5. Assume C2 holds. The asymptotic variance-covariance matrix of $\hat{\underset{\sim}{\theta}}_{TP}$, as defined in (2.24), is (Mekbunditkul, 2010)

$\left(X' W X \right)^{-1} = \left\{ \left(X_3' W_3 X_3 \right) + \left(X_2' W_2 X_2 \right) \right\}^{-1}$, where $W_3 = G_3(\lambda) \Sigma_3^{-1}$, and $W_2 = \Sigma_2^{-1}$ (Mekbunditkul, 2010).

Proof. It is available in Mekbunditkul's dissertation.

Theorem 6. Let C1 hold, then the ML estimators of each σ_a^2 and σ_b^2 in a TP regression model are obtained using (Mekbunditkul, 2010)

$$\max\left(c_j^2, f_j \right) < \hat{\sigma}_j^2 < \left(d_j + \sqrt{\frac{1}{2} d_j^2 + 2 f_j} \right)^2, \quad \text{where } j = a, b,$$

$c_j = 0$ if $d_j - \sqrt{\frac{1}{2} d_j^2 + 2 f_j} < 0$ and $c_j = d_j - \sqrt{\frac{1}{2} d_j^2 + 2 f_j}$ otherwise,

$$d_j = \frac{\underset{\sim}{1}' \left(\hat{\underset{\sim}{Y}}_{j1} - \underset{\sim}{L}_{j1} \right)}{n_{j2}}, \text{ and } f_j = \frac{\left(\hat{\underset{\sim}{Y}}_{j1} - \underset{\sim}{L}_{j1} \right)' \left(\hat{\underset{\sim}{Y}}_{j1} - \underset{\sim}{L}_{j1} \right)}{n_{j2}} + \frac{\left(\underset{\sim}{Y}_{j2} - \hat{\underset{\sim}{Y}}_{j2} \right)' \left(\underset{\sim}{Y}_{j2} - \hat{\underset{\sim}{Y}}_{j2} \right)}{n_{j2}}.$$

Proof. It is available in Mekbunditkul's dissertation.

Theorem 7. Let C2 hold, then ML estimators of each σ_a^2 and σ_b^2 in the TP regression model are obtained by (Mekbunditkul, 2010)

$$\max\left(p_j^2, r_j\right) < \hat{\sigma}_j^2 < \left(q_j + \sqrt{\frac{1}{2}q_j^2 + 2r_j}\right)^2, \quad \text{where} \quad j = a, b,$$

$$p_j = 0 \text{ if } q_j - \sqrt{\frac{1}{2}q_j^2 + 2r_j} < 0 \text{ and } p_j = q_j - \sqrt{\frac{1}{2}q_j^2 + 2r_j} \text{ otherwise,}$$

$$q_j = \frac{\underset{\sim}{1}'\left(\underset{\sim}{U}_{j3} - \hat{\underset{\sim}{Y}}_{j3}\right)}{n_{j2}} \text{ and } r_j = \frac{\left(\underset{\sim}{U}_{j3} - \hat{\underset{\sim}{Y}}_{j3}\right)'\left(\underset{\sim}{U}_{j3} - \hat{\underset{\sim}{Y}}_{j3}\right)}{n_{j2}} + \frac{\left(\underset{\sim}{Y}_{j2} - \hat{\underset{\sim}{Y}}_{j2}\right)'\left(\underset{\sim}{Y}_{j2} - \hat{\underset{\sim}{Y}}_{j2}\right)}{n_{j2}}.$$

Proof. It is available in Mekbunditkul's dissertation.

Whenever the same sample sizes n_1 and n_2 are assumed then an estimate of variance σ_{TP}^2 is as their pooled average of σ_a^2 and σ_b^2 (Snedecor and Cochran, 1989 and Welch, 1947). That is the ML estimator of σ_{TP}^2 is $\hat{\sigma}_{TP}^2 = \frac{\hat{\sigma}_a^2 + \hat{\sigma}_b^2}{2}$, where $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$ as shown in Theorems 6 and 7.

CHAPTER 3

A JOINED POINT ESTIMATION IN TOBIT-PIECEWISE REGRESSION MODEL

The TP regression model which was first introduced by Mekbunditkul, the joined point was assumed to be fixed. It is quite simple to fix the joined point in TP regression model if it is known in advance where it is. Thus this research deals with the more difficult case where the joined point has to be estimated from the data. Two estimation methods are introduced to investigate the joined point in TP regression model in this research as described below.

3.1 The Maximum Likelihood Fashion

In this section, the particular case that a single one regressor is assumed to simplify. The combination of the simple Tobit (2.1) and simple piecewise (2.4) regression models to be the TP regression model is shown in the model (3.1):

$$Y_i = \begin{cases} L_i & ; \quad Y_i^* \le L_i \\ Y_i^* & ; \quad L_i < Y_i^* < U_i \\ U_i & ; \quad Y_i^* \ge U_i, \end{cases} \tag{3.1}$$

where $Y_i^* = \alpha_1 + \beta_1 x_i + \beta_2 x_i^* + \varepsilon_i$, the regressor variables are x_i and x_i^*, $x_i^* = \left(x_i - \upsilon\right) D_i$, υ is an unknown joined point of two regression lines, and ε_i 's is i.d. $N\left(0, \sigma_i^2\right)$. Note $\sigma_i^2 = \begin{cases} \sigma_a^2 & \text{if } x_i \le \upsilon \\ \sigma_b^2 & \text{if } x_i > \upsilon \end{cases}$. The locally lower and upper limits are $L_i = \begin{cases} L_a & ; \ x_i \le \upsilon \\ L_b & ; \ x_i > \upsilon \end{cases}$, and $U_i = \begin{cases} U_a & ; \ x_i \le \upsilon \\ U_b & ; \ x_i > \upsilon \end{cases}$. The probability density function (p.d.f.) of Y is determined by

$$f_Y(y_i) = \Phi\left(\frac{L_i - \alpha_1 - \beta_1 x_i - \beta_2 x_i^*}{\sigma_i}\right) \text{ if } y_i = L_i,$$

$$f_Y(y_i) = \frac{1}{\sigma_i}\phi\left(\frac{y_i - \alpha_1 - \beta_1 x_i - \beta_2 x_i^*}{\sigma_i}\right) \text{ if } L_i < y_i < U_i,$$

$$\text{and } f_Y(y_i) = 1 - \Phi\left(\frac{U_i - \alpha_1 - \beta_1 x_i - \beta_2 x_i^*}{\sigma_i}\right) \text{ if } y_i = U_i.$$

Some notations are indicated for the derivation of joined point estimator as follows:

$I_{aL} = \{i \mid Y_{i1} = L_{i1} \text{ and } \upsilon_{i1} \le \upsilon,\ i = 1,\ ...,\ n_{a1}\}$,

$I_{bL} = \{i \mid Y_{i1} = L_{i1} \text{ and } \upsilon_{i1} > \upsilon,\ i = 1,\ ...,\ n_{b1}\}$,

$I_{aY} = \{i \mid Y_{i2} > L_{i2} \text{ and } \upsilon_{i2} \le \upsilon,\ i = 1,\ ...,\ n_{a2}\}$,

$I_{bY} = \{i \mid Y_{i2} > L_{i2} \text{ and } \upsilon_{i2} > \upsilon,\ i = 1,\ ...,\ n_{b2}\}$,

$I_{aU} = \{i \mid Y_{i3} > U_{i3} \text{ and } \upsilon_{i3} \le \upsilon,\ i = 1,\ ...,\ n_{a3}\}$,

and $I_{bU} = \{i \mid Y_{i3} > L_{i3} \text{ and } \upsilon_{i3} > \upsilon,\ i = 1,\ ...,\ n_{b3}\}$.

Note that $n_1 = n_{a1} + n_{b1}$, $n_2 = n_{a2} + n_{b2}$ and $n_3 = n_{a3} + n_{b3}$. In addition, n = $\sum_{j=1}^{3} n_j$.

The TP estimator of $\underset{\sim}{\theta}$ can be achieved by the ML method when the log-likelihood function of $\underset{\sim}{\theta} = (\alpha_1, \beta_1, \beta_2)'$ given $\underset{\sim}{Y}$ for some fixed values of $L_a, L_b, U_a, U_b,$ and σ^2 known, can be written as

$$\begin{aligned} \ln L\left(\upsilon; \hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2, \underset{\sim}{Y}\right) = & \sum_{i \in I_L} \left\{ \ln \Phi\left(\frac{L_i - \hat{\alpha}_1 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^*}{\hat{\sigma}_i}\right)\right\} \\ & + \sum_{i \in I_Y} \left\{ \ln\left(\frac{1}{\hat{\sigma}_i}\phi\left(\frac{y_i - \hat{\alpha}_1 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^*}{\hat{\sigma}_i}\right)\right)\right\} \\ & + \sum_{i \in I_U} \left\{ \ln\left(1 - \Phi\left(\frac{U_i - \hat{\alpha}_1 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^*}{\hat{\sigma}_i}\right)\right)\right\}. \end{aligned} \tag{3.2}$$

$$\frac{\partial \ln L\left(\upsilon;\hat{\alpha}_1,\hat{\beta}_1,\hat{\beta}_2,\underset{\sim}{Y}\right)}{\partial \upsilon}=\sum_{i\in I_L}\left\{\frac{\hat{\beta}_2 D_i \phi\left(\dfrac{L_i-\hat{\alpha}_1-\hat{\beta}_1 x_i-\hat{\beta}_2 x_i^*}{\hat{\sigma}_i}\right)}{\sigma_i\,\Phi\left(\dfrac{L_i-\hat{\alpha}_1-\hat{\beta}_1 x_i-\hat{\beta}_2 x_i^*}{\hat{\sigma}_i}\right)}\right\}$$

$$+\sum_{i\in I_Y}^{n_2}\left\{\frac{\hat{\beta}_2 D_i\left(y_i-\hat{\alpha}_1-\hat{\beta}_1 x_i-\hat{\beta}_2 x_i^*\right)}{\hat{\sigma}_i^2}\right\} \qquad (3.3)$$

$$+\sum_{i\in I_U}\left\{\frac{\hat{\beta}_2 D_i \phi\left(\dfrac{U_i-\hat{\alpha}_1-\hat{\beta}_1 x_i-\hat{\beta}_2 x_i^*}{\hat{\sigma}_i}\right)}{\hat{\sigma}_i\left(1-\Phi\left(\dfrac{U_i-\hat{\alpha}_1-\hat{\beta}_1 x_i-\hat{\beta}_2 x_i^*}{\hat{\sigma}_i}\right)\right)}\right\}$$

Accordingly, the score statistic for υ, the function (3.3), is always positive and it proves to be inappropriate the traditional way to find the value of υ which maximizes $\ln L\left(\upsilon;\hat{\alpha}_1,\hat{\beta}_1,\hat{\beta}_2,\underset{\sim}{Y}\right)$ by differentiating $\ln L\left(\upsilon;\hat{\alpha}_1,\hat{\beta}_1,\hat{\beta}_2,\underset{\sim}{Y}\right)$ with respect to υ and setting the derivation equal to zero. Quandt (1958) suggested a procedure to calculate the value of a switching point (a special case of joined point) by selecting t which gives the maximum likelihood function, where t is the time period. Nevertheless, the assumption of Quandt is without one joined point. By then, the estimate of a switching point was just introduced in piecewise regression model. Subsequently, Hudson (1966) suggested a parameter estimate based on the LS method and the joined point is assumed. Thus, we can apply the procedure of Quandt by assuming the joined point to find the value of υ in TP regression model. This procedure can be expressed as followed:

First, order the observation according to the value of x_i and split the data into two groups, i.e. left hand group and right hand group.

Second, determine the initial value of υ with υ as being in the range of X and put υ in the model (3.1).

Third, estimate remaining parameters in the model (3.1) by $\hat{\underset{\sim}{\theta}}_{TP}$ as shown in the equation (2.14) or (2.15).

Forth, substitute $\hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2$ back to the log-likelihood function (3.2) and calculate its value, after that move the point of υ between the two groups by one unit at a time to the right and one unit at a time to the left.

Fifth, calculate the log-likelihood function for each value of υ and then choose the value of υ which maximizes the log-likelihood function. Then, the ML estimators $\hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2$ are obtained.

In sum, this way can be generalized to the case that multiple linear regressions are taken into consideration.

3.2 The Nonlinear Least Square Fashion

The Tobit-piecewise regression model can be considered as one of nonlinear regression models evident in the figure 1.2 thus in the case that the data should truly be fitted by nonlinear regression models rather than linear models some nonlinear least square solving based have been recommended. A nonlinear regression model (Seber and Wild, 1988: 21) can be written as

$$Y_i = f\left(\underset{\sim}{x}_i ; \underset{\sim}{\theta}^*\right) + \varepsilon_i , \tag{3.4}$$

where i=1, 2,…, n, $f\left(\underset{\sim}{x}_i ; \underset{\sim}{\theta}^*\right)$ is a known regression function as defined in the equation (3.1), $\underset{\sim}{x}_i$ is a k x 1 vector, $\underset{\sim}{\theta}^*$ is a vector of k unknown parameters and the $E\left(\varepsilon_i\right) = 0$. The true value $\underset{\sim}{\theta}^*$ of $\underset{\sim}{\theta}$ is known to belong to Θ, a subset of p-dim Euclidian space $\Re^p$. From these statements, we can state that the ith element, Y_i^*, of $\underset{\sim}{Y}^*$ as shown in the model (2.15) can be served as the model (3.4). The least square estimate of $\underset{\sim}{\theta}^*$, denoted by $\hat{\underset{\sim}{\theta}}$, minimizes the error sum of squares. Thus, we can state the definition of nonlinear least square (NLS) estimator by the following definition.

Definition 1. The nonlinear least square (NLS) estimator for the nonlinear regression model (3.4) is defined by

$$
\begin{aligned}
\hat{\underset{\sim}{\theta}}^{NLS} &= \underset{\underset{\sim}{\theta}^{*} \in \Re^{p}}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\left(\underset{\sim}{\theta}^{*}\right) \right)^2 \\
&= \underset{\underset{\sim}{\theta}^{*} \in \Re^{p}}{\arg\min} \sum_{i=1}^{n} \left(Y_i - f\left(\underset{\sim}{x}_i ; \underset{\sim}{\theta}^{*}\right) \right)^2 .
\end{aligned}
$$

Unlike the linear least square estimator, the analytical solution of this solving for a general function $f\left(\underset{\sim}{x}_i ; \underset{\sim}{\theta}^{*}\right)$ can not be expressed. The Taylor's series expansion has been recommended to approximate the nonlinear objective function because the first two derivatives of $f\left(\underset{\sim}{x}_i ; \underset{\sim}{\theta}^{*}\right)$ exist. Let

$$
S\left(\underset{\sim}{\theta}\right) = \sum_{i=1}^{n} \left(Y_i - f\left(\underset{\sim}{x}_i ; \underset{\sim}{\theta}\right) \right)^2 , \tag{3.5}
$$

whenever each $f\left(\underset{\sim}{x}_i ; \underset{\sim}{\theta}\right)$ is differentiable with respect to $\underset{\sim}{\theta}$, $\hat{\underset{\sim}{\theta}}$ is satisfied (Seber and Wild, 1988: 21)

$$
\left. \frac{\partial S\left(\underset{\sim}{\theta}\right)}{\partial \theta_r} \right|_{\underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}} = \underset{\sim}{0}, \quad r = 1, 2, \ldots, p, \tag{3.6}
$$

The $f\left(X ; \underset{\sim}{\theta}\right)$ is defined as $f\left(X ; \underset{\sim}{\theta}\right) = \left(f\left(\underset{\sim}{x}_1 ; \underset{\sim}{\theta}\right), f\left(\underset{\sim}{x}_2 ; \underset{\sim}{\theta}\right), \ldots, f\left(\underset{\sim}{x}_n ; \underset{\sim}{\theta}\right) \right)'$

and

$$
F.\left(\underset{\sim}{x}_i ; \underset{\sim}{\theta}\right) = \frac{\partial f\left(\underset{\sim}{x}_i ; \underset{\sim}{\theta}\right)}{\partial \underset{\sim}{\theta}'} = \left[\left(\frac{\partial f_i\left(\underset{\sim}{x}_i ; \underset{\sim}{\theta}\right)}{\partial \theta_r} \right) \right], \tag{3.7}
$$

where $F.\left(\underset{\sim}{x}_i ; \underset{\sim}{\theta}\right)$ represents the first derivative.

Rewrite the equation (3.5) as

$$S(\underset{\sim}{\theta}) = \left\| \underset{\sim}{Y} - f\left(\underset{\sim}{x}_i; \underset{\sim}{\theta}\right) \right\|^2 . \tag{3.8}$$

The equation (3.6) induces to the following equation

$$\sum_{i=1}^{n} \left(Y_i - f\left(\underset{\sim}{x}_i; \underset{\sim}{\theta}\right) \right) \frac{\partial f\left(\underset{\sim}{x}_i; \underset{\sim}{\theta}\right)}{\partial \theta_r} \Bigg|_{\underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}} = 0 , \;\; r = 1, 2, \ldots, p, \tag{3.9}$$

or

$$\begin{aligned} \underset{\sim}{0} &= \hat{\underset{\sim}{F}}'.\left(\underset{\sim}{Y} - f\left(X; \hat{\underset{\sim}{\theta}} \right) \right) \\ &= \hat{\underset{\sim}{F}}'.\hat{\underset{\sim}{\varepsilon}} \end{aligned} \tag{3.10}$$

This is called the normal equation (Seber and Wild, 1988: 22) for the nonlinear model. The numerical methods, namely Gauss-Newton method, steepest descent method and Levenberg-Marquardt method as described in the section 2.5, are utilized to find the value of $\hat{\underset{\sim}{\theta}}$ because the most nonlinear estimators of nonlinear model can not be solved explicitly. In this research, only Levenberg-Marquardt method is provided in the simulation results.

CHAPTER 4

NUMERICAL ANALYSIS AND SIMULATION STUDIES

4.1 Numerical Analysis

Household income and household expenditure data on socio-economic surveys (SES) in Thailand for the years 2007 and 2009 are analyzed to particularly investigate the performance of four different regressions, namely LS, Tobit, piecewise (abbreviated by PW) and Tobit-piecewise (abbreviated by TP) regressions.

In this research, TP and PW where each of their joined points is estimated by two methods, i.e. ML based such as Quandt's method and nonlinear LS based namely Levenberg-Marquardt methods. The data on SES used in this application are household-expenditure and -income. First, some characteristics of the data, i.e. mean, standard deviation, minimum and maximum values are shown. After that, the suitable relationship between response variables as being household expenditure and explanatory variable as being household income are investigated by all four different regression models.

Where, household income data in SES data means average monthly total income per household and household expenditure data is average monthly total expenditure per household. From the Table 4.1, there exists the evidence that both income and expenditure data consist of outliers. Therefore, the LS regression might not be preferable. Instead of using the LS, we use other ways, i.e. Tobit, PW, and TP, to construct the relation of two variables. The results of this study are shown in the form of regression line of each of the four different methods and the average sum of square (ASSR) of them. The ASSR for this application is referred to Theorems 5 and 6 as shown in Chapter 2. RE is the ratio of the ASSR of the TP, PW and Tobit regressions to the LS regression.

Graphs of four fittings and their interpretations were only to present the data in year 2009 meanwhile the ASSR and RE values were calculated for data in both 2007 and 2009.

Table 4.1 Minimum, Maximum, Mean and Standard Deviation values of Household Income and Household Expenditure for Data on SES in Thailand duringYear 2009

	Characteristics			
Region	**Min**	**Max**	**Mean**	**S.D.**
Whole Kingdom				
Income	617	558,365	17,032	16,085
Expenditure	21	2,821,572	22,426	38,031
Bangkok Metropolis				
Income	3,165	393,229	31,114	26,726
Expenditure	574	2,062,805	44,471	80,068
Central				
Income	1,108	472,941	18,576	16,422
Expenditure	21	2,821,572	23,178	36,879
North				
Income	617	273,571	13,335	12,396
Expenditure	131	888,539	17,816	21,407
Northeast				
Income	776	558,365	14,853	13,557
Expenditure	115	23,804,30	19,900	36,693
South				
Income	1,000	345,458	17,951	15,124
Expenditure	448	1,005,000	23,692	32,782

Source of Data: National Statistical Office

The household expenditure and income in Bangkok Metropolis region is analyzed. Mean and standard deviation of income data are 44,471 baht and 80,068 baht, respectively. Their values of expenditure are 31,114 baht and 26,726 baht, respectively.

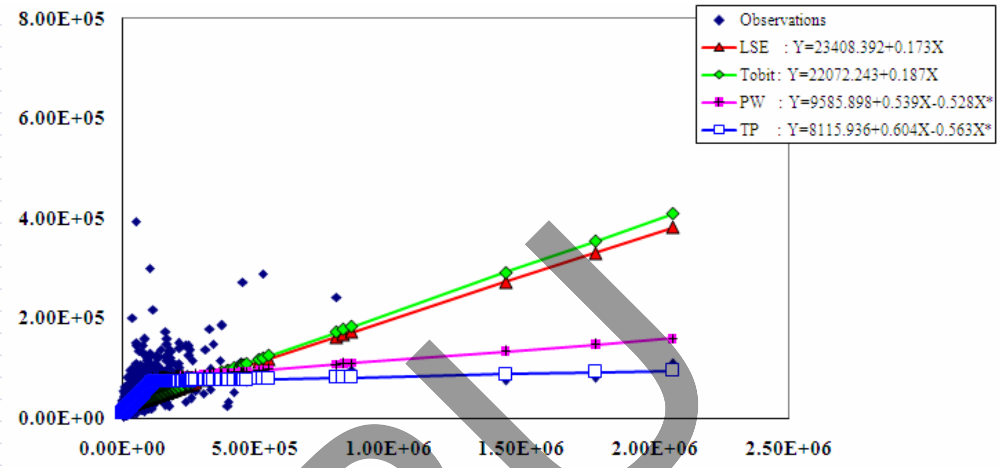


Figure 4.1 Observation and Four Regression Lines for Household-Expenditure and -Income Data for Bangkok Metropolis region on SES in year 2009

Source of Data : National Statistical Office

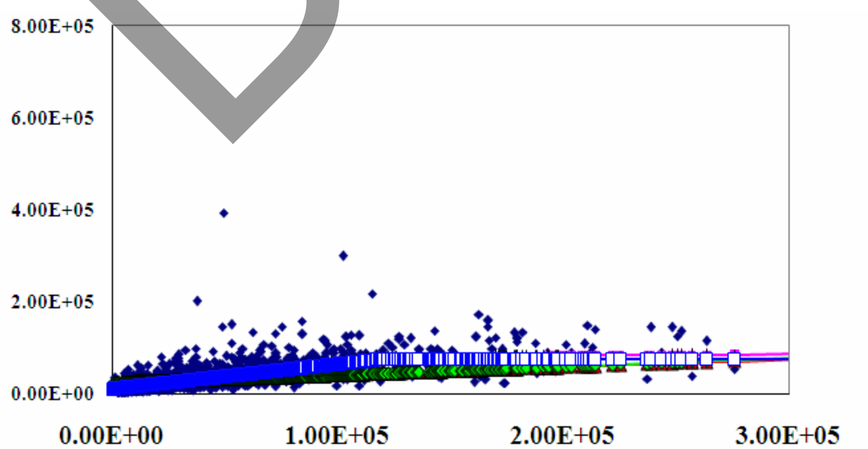


Figure 4.2 The Expansion of Figure 4.1 for the Range of Household Income as being between 0 and 300,000 Baht

When we test the normal assumption of expenditure data for whole kingdom and for each of region, we found that it is significantly violated. This mean that the data not normally distributed ($p<0.001$), might be caused by outliers, thus the LS regression does not properly act as evidence shown in Table 4.2 with largest ASSR, in the case of Bangkok Metropolis region, as 362×10^6 and RE as 1.0000. Tobit, PW and TP regression can be taken into account to cope with the case that data consist of outliers. When the data is limited in the space of dependent variables, Tobit regression is used to construct the linear relation. Nevertheless, in this case the Tobit regression is seem to be not appropriate as shown fitting line in Figure 4.1 with ASSR and RE by about 357×10^6 and 0.9864, respectively. Meanwhile, if the data are divided into two groups and fitted by PW regression, its result is better than both of Tobit and LS with ASSR and RE of it as 138×10^6 and 0.3823, respectively. Because the first regression regime properly fits the subsample, as shown in Figure 4.2, but the second regime of PW is still be affected by outlier data. Considering, therefore, TP regression is particularly best among all fours different method with ASSR 134×10^6 and RE by about 0.3709 for Levenberg-Marquardt method and with 136×10^6 and 0.3750 for Quandt's method, it means that the estimation method of the joined point in TP regression by the nonlinear LS based, for example Levenberg-Marquardt method is slightly better than by ML based such as Quandt's method.

In addition, we found that the joined point occurring on household income data estimated by Levenberg-Marquardt has a value of 118,213 baht while by Quandt's equals 122,500 baht.

Mean and standard deviation of income data in Central region are 23,178 baht and 36,879 baht, respectively. Their values of expenditure are 18,576 baht and 16.422 baht, respectively. Both the first and second regimes seem to be the TP and PW better than LS and Tobit.

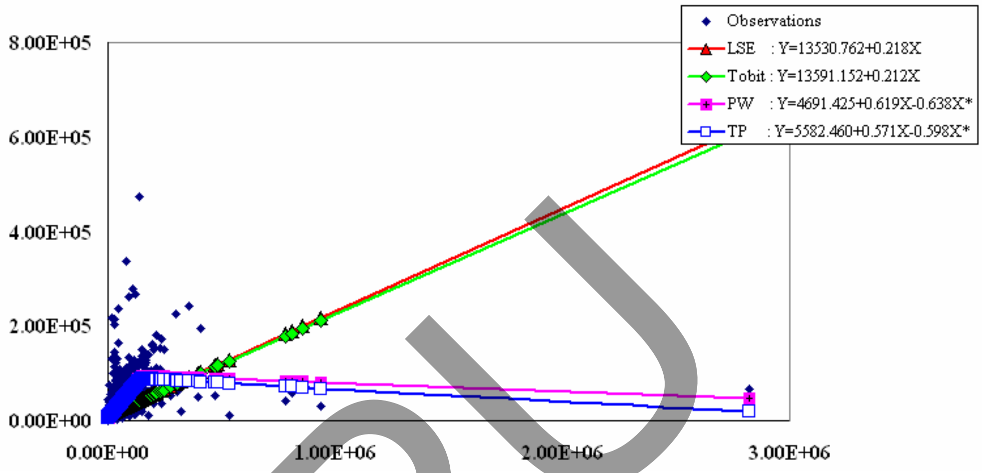


Figure 4.3 Observation and Four Regression Lines for Household-Expenditure and -Income Data for Central region on SES in year 2009

Source of Data : National Statistical Office

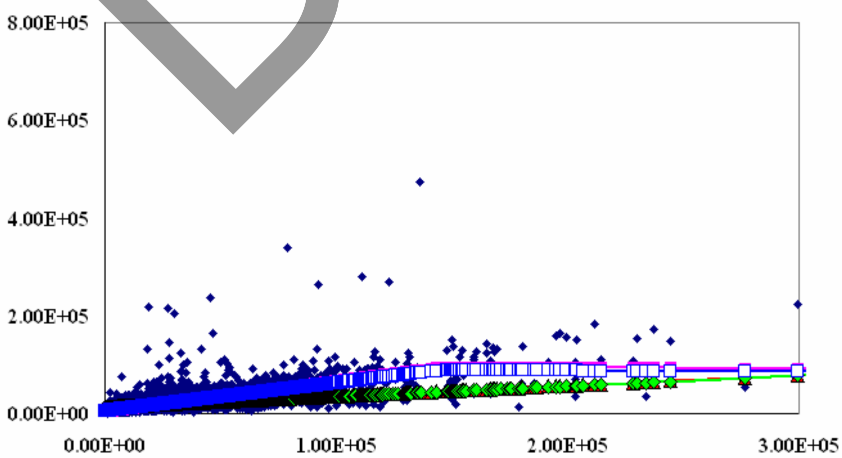


Figure 4.4 The Expansion of Figure 4.3 for the Range of Household Income as being between 0 and 300,000 Baht

In the sense that they can preferably represent the bulk of the data, when considering Figure 4.3 the Tobit regression, the dependent variable being limited by one value of upper limit whole the observed data is inappropriate. Nevertheless, if we divide the data into two groups and limit the dependent variable by upper limits for each group, i.e. fitting data by TP regression, the result yields better than both Tobit and LS with ASSR and RE as 65.30×10^{6} and 0.3727 for Levenberg-Marquardt method and as 65.38×10^{6} and 0.3731 for Quandt's method. Meanwhile, PW is slightly larger the value of ASSR 67.58×10^{6} and RE by about 0.3857 than TP.

Thus in the particular case, we can conclude that TP and PW can down-weigh value (reduce effect) of outliers than LS and Tobit regression. When considering the joined point in TP regression which is estimated by the nonlinear LS based, Levenberg-Marquardt method is slightly better than ML based such as Quandt's method.

In addition, we found that the joined point occurring in the space of household income data, which is estimated by Levenberg- Marquardt's has a value of 146,221 baht and by Quandt's equals 146,988 baht.

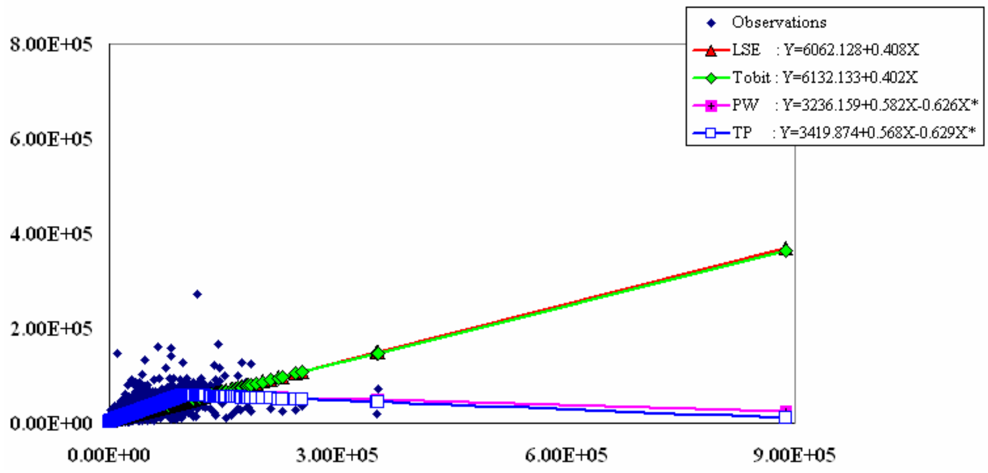


Figure 4.5 Observation and four Regression Lines for Household-Expenditure and -Income Data for North region on SES in year 2009

Source of Data : National Statistical Office

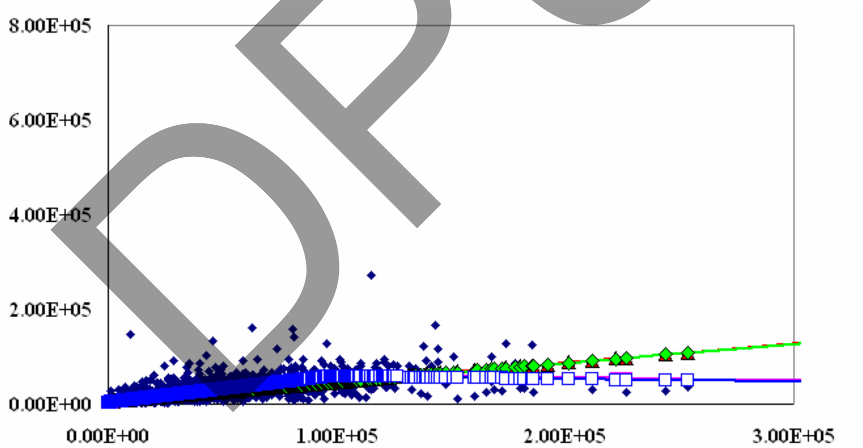


Figure 4.6 The Expansion of Figure 4.5 for the Range of Household Income as being between 0 and 300,000 Baht

Income data of households in North of Thailand 2009 have the mean and standard deviation as 17,816 baht and 21,407 baht, respectively. Expenditure data have them as 13,335 baht and 12,396 baht, respectively. From Figure 4.6, we found that the observed data consist of outliers both in *y-direction* and *x-direction* as same as the data in Central and Bangkok Metropolis. Therefore, the application of TP and PW regression is preferable and they gave better results than Tobit and LS regression as shown in the performance of Tables 4.2 and 4.3. ASSR and RE of TP regressions are each smallest as 39.96×10^6 and 0.5683 for Levenberg-Marquardt method and as 38.97×10^6 and 0.5685 for Quandt method, meanwhile, ASSR of each PW, Tobit and LS are 40.99×10^6, 69.51×10^6 and 70.31×10^6 , in that order. RE of each PW, Tobit and LS are 0.5830, 0.9886 and 1.0000, in that order. It was found that the estimation method of the joined point in TP regression by the nonlinear LS based, for example Levenberg-Marquardt method is slightly better than by ML based such as Quandt's method.

In addition, we found that the joined point appearing on the range of household income which is estimated by Levenberg- Marquardt has a value of 97,281 baht and by Quandt's equals 97,403 baht.

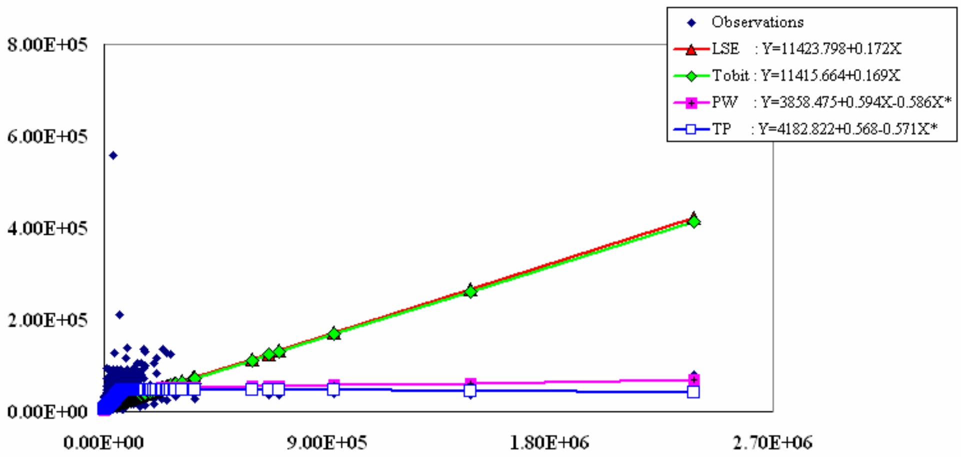


Figure 4.7 Observation and four Regression Lines for Household-Expenditure and -Income Data for Northeast region on SES in year 2009

Source of Data : National Statistical Office

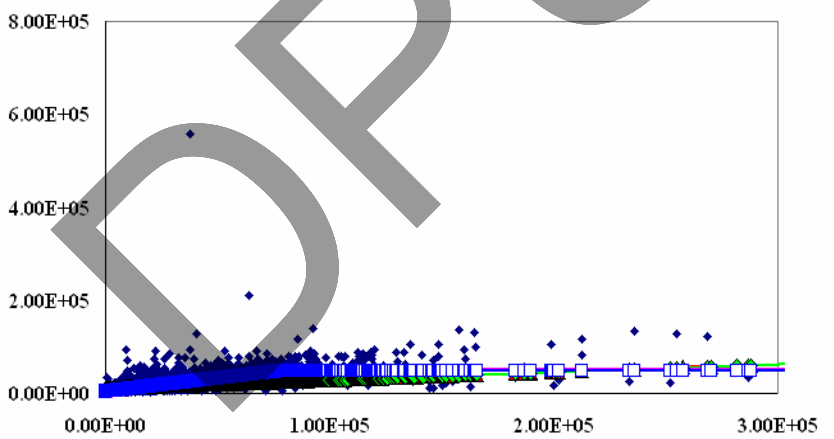


Figure 4.8 The Expansion of Figure 4.7 for the Range of Household Income as being between 0 and 300,000 Baht

Mean and standard deviation of income data in Northeast region are 19,900 baht and 36,693 baht, respectively. Their values of expenditure are 14,853 baht and 13,557 baht, respectively. From Figure 4.8 for the range of explanatory variable or household income as being between 0 to 300,000 baht, four different regression methods yield nearly the same result. This means that *y-direction* outliers appearing in the data do not much affect all the four regression lines. Whilst the *x-direction* outliers occurring on the second regime are much affect to LS and Tobit regression drawn far away from the true value, meanwhile, TP and PW seem to be more suitable than Tobit and LS. As evidence shown in Figure 4.7 and Tables 4.2 and 4.3, ASSR of each PW, Tobit and LS regression are 43.36×10^6, 112.43×10^6 and 113.82×10^6, respectively. RE of each PW, Tobit and LS regression are 0.3809, 0.9879 and 1.0000, respectively.

When considering the estimation method of the joined point in TP regression, we found that the nonlinear LS based, namely Levenberg-Marquardt method is slightly better than ML based such as Quandt's method with ASSR of each being as 41.49×10^6 and 41.56×10^6, respectively. RE of each estimator are as 0.3645 and 0.3651, respectively.

In addition, we found that the joined point which appears on the space of income data estimated by Levenberg- Marquardt's has a value of 77,965 baht and by Quandt's equals 78,081 baht.

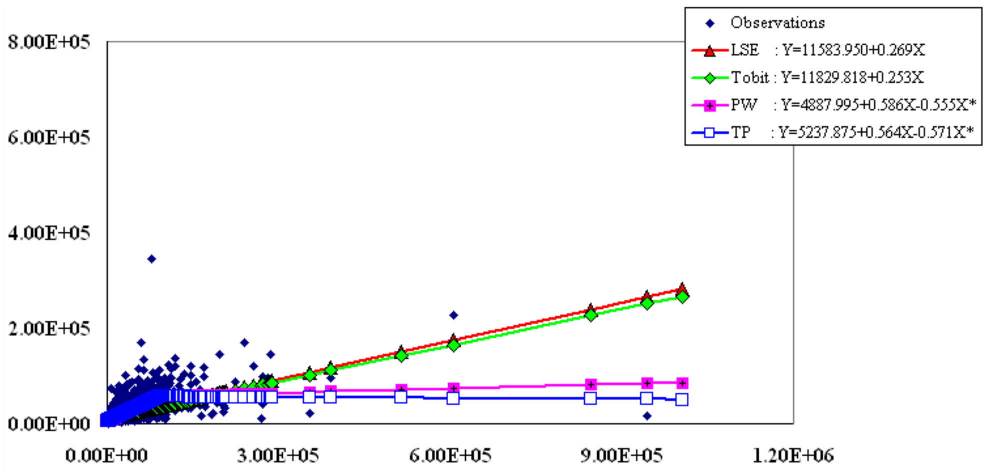


Figure 4.9 Observation and four Regression Lines for Household-Expenditure and -Income Data for South region on SES in year 2009

Source of Data : National Statistical Office

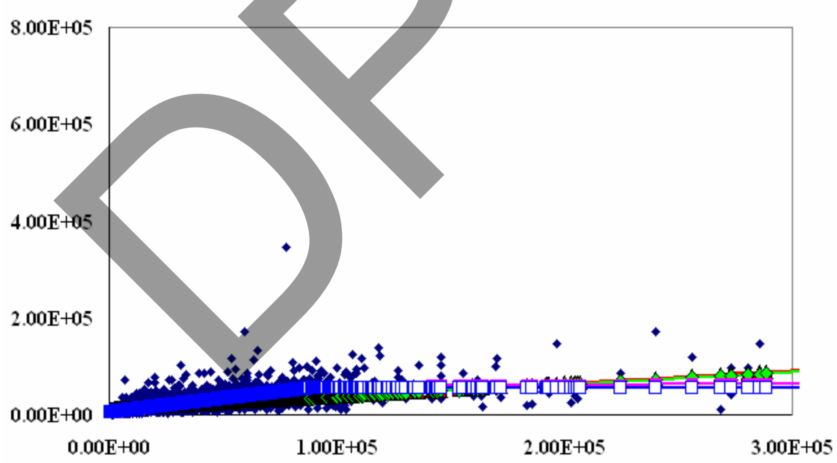


Figure 4.10 The Expansion of Figure 4.7 for the Range of Household Income as being between 0 and 300,000 Baht

The mean and standard deviation of household expenditure are respectively 17,951 baht and 15,124 baht and of income are 23,692 baht and 32,782 baht. The results of all the four regression models look like the observed data of Central, Bangkok Metropolis, North and Northeast in Thailand. The LS and Tobit, in this particular case, are much affected by *x-direction y*outliers. Nevertheless, TP and PW yield better results than Tobit and LS regression as shown the performance by ASSR and RE in Tables 4.2 and 4.3. The smallest value of ASSR by about 63.8×10^6 is of TP with joined point estimated by Levenberg Marquardt method and followed by of PW as 65.13×10^6, Tobit as 12.63×10^6 and LS as 157.11×10^6, in that order. Therefore the smallest RE is also of TP as 0.4064 and followed by PW as 0.4145, Tobit as 0.8039 and LS as 1.0000, in that order.

When considering the estimation method of the joined point in TP regression, we found that the nonlinear LS based, for example Levenberg-Marquardt method is slightly better than ML based such as Quandt's method with ASSR of each being as 63.85×10^6 and 64.05×10^6, respectively. RE of each estimator is as 0.4064 and 0.4077, respectively.

In addition, we found that the joined point which appears on the space of income data is estimated by Levenberg- Marquardt's has value of 90,790 baht and by Quandt's equals 91,818 baht.

Table 4.2 ASSR for four different regression models on SES Data in Thailand, Year 2009

Region	Joined Point in TP	ASSR			
		LS	Tobit	PW	TP
Bangkok Metropolis					
Levenberg Marquardt Method	118,213	361,611,477	356,678,962	138,229,145	134,137,380
Quandt's Method	122,500			138,262,874	135,614,342
Central					
Levenberg Marquardt Method	146,221	175,233,047	172,454,090	67,579,056	65,302,540
Quandt's Method	146,988			67,579,975	65,380,687
North					
Levenberg Marquardt Method	97,281	70,309,269	69,505,579	40,990,838	39,958,789
Quandt's Method	97,403			40,990,875	39,967,927
Northeast					
Levenberg Marquardt Method	77,965	113,816,058	112,433,338	43,358,120	41,487,126
Quandt's Method	78,081			43,680,080	41,558,970
South					
Levenberg Marquardt Method	90,790	157,113,391	126,298,135	65,125,721	63,845,878
Quandt's Method	91,818			66,127,146	64,048,316

From the Tables 4.2 - 4.3 and Figures 4.1 – 4.10, it was found that outliers in *y-direction* and in *x-direction* for the data on SES in Thailand can be made "down-weight" the values or reduce its effect by both TP and PW regressions.

Table 4.3 RE of four different regression models on SES Data in Thailand, Year 2009

Region	ASSR			
	LS	**Tobit**	**PW**	**TP**
Bangkok Metropolis				
Levenberg Marquardt Method	1.0000	0.9864	0.3823	0.3709
Quandt's Method			0.3824	0.3750
Central				
Levenberg Marquardt Method	1.0000	0.9841	0.3857	0.3727
Quandt's Method			0.3857	0.3731
North				
Levenberg Marquardt Method	1.0000	0.9886	0.5830	0.5683
Quandt's Method			0.5830	0.5685
Northeast				
Levenberg Marquardt Method	1.0000	0.9879	0.3809	0.3645
Quandt's Method			0.3838	0.3651
South				
Levenberg Marquardt Method	1.0000	0.8039	0.4145	0.4064
Quandt's Method			0.4209	0.4077

From the Tables 4.4 - 4.5, there is the evidence that outliers in *y-direction* and in *x-direction* for the data on SES in Thailand year 2007 can also be made to "down-weight" the values or reduced their effect by both TP and PW regressions. This supports the results of SES data in year 2009.

Table 4.4 ASSR for four different regression models on SES Data in Thailand, Year 2007

Region	Joined Point in TP	ASSR			
		LS	Tobit	PW	TP
Bangkok Metropolis					
Levenberg Marquardt Method	87,357	305,424,398	265,511,268	99,005,602	91,996,875
Quandt's Method	101,734			101,472,751	99,386,841
Central					
Levenberg Marquardt Method	108,823	135,527,007	134,502,303	49,467,872	49,423,115
Quandt's Method	107,172			58,816,868	57,362,815
North					
Levenberg Marquardt Method	74,418	60,414,527	58,928,011	30,430,406	29,269,428
Quandt's Method	79,802			32,289,353	31,653,535
Northeast					
Levenberg Marquardt Method	64,712	89,959,059	89,453,915	35,973,115	34,048,778
Quandt's Method	65,257			36,569,438	35,171,241
South					
Levenberg Marquardt Method	64,403	124,510,620	88,043,683	50,917,754	49,742,225
Quandt's Method	51,973			53,187,474	52,236,496

Table 4.5 RE of four different regression models on SES Data in Thailand, Year 2007

Region	ASSR			
	LS	Tobit	PW	TP
Bangkok Metropolis				
Levenberg Marquardt Method	1.0000	0.8693	0.3242	0.3012
Quandt's Method			0.3322	0.3254
Central				
Levenberg Marquardt Method	1.0000	0.9924	0.3650	0.3647
Quandt's Method			0.4340	0.4233
North				
Levenberg Marquardt Method	1.0000	0.9754	0.5037	0.4845
Quandt's Method			0.5345	0.5239
Northeast				
Levenberg Marquardt Method	1.0000	0.9944	0.3999	0.3785
Quandt's Method			0.4065	0.3910
South				
Levenberg Marquardt Method	1.0000	0.7071	0.4089	0.3995
Quandt's Method			0.4272	0.4195

4.2 Simulation Studies

The performance of Tobit-piecewise (TP) regression model is investigated in term of the average sum of squares of residuals (ASSR) (Mekbunditkul, 2011) by simulation studies. There are two situations to be considered, namely the *y-direction*, and *xy-direction*. Nevertheless, other two situations are not taken into account. As without the existence of outliers, it is found that the ASSR of Tobit is equal to the LS regression model while the ASSR for PW and TP are the same. However, both Tobit and LS results were significantly different from PW and TP methods. The data fitted by PW and TP regression models yielded the value of ASSR that were smaller than the Tobit and LS methods' by about RE equal to 0.35 (Mekbunditkul, 2010). This mean that the PW and TP regressions are more suitable than LS and Tobit models. In the existence of *x-direction* outliers, numerical examples and simulation results as studied in Mekbunditkul's research provided the evidence that Tobit and LS were identical. Meanwhile PW and TP were the same and they were significantly better than Tobit and LS regressions.

However, there has been no study in terms of joined point estimation so that the simulation is needed to compare the potential of four estimators again. Attributes to the Monte Carlo technique are specified as followed: Sample sizes are varied, namely 10, 20, 30,…, 100 and the percentage of outliers considered are 5%, 10%, 15% and 20%. The ASSR and RE of each estimator are determined.

Case 1: Outliers in the *y-direction*

1. Generate $x_i \sim N(2.5,4)$, for i=1, 2, …, $\frac{n}{2}$, and $x_i \sim N(7.5,4)$, for i = $\frac{n}{2}+1, \frac{n}{2}+2, \ldots, n$

2. Generate $\varepsilon_i \sim N\left(0,\sigma_i^2\right)$, for $i = 1, 2, \ldots, (1-\alpha)n$, where

$$\sigma_i^2 = \begin{cases} 4 & \text{if } \upsilon_i \le 5, \\ 16 & \text{if } \upsilon_i > 5. \end{cases}$$

3. Generate $\varepsilon_i \sim N(0,144)$, for $i = (1-\alpha)n+1,\ (1-\alpha)n+2,\ \ldots,\ n$, for αn outliers, where α is given in advance

4. Calculate y_i as indicated in Case 1

Case 2: Outliers in the *xy-direction*

1. Generate $x_i \sim N(2.5,4)$, for $i=1,\ 2,\ \ldots,\ \left\lfloor \frac{(1-\alpha)n}{2} \right\rfloor + 1$, and $x_i \sim N(7.5,4)$, for $i = \left\lfloor \frac{(1-\alpha)n}{2} \right\rfloor + 2, \ldots, (1-\alpha)n$

2. Generate $x_i \sim N(15,16)$, for $i = (1-\alpha)n+1, \ldots, n$, for αn outliers

3. Generate $\varepsilon_i \sim N\left(0,\sigma_i^2\right)$, for $i = 1, 2, \ldots, (1-\alpha)n$, where

$$\sigma_i^2 = \begin{cases} 4 & \text{if } \upsilon_i \le 5, \\ 16 & \text{if } \upsilon_i > 5. \end{cases}$$

Each statistic, namely ASSR and RE, was obtained for the four methods of estimation. For each method, the statistic was calculated 1,000 times. The average of each ASSR and RE from the four methods is compared. For each of the above cases, a random sample of size n, where k=1 for a simple linear regression.

4.2.1 Outliers in the *y-direction*

For *y-direction* outliers, the average values of ASSR and RE for 1,000 samples, with a certain percentage of outliers and different estimates of the regression coefficients, i.e. LS, Tobit, PW and TP, are presented in Tables 4.6 and 4.7 and Figure 4.11. The value of ASSR of TP and of PW regression models with each joined point estimated by Levenberg-Marquardt method, nonlinear LS based, is shown in Table 4.6 corresponding to Figure 4.11. In addition, the value of each ASSR and RE for joined point in TP and PW estimated by Quandt's method, ML based, are presented in Tables 4.8 and 4.9 and also for Figure 4.12.

Table 4.6 ASSR of four different regression models for Levenberg-Marquardt method in cases of *y-direction* outliers

Sample Size	% of Y-Outliers	ASSR[1]				
		LS	Tobit	PW	TP	Joint Point in TP
10	5	-	-	-	-	-
	10	7,691	5,657	6,685	3,253	4.02
	15	-	-	-	-	-
	20	12,041	6,005	10,953	4,454	3.97
20	5	4,519	2,280	4,116	1,463	4.37
	10	7,841	5,233	7,237	3,027	4.07
	15	10,286	5,546	9,655	3,825	4.05
	20	11,954	5,903	10,874	4,247	4.01
40	5	4,569	2,251	4,243	1,599	4.45
	10	7,968	5,147	7,641	3,132	4.15
	15	10,436	5,525	10,097	3,651	4.09
	20	12,009	5,633	11,615	4,010	4.04
60	5	4,588	2,184	4,280	1,521	4.71
	10	7,998	4,869	7,620	3,108	4.34
	15	10,490	5,463	10,149	3,200	4.25
	20	12,067	5,636	11,641	3,240	4.17
100	5	4,594	2,116	4,259	1,440	4.74
	10	8,009	4,873	7,426	2,849	4.41
	15	10,550	5,435	9,779	2,993	4.28
	20	12,202	5,738	11,344	3,052	4.19

[1] Average sum of squares residual (ASSR) is used and recommended to be used as a measure of model precision. Caution should be noted. The MSE in regression under classical assumption is usually an estimator of the variance of error (σ^2) and of dependent variable as well. As number of observation n approaches infinity, such MSE should converge in probability to σ^2 but not zero (Mekbunditkul, 2010).

Table 4.7 RE of four different regression models for Levenberg-Marquardt method in cases of *y-direction* outliers

Sample Size	% of Y-Outliers	ASSR LS	Tobit	PW	TP
10	5	-	-	-	-
	10	1.0000	0.7355	0.8691	0.4230
	15	-	-	-	-
	20	1.0000	0.4987	0.9096	0.3699
20	5	1.0000	0.5046	0.9109	0.3238
	10	1.0000	0.6674	0.9229	0.3861
	15	1.0000	0.5392	0.9387	0.3718
	20	1.0000	0.4938	0.9096	0.3552
40	5	1.0000	0.4927	0.9286	0.3499
	10	1.0000	0.6460	0.9590	0.3931
	15	1.0000	0.5294	0.9675	0.3498
	20	1.0000	0.4691	0.9672	0.3339
60	5	1.0000	0.4760	0.9327	0.3314
	10	1.0000	0.6088	0.9527	0.3886
	15	1.0000	0.5208	0.9675	0.3051
	20	1.0000	0.4670	0.9647	0.2685
100	5	1.0000	0.4605	0.9271	0.3134
	10	1.0000	0.6084	0.9271	0.3557
	15	1.0000	0.5152	0.9270	0.2837
	20	1.0000	0.4702	0.9296	0.2501

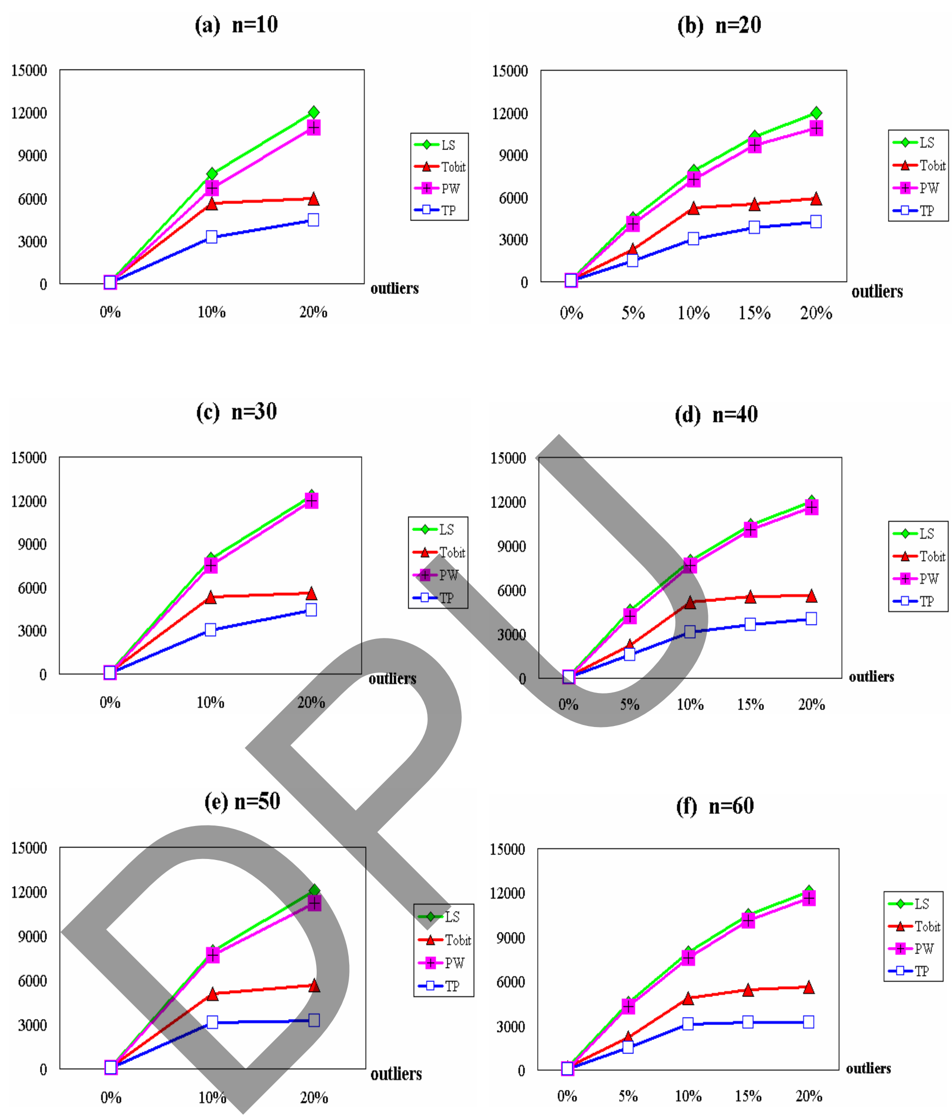


Figure 4.11 ASSR of four different regression models for Levenberg-Marquardt method varied by percentage of outliers when n=10, 20,…, 100 where *y-direction* outliers exist

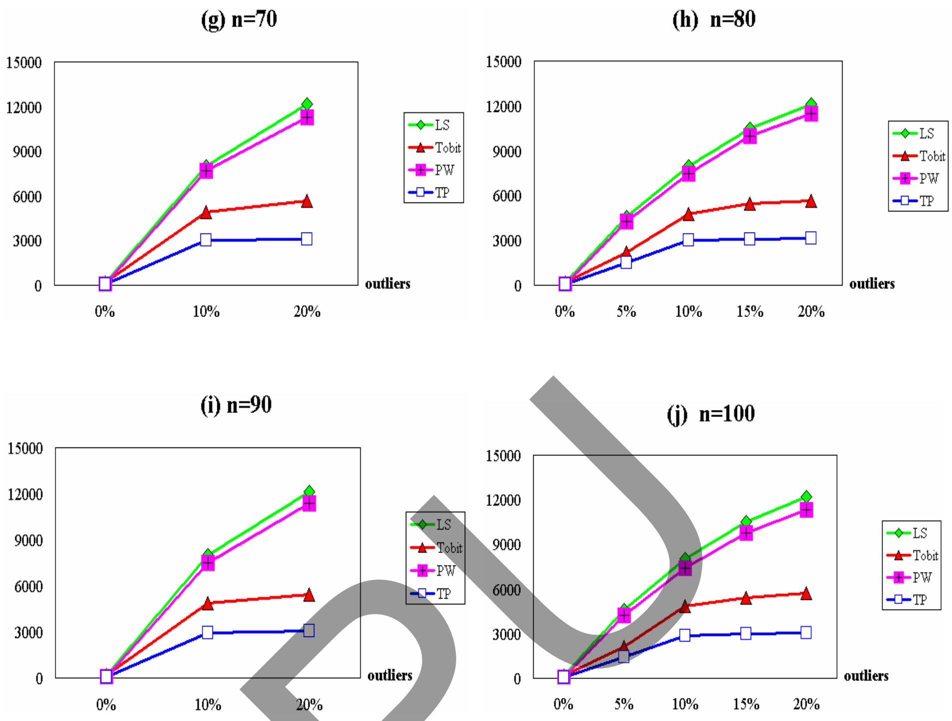


Figure 4.11 ASSR of four different regression models for Levenberg-Marquardt method varied by percentage of outliers when n=10, 20,…, 100 where *y-direction* outliers exist (continued)

From Table 4.6 and Figure 4.11, for all percentages of outliers we find that the significantly smallest ASSR is of TP regression model, followed by of Tobit, PW and LS regression models, in that order.

Considering the information in Table 4.6, we can see that when the percentage of outliers increases, TP regression model with joined point estimated by Levenberg-Marquardt method is preferable to LS model. Furthermore, it was found that both Tobit and PW regression models fit better than LS model for all values of percentage of outliers. Thus, in this particular case, it can be concluded that the TP regression model yields the best results followed by the Tobit, PW and LS regression models, in that order. Moreover, in the case where outliers exist in the *y-direction,* not only TP but also the Tobit regression model is preferable to both PW and LS. The results correspond to findings in Mekbunditkul's research when the joined point is assumed to be known.

In Figure 4.11, it was found that the values of ASSR of four different regression models increase when the percentages of outliers increase for all sample sizes are considered. For the value of joined point (see Tables 4.6 and 4.8) that is estimated by both Levenberg-Marquardt and Quandt's method, we find that it is biased downward from the true value as fixed in advance by 5. The mean of joined point estimated by Levenberg-Marquardt is 4.23 and its standard deviation is 0.2288. Whilst, the mean of joined point estimated by Quandt's method is 3.97 and its standard deviation is 0.3781.

In addition, when the percentage of outliers increases, the bias increases for all sample sizes considered. Meanwhile, the sample size increases then the bias decreases for all percentages of outliers.

Next, Table 4.8 and Figure 4.12 exhibit the results of simulation studies for the *y-direction* outliers, when unknown joined points in TP and in PW regression models are estimated by Quandt's method, ML based.

Table 4.8 ASSR of four different regression models for Quandt's method in cases of *y-direction* outliers

Sample Size	% of Y-Outliers	ASSR LS	Tobit	PW	TP	Joined Point in TP
10	5	-	-	-	-	-
	10	7,691	5,657	6,789	3,355	3.42
	15	-	-	-	-	-
	20	12,041	6,005	11,175	4,668	3.29
20	5	4,519	2,280	4,185	1,530	4.14
	10	7,841	5,233	7,339	3,129	3.63
	15	10,286	5,546	9,829	3,997	3.49
	20	11,954	5,903	11,085	4,457	3.28
40	5	4,569	2,251	4,308	1,664	4.29
	10	7,968	5,147	7,743	3,233	4.05
	15	10,436	5,525	10,271	3,822	4.03
	20	12,009	5,633	11,828	4,221	4.00
60	5	4,588	2,184	4,345	1,586	4.49
	10	7,998	4,869	7,721	3,209	4.10
	15	10,490	5,463	10,322	3,371	4.08
	20	12,067	5,636	11,852	3,450	4.02
100	5	4,594	2,116	4,328	1,506	4.58
	10	8,009	4,873	7,529	2,950	4.13
	15	10,550	5,435	9,958	3,165	4.27
	20	12,202	5,738	11,560	3,263	4.14

Table 4.9 RE of four different regression models for Quandt's method in cases of *y-direction* outliers

Sample Size	**% of Y-Outliers**	**ASSR**			
		LS	**Tobit**	**PW**	**TP**
10	5	-	-	-	-
	10	1.0000	0.7355	0.8828	0.4363
	15	-	-	-	-
	20	1.0000	0.4987	0.9106	0.3703
20	5	1.0000	0.5046	0.9261	0.3385
	10	1.0000	0.6674	0.9361	0.3990
	15	1.0000	0.5392	0.9556	0.3886
	20	1.0000	0.4938	0.9273	0.3729
40	5	1.0000	0.4927	0.9429	0.3641
	10	1.0000	0.6460	0.9719	0.4058
	15	1.0000	0.5294	0.9842	0.3662
	20	1.0000	0.4691	0.9849	0.3515
60	5	1.0000	0.4760	0.9470	0.3456
	10	1.0000	0.6088	0.9654	0.4012
	15	1.0000	0.5208	0.9839	0.3214
	20	1.0000	0.4670	0.9822	0.2859
100	5	1.0000	0.4605	0.9422	0.3278
	10	1.0000	0.6084	0.9400	0.3683
	15	1.0000	0.5152	0.9439	0.3000
	20	1.0000	0.4702	0.9474	0.2674

From Table 4.8 and 4.9, there is evidence that TP regression with the unknown joined point estimated by Quandt's method yields the smallest ASSR and RE among all the different estimations and followed by Tobit, PW with the unknown joined point estimated by Quandt's method and LS, in that order. In addition, from Figure 4.12, it is found that the value of ASSR for all types of estimations increases when the percentage of outliers increases for all sample sizes considered.

When the comparison of ASSR for joined point estimated by ML based and nonlinear LS based was considered, it was found that TP regression with the unknown joined point estimated by nonlinear LS based yields non-significantly smaller ASSR than by ML based for all values of percentage of outliers considered.

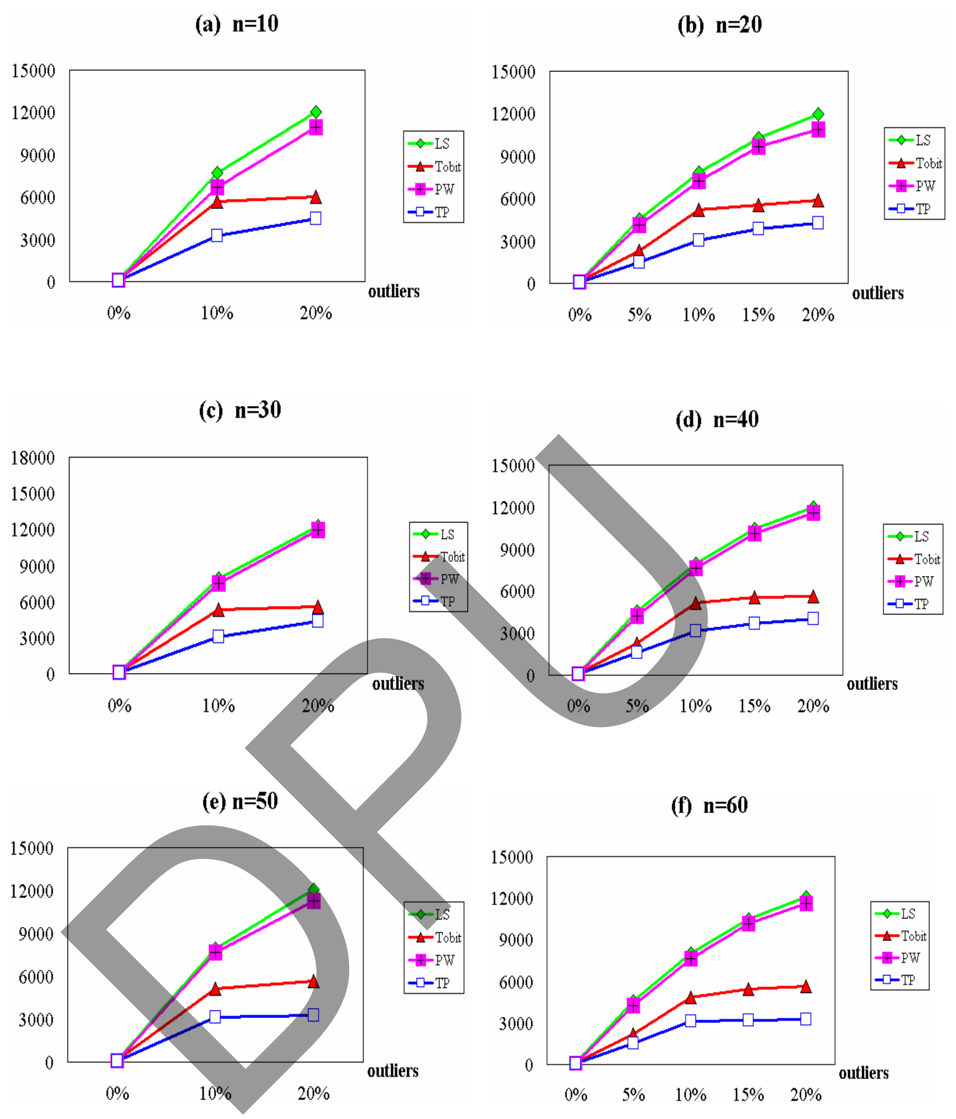


Figure 4.12 ASSR of four different Regression Models for Quandt's method varied by percentage of Outliers when n=10, 20,…, 100 where *y-direction* Outliers exist

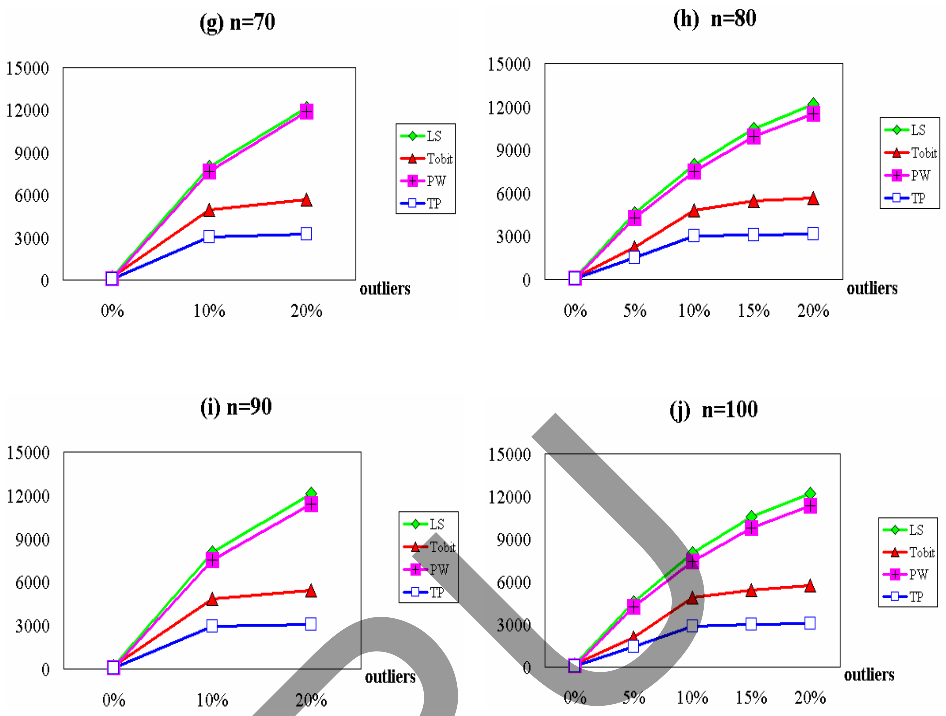


Figure 4.12 ASSR of four different Regression Models for Quandt's method varied by percentage of Outliers when n=10, 20,…, 100 where *y-direction* Outliers exist (continued)

4.2.2 Outliers in the *xy-direction*

For samples with *xy-direction* outliers, Tables 4.10 and 4.12 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 4.13 and 4.14.

The difference of Tables 4.10 and 4.12 is that ASSR value appearing on the Table 4.10 is from TP and PW in which their joined point estimated by Levenberg-Marquardt method. Meanwhile, the ASSR shown in Table 4.12 is obtained from TP model in which the joined point is estimated by Quandt's method.

Table 4.10 ASSR of four different regression models for Levenberg-Marquardt method in cases of *xy-direction* outliers

Sample Size	% of Y-Outliers	ASSR LS	Tobit	PW	TP
10	5	-	-	-	-
	10	4,157	1,416	2,126	992
	15	-	-	-	-
	20	5,070	4,837	2,803	1,317
20	5	3,069	581	1,617	481
	10	4,270	1,337	2,148	1,046
	15	5,204	4,769	2,718	1,171
	20	5,286	4,733	2,782	1,607
40	5	3,077	495	1,601	429
	10	4,636	1,044	2,362	479
	15	5,245	2,221	2,562	761
	20	5,497	4,442	2,708	1,321
60	5	3,154	551	1,717	296
	10	4,473	741	2,058	616
	15	5,102	1,559	2,303	815
	20	5,354	3,900	2,437	1,071
100	5	3,118	493	1,617	244
	10	4,552	838	2,094	471
	15	5,196	1,615	2,284	691
	20	5,459	3,641	2,429	894

Table 4.11 RE of four different regression models for Levenberg-Marquardt method in cases of *xy-direction* outliers

Sample Size	% of Y-Outliers	ASSR LS	Tobit	PW	TP
10	5	-	-	-	-
	10	1.0000	0.3406	0.5113	0.2386
	15	-	-	-	-
	20	1.0000	0.9541	0.5528	0.2597
20	5	1.0000	0.1892	0.5269	0.1566
	10	1.0000	0.3132	0.5030	0.2449
	15	1.0000	0.9163	0.5222	0.2250
	20	1.0000	0.8953	0.5263	0.3040
40	5	1.0000	0.1610	0.5204	0.1394
	10	1.0000	0.2252	0.5095	0.1034
	15	1.0000	0.4235	0.4884	0.1452
	20	1.0000	0.8081	0.4927	0.2404
60	5	1.0000	0.1748	0.5444	0.0940
	10	1.0000	0.1656	0.4601	0.1377
	15	1.0000	0.3055	0.4513	0.1597
	20	1.0000	0.7284	0.4552	0.2001
100	5	1.0000	0.1580	0.5186	0.0784
	10	1.0000	0.1840	0.4601	0.1036
	15	1.0000	0.3108	0.4396	0.1330
	20	1.0000	0.6670	0.4449	0.1638

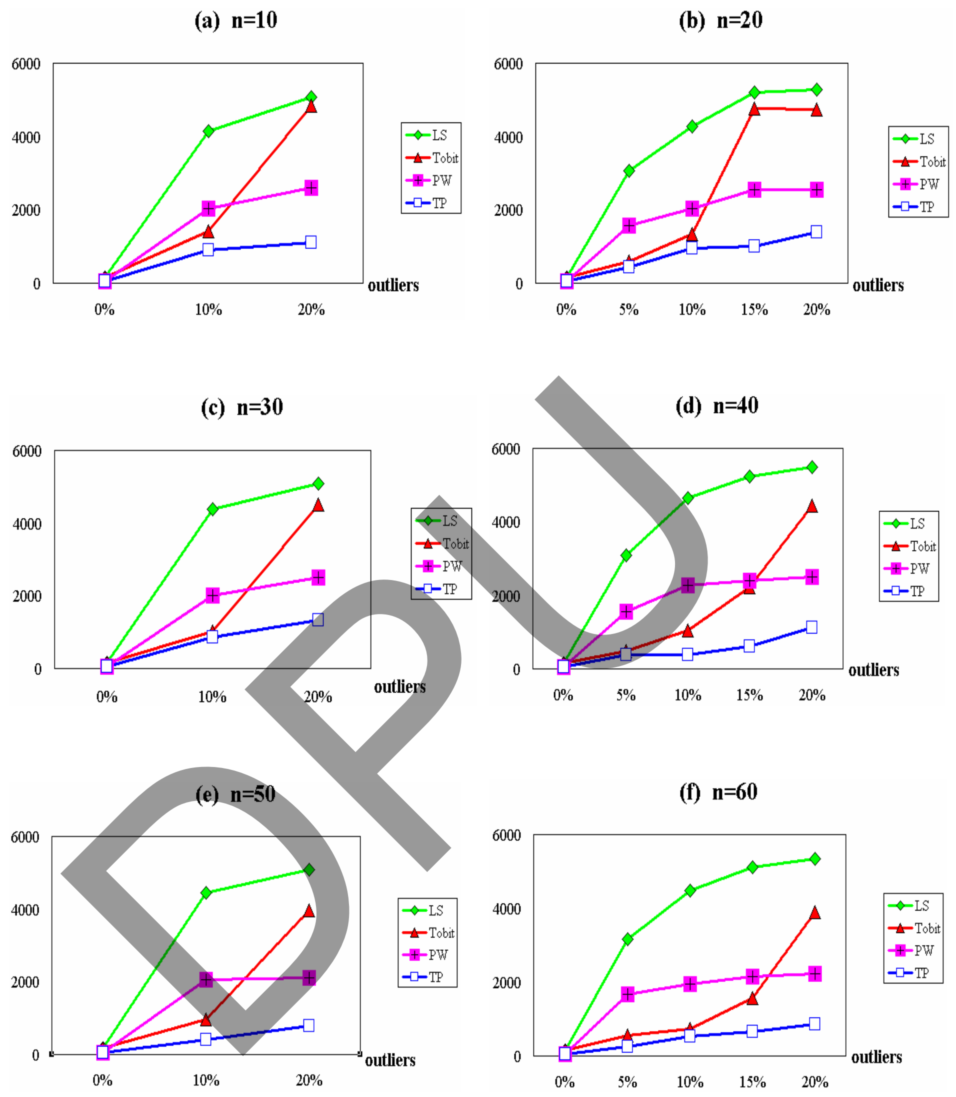


Figure 4.13 ASSR of four different regression models for Levenberg-Marquardt method varied by percentage of outliers when n=10, 20,…, 100 where *xy-direction* outliers exist

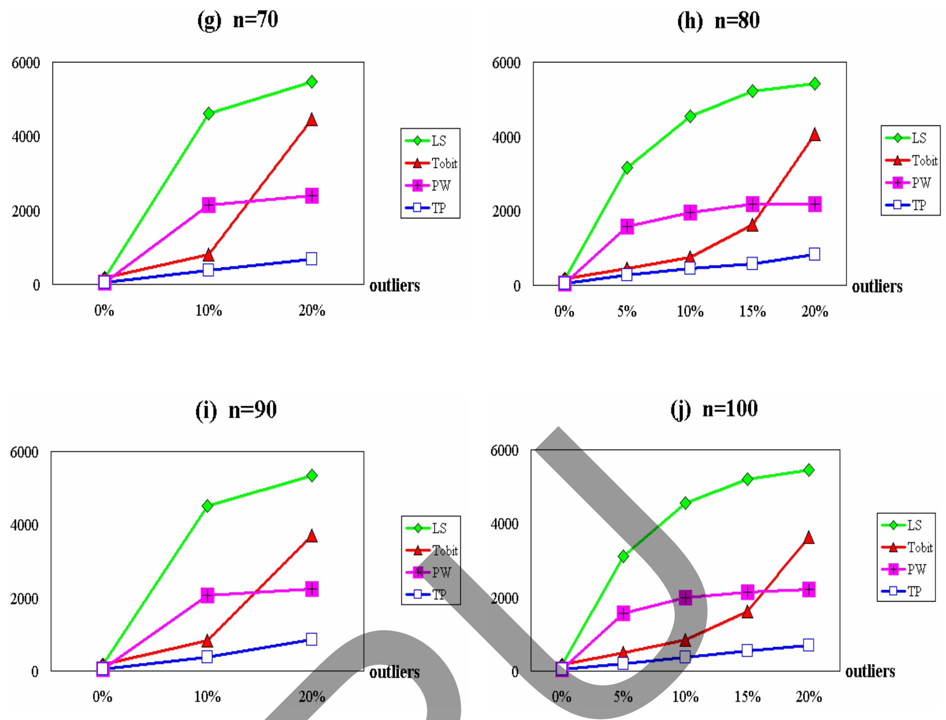


Figure 4.13 ASSR of four different regression models for Levenberg-Marquardt method varied by percentage of outliers when n=10, 20,…, 100 where *xy-direction* outliers exist (continued)

From Tables 4.10 and 4.12, where the percentages of outliers are 5%, 10% and 15%, it is found that the significantly smallest ASSR and RE (see Tables 4.11 and 4.13) are of the TP regression model then followed by that of Tobit, PW and LS regression models, in that order. Meanwhile, at 20% outliers, the smallest ASSR and RE (see Tables 4.11 and 4.13) are from the TP regression model then followed sequentially by that of PW, Tobit, and LS. These results indicate that, in the case where *xy-direction* outliers exist, the potential applicability of Tobit regression decreases when the percentage of outliers increases. It can be seen that the PW regression model slightly changes when the percentage of outliers increases, as evident in Figures 4.13 and 4.14. Moreover, in the case of outliers existing in the *xy-direction* between 5% and 15%, it was found that not only the TP regression model but also Tobit is preferable to both PW and LS. Meanwhile at 20% outliers, the PW regression is preferable to both Tobit and LS since, when the percentage of outliers is high, it indicates that data should be divided into two groups and ought to be fit by either PW or TP. These results are the same as those from Mekbunditkul's study in which each joined point in both TP and PW regression models are assumed to be known.

When looking at Figures 4.13 and 4.14, 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 is found that when the percentage of outliers increases, the ASSR of TP and PW slightly increases.

Considerably, the value of joined point in TP regression model, which is estimated by Levenberg-Marquardt and Quandt's method, is biased downward from the true value as fixed in advance by 5.

The mean of joined point estimated by Levenberg-Marquardt is 4.67 and its standard deviation is 0.2134. Whilst, the mean of joined point estimated by Qunadt's method is 4.29 and its standard deviation is 0.3426.

In addition, when the percentage of outliers increases the bias increases for all sample sizes considered. Meanwhile, when the sample size increases the bias decreases for all percentages of outliers.

Table 4.12 ASSR of four different regression models for Quandt's method in cases of *xy-direction* outliers

Sample Size	% of Y-Outliers	ASSR LS	Tobit	PW	TP
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

Table 4.13 RE of four different regression models for Quandt's method in cases of *xy-direction* outliers

Sample Size	% of Y-Outliers	ASSR LS	Tobit	PW	TP
10	5				
	10	1.0000	0.3406	0.4884	0.2157
	15				
	20	1.0000	0.9541	0.5116	0.2185
20	5	1.0000	0.1892	0.5138	0.1436
	10	1.0000	0.3132	0.4807	0.2227
	15	1.0000	0.9163	0.4936	0.1963
	20	1.0000	0.8953	0.4868	0.2645
40	5	1.0000	0.1610	0.5074	0.1264
	10	1.0000	0.2252	0.4891	0.0829
	15	1.0000	0.4235	0.4600	0.1168
	20	1.0000	0.8081	0.4547	0.2024
60	5	1.0000	0.1748	0.5317	0.0813
	10	1.0000	0.1656	0.4389	0.1164
	15	1.0000	0.3055	0.4221	0.1305
	20	1.0000	0.7284	0.4161	0.1611
100	5	1.0000	0.1580	0.5058	0.0655
	10	1.0000	0.1840	0.4392	0.0827
	15	1.0000	0.3108	0.4110	0.1043
	20	1.0000	0.6670	0.4066	0.1255

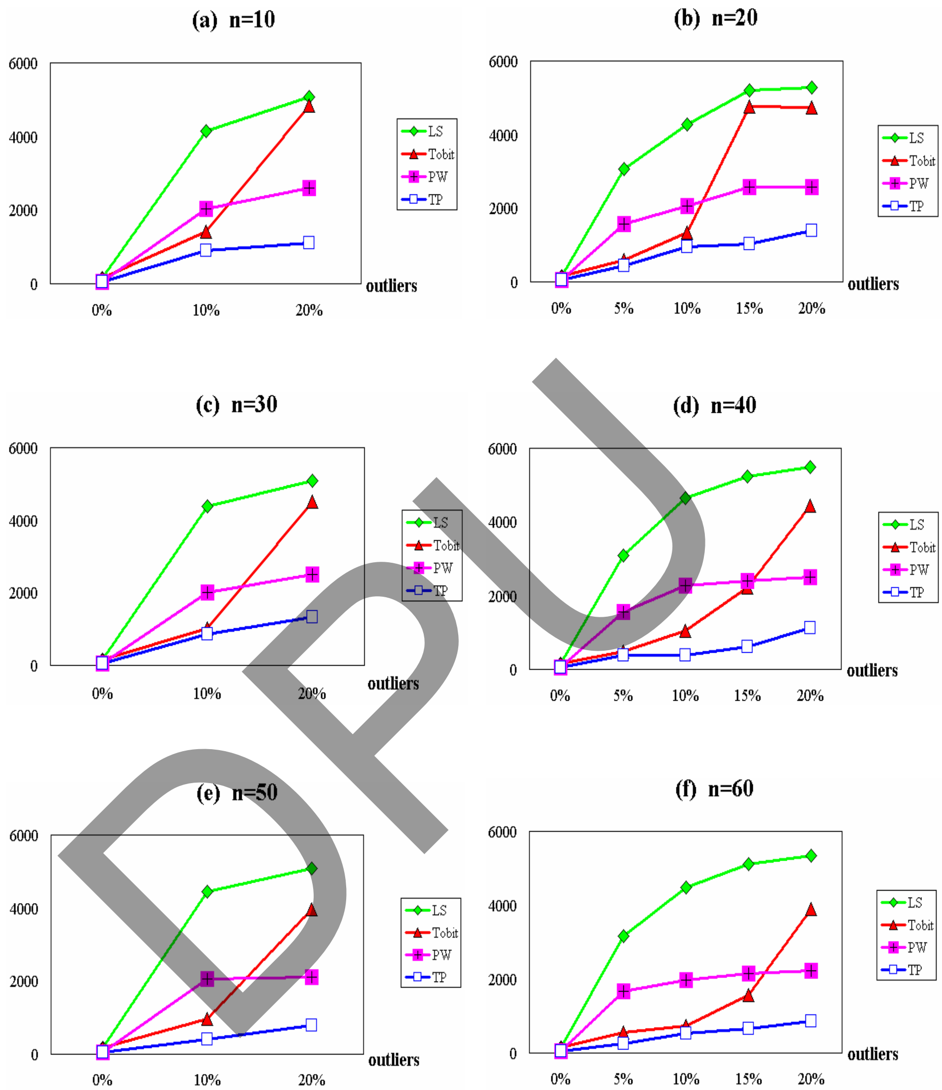


Figure 4.14 ASSR of four different regression models for Quandt's method varied by percentage of outliers when n=10, 20,…, 100 where *xy-direction* outliers exist

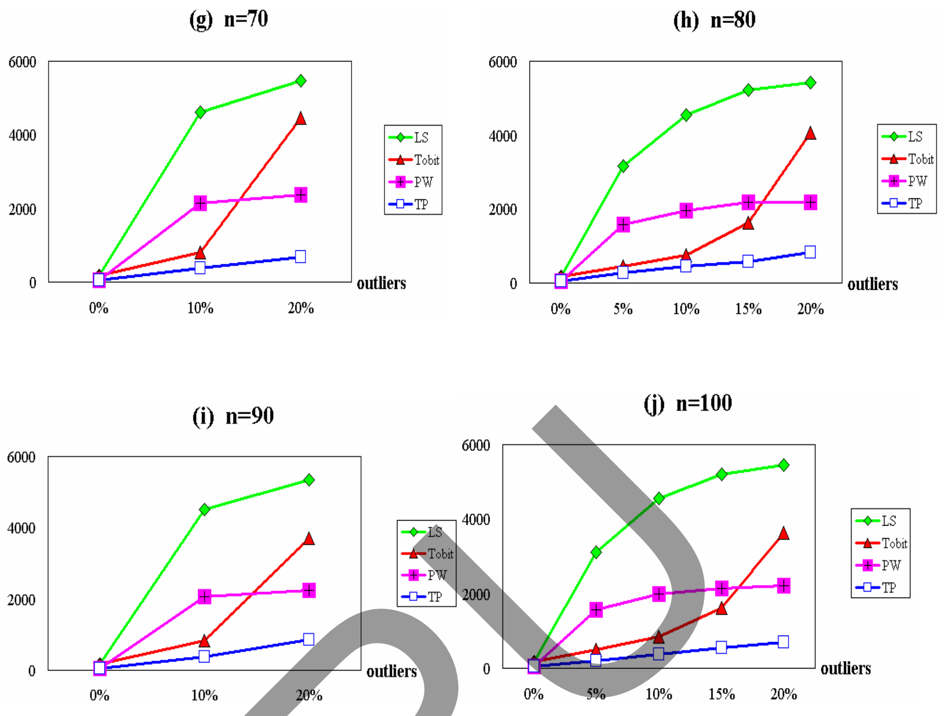


Figure 4.14 ASSR of four different regression models for Quandt's method varied by percentage of outliers when n=10, 20,…, 100 where *xy-direction* outliers exist (continued)

CHAPTER 5

CONCLUSION, DISCUSSION AND RECOMMENDATION

5.1 Conclusion

Throughout this research, the outliers problem is taken into account because these data might affect the hetoroscedasticity problem (Rousseeuw and Leroy, 1987). Therefore, the Gauss-Markov assumptions are violated, and consequently, LS estimator of $\underset{\sim}{\theta}$ for regression coefficient turn out not to be the best linear unbiased estimator (BLUE). This study has focused on the estimation of regression coefficients $\underset{\sim}{\theta}$ in multiple linear regression when sample data contain some usual outliers. Four different methods were first applied to cope with the outliers problem, namely LS, Tobit, piecewise (PW) and Tobit-piecewise (TP) (the construction of TP regression model by the combination of the Tobit and piecewise regression models was first introduced in Mekbunditkul (2010)). The joined point estimation of TP regression model is first interested, leading to an introduction of an estimation for the joined point in TP regression model based on nonlinear least square (LS) such as Levenberg-Marquardt method and on maximum likelihood (ML), for example Quandt's method. From simulation results, it is found that the TP regression model with its joined point estimated by Levenberg-Marquardt method can "down-weigh" the value or reduce the effect of outliers better than other remaining methods. This is followed by TP with its joined point estimated by Quandt's method, PW, Tobit, and LS, respectively.

Results of the analysis of four different regressions on household-income and –expenditure data in socio-economic surveys in Thailand in the year 2009 are presented as below:

1) For Bangkok and Metropolis regions, we obtain four regression models as the followings:

LSE : $\hat{Y} = 23,408.39 + 0.173X$

Tobit : $\hat{Y} = 22,072.24 + 0.187X$

PW : $\hat{Y} = 9,885.90 + 0.539X - 0.528X^*$

TP : $\hat{Y} = 8,115.94 + 0.604X - 0.563X^*$,

where Y represents expenditure data, X is income data and $X^* = (X - 118,213)D$, $D = 1$ if $X \geq 118,213$ and $D = 0$ if $X < 118,213$. The joined point value 118,213 was obtained by Levenberg-Marquardt method. The smallest value of each ASSR and RE is of TP as 134×10^6 and 0.3709, respectively. This means that, in this particular case, TP is the best. In addition, TP regression model with its joined point estimated by Levenberg-Marquardt yields smaller ASSR than with that by Quandt's method. Take into consideration the situation when the value of household income is as zero baht (any household have no income) then the expenditure is about 8,116 and 9,886 baht as predicted by TP and PW models, respectively. Meanwhile, it is 23,408 baht by LS model and is 22,072 baht by Tobit model. We can see that TP and PW yield more feasible results than both LS and Tobit. In addition, there exist the same results for other regions as shown from 2) to 5).

2) For the Central region, the four regression models can be constructed as the followings:

LSE : $\hat{Y} = 13,530.76 + 0.218X$

Tobit : $\hat{Y} = 13,591.15 + 0.212X$

PW : $\hat{Y} = 4,691.43 + 0.619X - 0.638X^*$

TP : $\hat{Y} = 5,582.46 + 0.571X - 0.598X^*$,

where $X^* = (X - 146,221)D$, $D = 1$ if $X \geq 146,221$ and $D = 0$ if $X < 146,221$. The joined point value $146,221$ estimated by Levenberg-Marquardt method makes TP regression model attain the smallest ASSR as 65.30×10^6 and RE as 0.3727. And we also found that TP regression model with its joined point estimated by Quandt's method yields slightly larger value of each ASSR and RE than Levenberg-Marquardt method.

3) For the Northern region, we obtain four different regression models as the followings:

LSE : $\hat{Y} = 6,062.13 + 0.408X$

Tobit : $\hat{Y} = 6,132.13 + 0.402X$

PW : $\hat{Y} = 3,236.19 + 0.582X - 0.626X^*$

TP : $\hat{Y} = 3,419.87 + 0.568X - 0.629X^*$,

where $X^* = (X - 97,281)D$, $D = 1$ if $X \geq 97,281$ and $D = 0$ if $X < 97,281$. The smallest value of each ASSR as 39.96×10^6 and RE as 0.5683 is of TP regression model with its joined point estimated by Levenberg-Marquardt method yields better result than Quandt's method in terms of ASSR and RE.

4) For the Northeastern region, four different regression models can be obtained as the followings:

LSE : $\hat{Y} = 11,423.80 + 0.172X$

Tobit : $\hat{Y} = 11,415.66 + 0.169X$

PW : $\hat{Y} = 3,858.48 + 0.594X - 0.586X^*$

TP : $\hat{Y} = 4,182.82 + 0.568X - 0.571X^*$,

where $X^* = (X - 77,965)D$, $D = 1$ if $X \geq 77,965$ and $D = 0$ if $X < 77,965$. The value $77,965$ came from Levenberg-Marquardt method and this method gave the smallest value of ASSR as 41.49×10^6 and RE as 0.3645.

5) For the Southern region, four different regression models are formed as the followings:

$$\text{LSE} \quad : \quad \hat{Y} = 11,583.95 + 0.269X$$

$$\text{Tobit} \quad : \quad \hat{Y} = 11,829.82 + 0.253X$$

$$\text{PW} \quad : \quad \hat{Y} = 4,888 + 0.586X - 0.555X^{*}$$

$$\text{TP} \quad : \quad \hat{Y} = 5,237.88 + 0.564X - 0.571X^{*},$$

where $X^{*} = (X - 90,790)D$, $D = 1$ if $X \geq 90,790$ and $D = 0$ if $X < 90,790$. The value $90,790$ was obtained by Levenberg-Marquardt method. This method yields better results than Quandt's method compared with Tobit, LS and PW, in senses of ASSR as 63.85×10^{6} and RE as 0.4064.

Therefore, from the results of both simulation study and numerical analysis, we can conclude that TP regression model with its joined point estimated by Levenberg-Marquardt method attains the best among all four different estimation methods including the joined point estimated by Quandt's method. Moreover, the finding in empirical data analysis is that outliers of both *y*- and *x-direction* existing in household expenditure and household income data can have their values down-weighted (reduced effect) by either TP or PW. It is obvious the evidence that PW gives slightly different results from TP. From figures 4.1 – 4.14, it can be seen that both TP and PW can represent the relationship of the bulk of the data more suitably than LS and Tobit models. The results of the analysis on SES data in year 2007 (see also Table 4.4) also support the above conclusion.

5.2 Discussion

1. The TP regression model is first proposed in Mekbunditkul's dissertation as one of alternative models to fit data consisting of outliers. It was derived by the combination of two advantageous principle ideas, namely the Tobit and piecewise regression models. In Tobit regression, putting appropriate limit(s) on dependent variables may down-weigh the effects of outliers. Since the TP model can be considered as one of the factored likelihood functions, the TP estimator was derived by ML method as shown in Chapter 2. Therefore, TP estimator retains good properties of a MLE, e.g. consistency and best asymptotically normal (B.A.N). Further, benefits of piecewise regression can be obtained when structural change in regression is present.

Findings, in this research, indicate that TP regression with the joined point estimated by Levenberg-Marquardt method is most suitable for data with outliers. From simulation results, we can indicate that the ranking of four estimators from the best to the poorest, where the best and poorest are measured by ASSR and RE, as shown in the following table,

<table>
<tr><th colspan="2">Data</th><th>LS</th><th>Tobit</th><th>PW</th><th>TP</th></tr>
<tr><td colspan="2">➢ no outlier</td><td>①</td><td>①</td><td>①</td><td>①</td></tr>
<tr><td colspan="2">➢ y-direction</td><td>④</td><td>③</td><td>②</td><td>①</td></tr>
<tr><td rowspan="2">➢ xy-direction</td><td>Small % of outliers</td><td>④</td><td>②</td><td>③</td><td>①</td></tr>
<tr><td>Large % of outliers</td><td>④</td><td>③</td><td>②</td><td>①</td></tr>
<tr><td colspan="2">➢ x-direction</td><td>②</td><td>②</td><td>①</td><td>①</td></tr>
</table>

2. In addition, another findings show that the Levenberg-Marquardt method does not significantly differ from Quandt's method regarding the results from both simulation study and numerical data analysis, namely household income and

household expenditure in socio-economic survey data. Nevertheless, Levenberg-Marquardt method yields slightly better results than Quandt's method in terms of smaller ASSR and RE. The existence of this insignificant difference might be caused by two reasons: (1) the application in this study is considered only in case of a simple regression. That is one explanatory variable is chosen for done simulation and numerical study, (2) Levenberg-Marquardt method, in this research, is utilized for analyzing data described by model which is linear in parameters, such as TP regression model.

5.3 Recommendation

There are three recommendations for further study arising from this research. First, estimators of the desired limited value in TP regression model have not been included under the scope of this study but it could be conducted 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 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. Third, the factored likelihood, as shown particularly in model (2.10), is still powerful for certain cases such as the Tobit and TP models. Thus, the estimator constructed by the factored likelihood method could be utilized for other situations because it is a good estimator and retains all of the properties of MLE, such as consistency and best asymptotically normal (B.A.N).

BIOGRAPHY

NAME	Titirut Thipbharos, Ph.D.
ACADEMIC BACKGROUND	B.Sc. (Statistics) Khon Kaen University, Khon Kaen , Thailand, 2001. M.S.(Statistics), Chulalongkorn University, Bangkok, Thailand, 2003. Ph.D. (Statistics), National Institute of Development Administration, 2010.
PRESENT POSITION	Lecturer, Department of Mathematics and Statistics, Faculty of Arts and Sciences, Dhurakij Pundit University Bangkok, Thailand.
RESEARCHS :	1. PARMETERS-ESTIMATION OF POLYNOMIAL REGRESSION MODELS WITH ERRORS IN INDEPENDENT VARIABLES, 2003 2. Burden of Household Expenditures for Education , 2005 3. AN ALTERNATIVE ESTIMATOR FOR REGRESSION COEFFICIENTS WITH OUTLIERS, 2010

PUBLICATIONS:

1. Mekbunditkul, M. and Siripanich, P. An Alternative Model used for Linear Relations with Outliers. **Proceedings of the 11th Conference on Statistics and Applied Statistics,** May 28-29, 2010. Chiangmai, Thailand.

2. Mekbunditkul, M. and Siripanich, P. Tobit-Piecewise Estimator of the Regression Coefficient. **Proceedings of the 4th International Conference of Operations and Supply Chain Management**, July 25 – 27, 2010. Shatin, Hong Kong.

3. Mekbunditkul, M. and Siripanich, P. Maximum Likelihood Estimation of Parameters in Tobit-Piecewise Regression Model. **Proceedings of the 6th IMT-GT International Conference on Mathematics, Statistics, and its Applications**, Nov. 3-4, 2010. Kuala Lumpur, Malaysia.

REFERENCES

Amphanthong, P. 2009. Estimation of Regression Coefficients with Outliers. Dissertation. Bangkok: National Institute of Development Administration.

Barnett, V. and Lewis, T. 1994. **Outliers in Statistical Data**. John Wiley & Sons. Available at http://en.wikipedia.org/wiki/Outlier.

Gill, P.E. 1991. **Numerical Linear algebra and Optimizatin**. Adison-Wesley.

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).

Greene, W.H. 1981. On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model. **The Econometric Society**. 42(March): 505-513.

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.

Moronna, R.A., Martin, R.D. and Yohai, U.J. 2006 . **Robust Statistics: Theory and. Methods.** West Sussex: John Wiley & Sons.

Mekbunditkul, T. 2010. An Alternative Estimation for Regression Coefficients with Outliers. Dissertation. Bangkok: National Institute of Development Administration.

Monhor, D. and Takemoto, S. 2005. Understanding the concept of outlier and its relevance to the assessment of data quality: Probabilistic background theory. **Earth Planet Space.** 57 : 1009-1018.

Montgomery, D.C. and Peck, E.A. 1982. **Introduction to Linear Regression Analysis.** New York : John Wiley & Sons.

Moore, D.S. et al. 1999. **Introduction to the Practice of Statistics.** New York : W.H.Freeman.

Poon, K.S., Fan, G., Poon, J., and Young, R. 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.

Rousseew, P.J. and Zomeren, B.C. 1990. Unmasking multivariate outliers and leverage points. **Journal of the American Statistical Association**. 85 : 633-651.

Ryan, T.P. 1997. **Modern Regression Methods**. New York: John Wiley & Sons.

Seber, G.A.F. and Wild, C.J. 2003. **Nonlinear Regression**. New York: John Wiley & Sons.

Snedecor, G.W. and Cochran, W.G. 1989. **Statistical Methods**. 8th ed. Iowa: Blackwell.

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

Tishler, A., and Zang, I. 1981. A Maximum Likelihood Method for Piecewise Regression Models with a Continuous Dependent Variable . **Applied Statistics**. 30(April): 116-124.

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

Walfish, S. 2006. A Review of Statistical Outlier Methods. **Pharmaceutical Technology.** Available at http://www.pharmtech.com/pharmtech/content/print/content/printContentPopup.jsp?id=384716