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.

DPU

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.

DPU

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 $\underset{\sim}{x}_i = (x_{i1}, x_{i2}, ..., x_{ik})$, 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. $\hat{\beta}_1$, $\hat{\beta}_2$, ..., $\hat{\beta}_k$ will have $\underset{\hat{\underset{\sim}{\theta}}}{\text{Min}} \sum_{i=1}^{n} r_i^2 = \underset{\hat{\underset{\sim}{\theta}}}{\text{Min}} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$. 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

DPU

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

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 $f_Y\left(Y_i\right) = \Phi\left(\frac{L - \underset{\sim}{x}_i\theta}{\sigma_i}\right)$ if $Y_i = L_i$, $f_Y\left(Y_i\right) = \frac{1}{\sigma_i}\phi\left(\frac{Y_i - \underset{\sim}{x}_i\theta}{\sigma_i}\right)$ if $Y_i = Y_i^*$, and $f_Y\left(Y_i\right) = 1 - \Phi\left(\frac{U_i - \underset{\sim}{x}_i\theta}{\sigma_i}\right)$ 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.

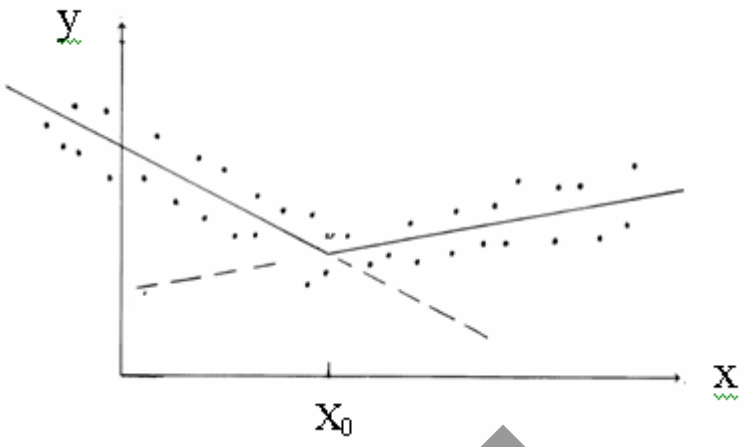
DPU

2.4 Piecewise Regression Model

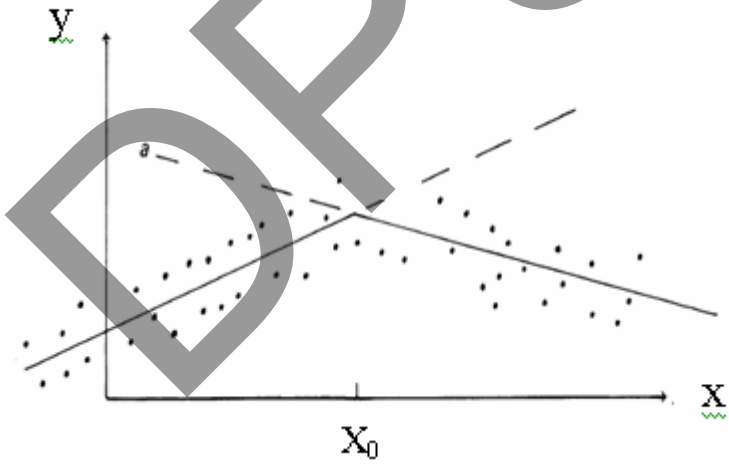
In this research, the structural change in the regression model is taken into account, thus the piecewise regression model (Quandt, 1958: 874, Hudson, 1966: 1097-1129, Goldfeld, Kelejian and Quandt, 1971, Suits, Mason and Chan 1978: 132-133) is considered. Quandt first introduced that economic variables 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.

DPU



(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.

DPU

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 $D_i = \begin{cases} 0 & ; x_i \le \upsilon \\ 1 & ; x_i > \upsilon \end{cases}$.

DPU

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

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

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

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

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

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

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

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

DPU

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

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

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 $h\left(\underset{\sim}{\theta}\right)$, a real-valued function of p parameters vector $\underset{\sim}{\theta} = \left(\theta_1, \theta_2, ..., \theta_p\right)'$, 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 $h\left(\underset{\sim}{\theta}\right)$ is defined as $S\left(\underset{\sim}{\theta}\right) = \sum_{i=1}^{n}\left(Y_i - f\left(\underset{\sim}{x}_i; \underset{\sim}{\theta}\right)\right)^2$.

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 a^{th} step, is differently defined as

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

DPU

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

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

DPU

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)

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

where $D^{(a)}$ is a diagonal matrix with positive diagonal element frequently defined to be the same as $\underset{\cdot}{F}^{(a)'} \underset{\cdot}{F}^{(a)}$ 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 $\underset{\cdot}{F}^{(a)'} \underset{\cdot}{F}^{(a)} + \eta^{(a)} D^{(a)}$ 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

DPU

$\eta^{(a)} \to \infty$ then $\underset{\sim}{\delta}^{(a)}$ tend to zero. Therefore if we select very large $\eta^{(a)}$ then we can so fast reduce the function $S(\underset{\sim}{\theta}) = \left\| \underset{\sim}{Y} - f(\underset{\sim}{x}_i; \underset{\sim}{\theta}) \right\|^2$. 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

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

where Y_i^* is a dependent variable, x_{ij} represents the i^{th} observation of the j^{th} independent variable, for j=1, ..., k and i=1, ..., n . In addition, $\upsilon_i = \underset{\sim}{x}_i \underset{\sim}{\vartheta}$ (Quandt,

DPU

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 $\underset{\sim}{x}_0 = \left(x_{01},...,x_{0k}\right)$ 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

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

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

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

DPU

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

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

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

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

DPU

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

Therefore, the inverse matrix of Σ is $\Sigma^{-1} = \begin{bmatrix} \Sigma_1^{-1} & 0 & 0 \\ 0 & \Sigma_2^{-1} & 0 \\ 0 & 0 & \Sigma_3^{-1} \end{bmatrix}$.

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

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

DPU

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 $i = 1, ..., n_m$ 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 $i = 1, ..., n_1$:

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

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

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

DPU

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

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

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 $i = 1, \ ..., \ n_3$:

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

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

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

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

DPU

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

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

Thus, the log-likelihood function is thus given by

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

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

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

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

DPU

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

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

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

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

where $\hat{\lambda}^{-}_{n_1 1}$ and $\hat{\lambda}_{i3}$ are estimators of $\lambda^{-}_{n_1 1}$ 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 $\underset{\sim}{Y} = \begin{bmatrix} \underset{\sim}{L}_1 \\ \cdots \\ \underset{\sim}{Y}_2 \end{bmatrix}_{n\times 1}$, where $n = n_1 + n_2$.

DPU

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

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

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

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

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)

DPU

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

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

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)

$\left(X'WX \right)^{-1} = \left\{ \left(X_3' W_3 X_3 \right) + \left(X_2' W_2 X_2 \right) \right\}^{-1}$, 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.

DPU

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)

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

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

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

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.

DPU