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

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

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

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

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

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

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

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

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

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

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

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 $\underset{\sim}{\theta} = \left(\alpha_1, \beta_1, \beta_2\right)'$ 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

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

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

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

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

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

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

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.