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:

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

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

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

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

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,

Data		**LS**	**Tobit**	**PW**	**TP**
➢ no outlier		①	①	①	①
➢ *y-direction*		④	③	②	①
➢ *xy-direction*	Small % of outliers	④	②	③	①
	Large % of outliers	④	③	②	①
➢ *x-direction*		②	②	①	①

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).