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.

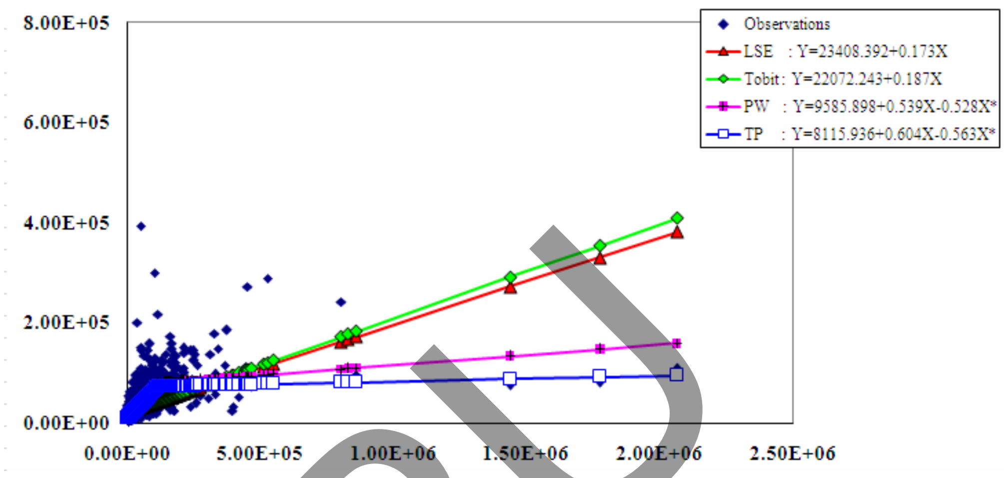

Figure 4.1 Observation and Four Regression Lines for Household-Expenditure and -Income Data for Bangkok Metropolis region on SES in year 2009

Source of Data : National Statistical Office

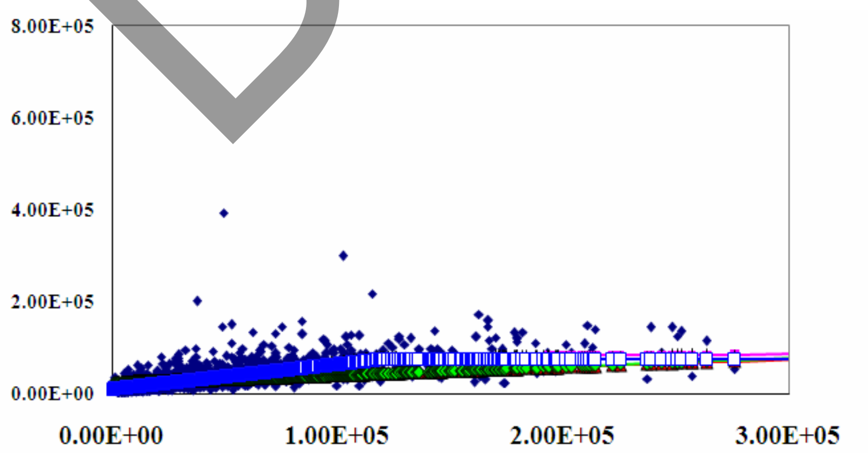

Figure 4.2 The Expansion of Figure 4.1 for the Range of Household Income as being between 0 and 300,000 Baht

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.

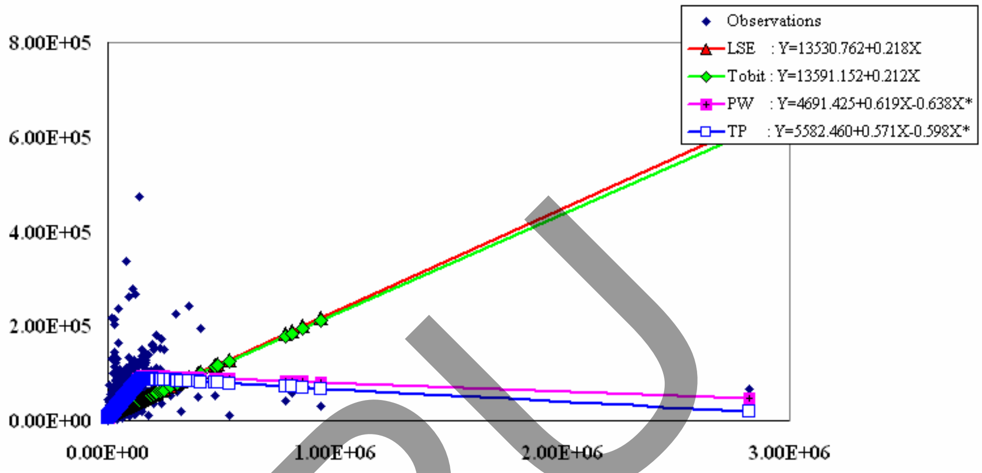

Figure 4.3 Observation and Four Regression Lines for Household-Expenditure and -Income Data for Central region on SES in year 2009

Source of Data : National Statistical Office

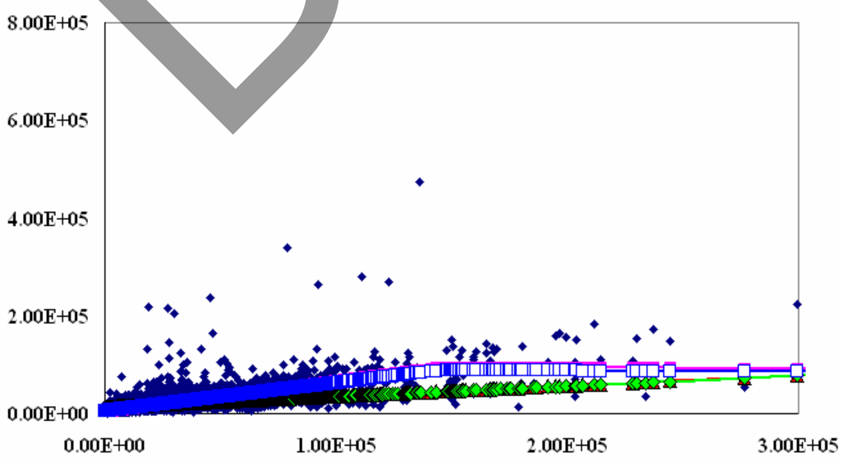

Figure 4.4 The Expansion of Figure 4.3 for the Range of Household Income as being between 0 and 300,000 Baht

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.

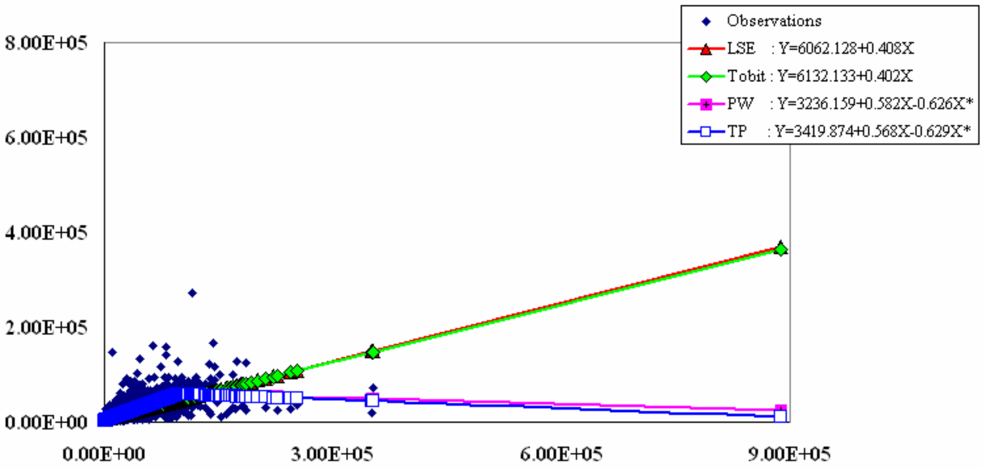

Figure 4.5 Observation and four Regression Lines for Household-Expenditure and -Income Data for North region on SES in year 2009

Source of Data : National Statistical Office

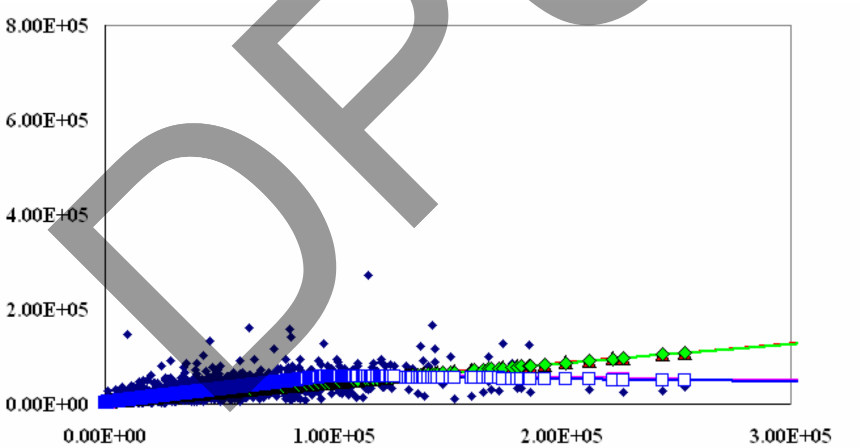

Figure 4.6 The Expansion of Figure 4.5 for the Range of Household Income as being between 0 and 300,000 Baht

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.

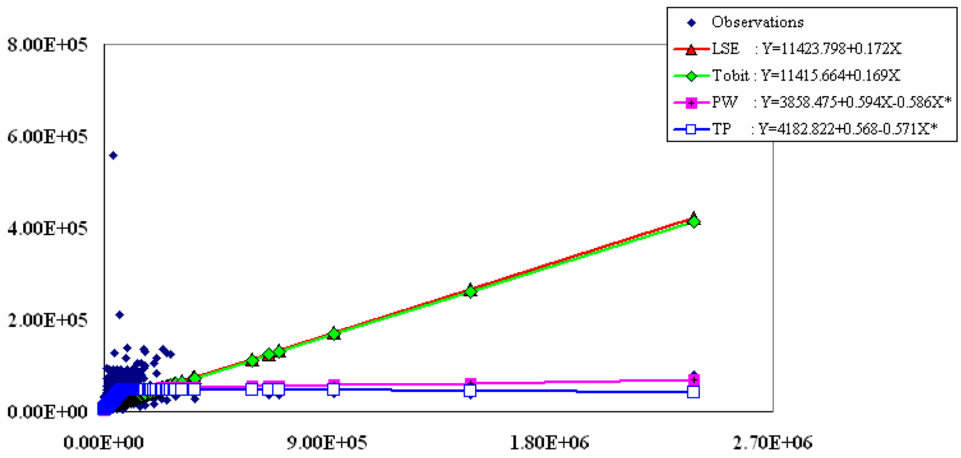

Figure 4.7 Observation and four Regression Lines for Household-Expenditure and -Income Data for Northeast region on SES in year 2009

Source of Data : National Statistical Office

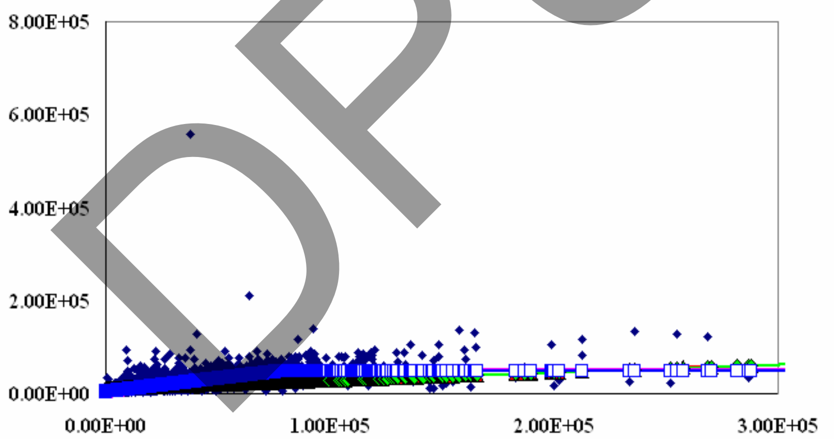

Figure 4.8 The Expansion of Figure 4.7 for the Range of Household Income as being between 0 and 300,000 Baht

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.

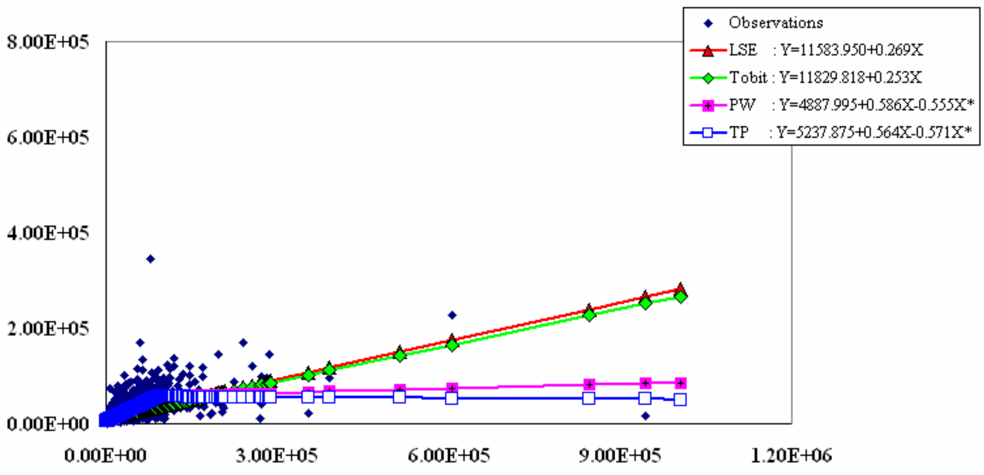

Figure 4.9 Observation and four Regression Lines for Household-Expenditure and -Income Data for South region on SES in year 2009

Source of Data : National Statistical Office

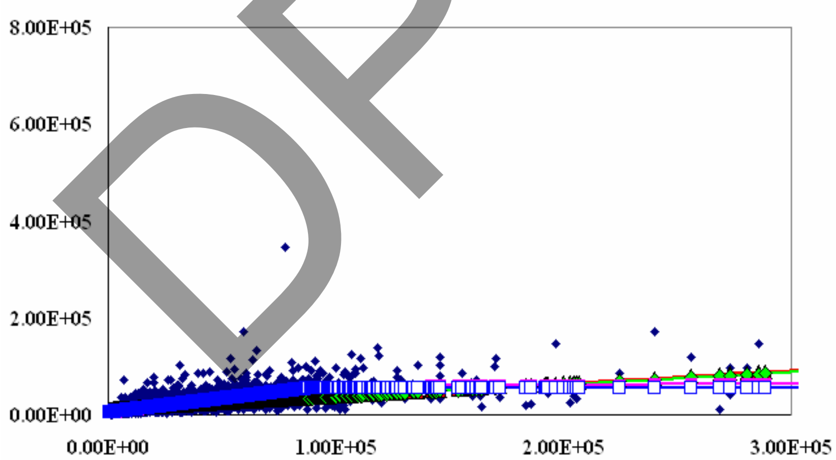

Figure 4.10 The Expansion of Figure 4.7 for the Range of Household Income as being between 0 and 300,000 Baht

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, …, $\left\lfloor \frac{(1-\alpha)n}{2} \right\rfloor + 1$, 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

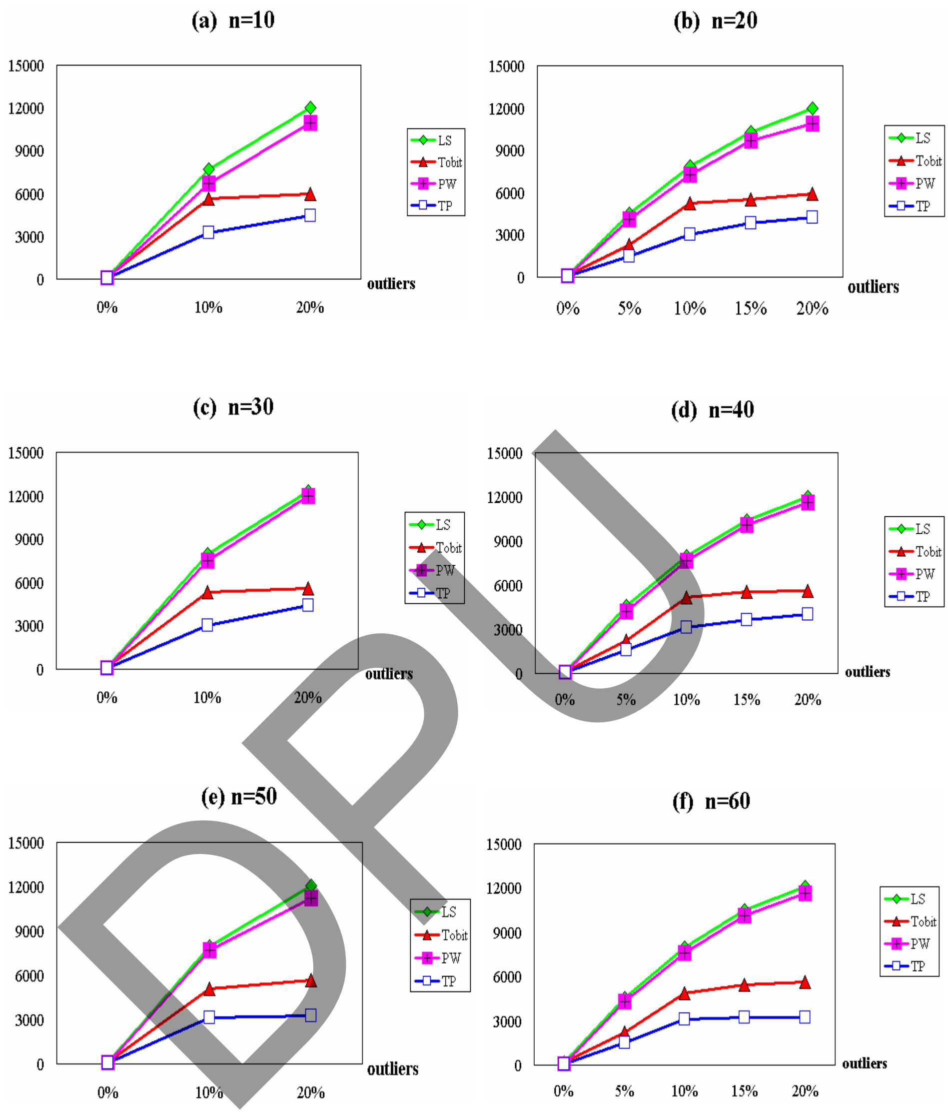

Figure 4.11 ASSR of four different regression models for Levenberg-Marquardt method varied by percentage of outliers when n=10, 20,…, 100 where *y-direction* outliers exist

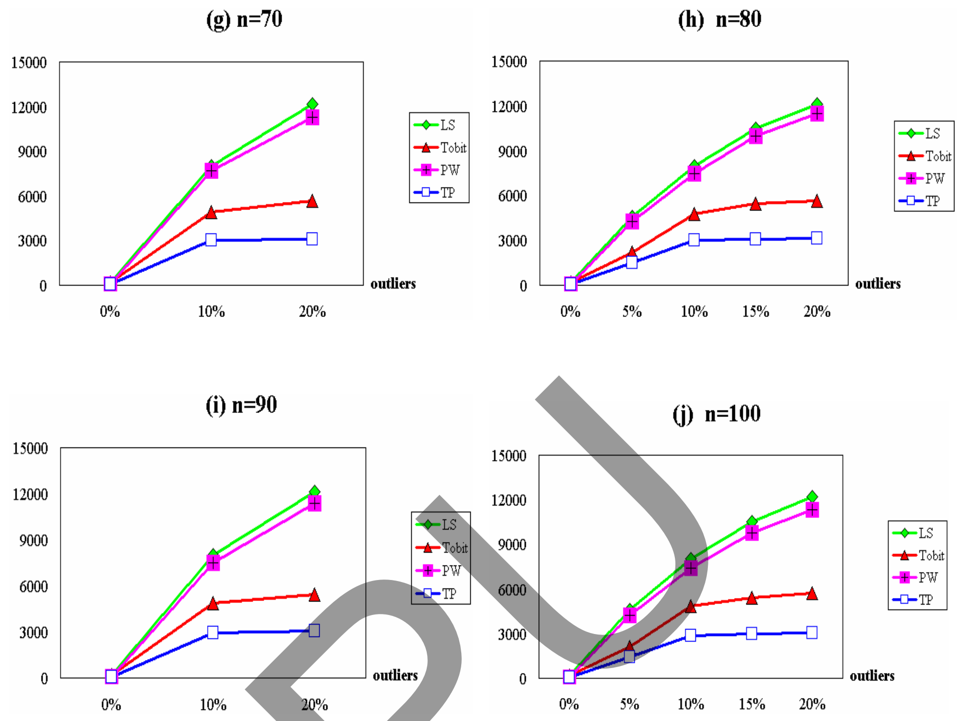

Figure 4.11 ASSR of four different regression models for Levenberg-Marquardt method varied by percentage of outliers when n=10, 20,…, 100 where *y-direction* outliers exist (continued)

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	ASSR Tobit	ASSR PW	ASSR 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.

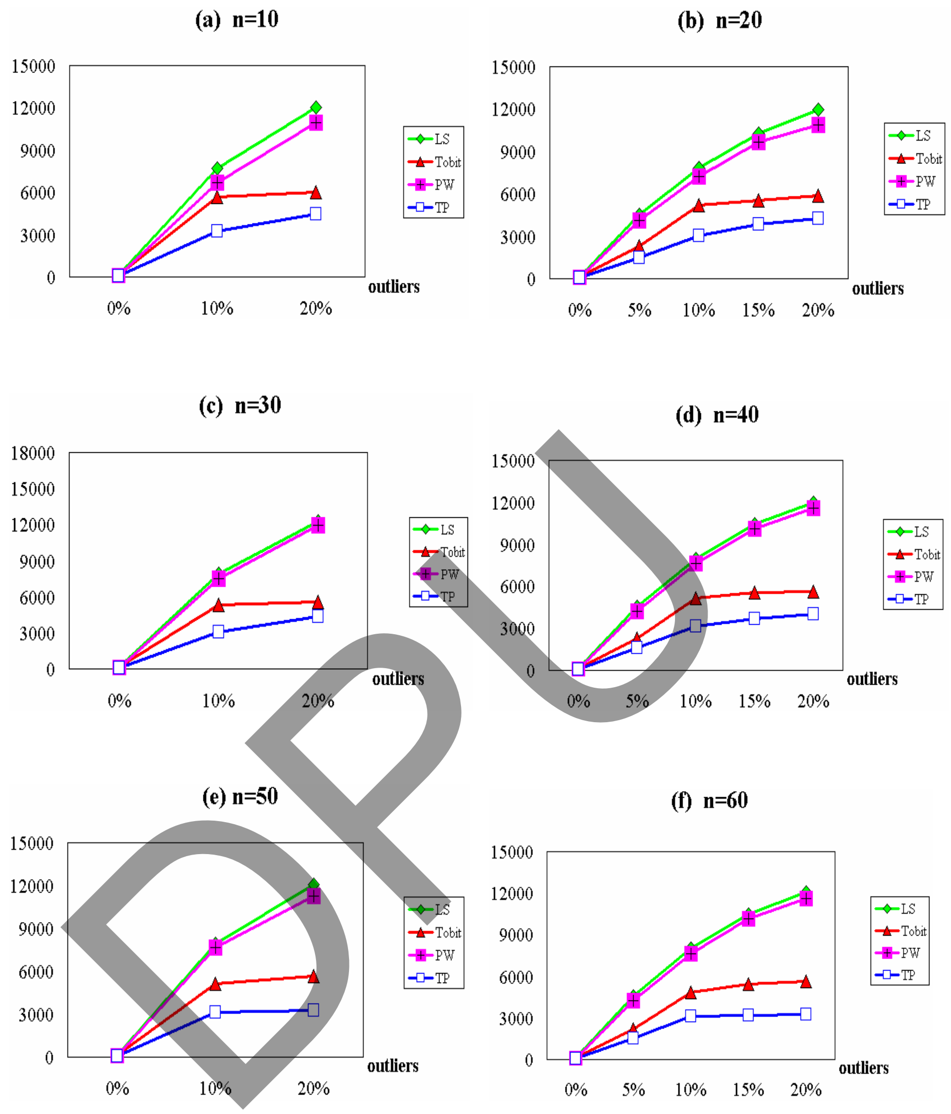

Figure 4.12 ASSR of four different Regression Models for Quandt's method varied by percentage of Outliers when n=10, 20,…, 100 where *y-direction* Outliers exist

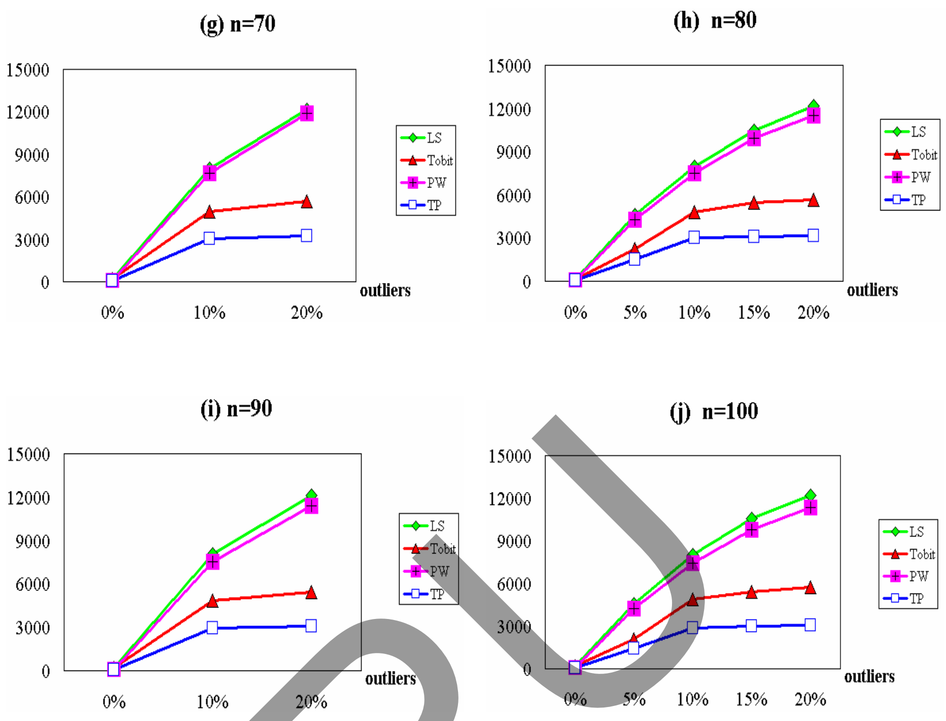

Figure 4.12 ASSR of four different Regression Models for Quandt's method varied by percentage of Outliers when n=10, 20,…, 100 where *y-direction* Outliers exist (continued)

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

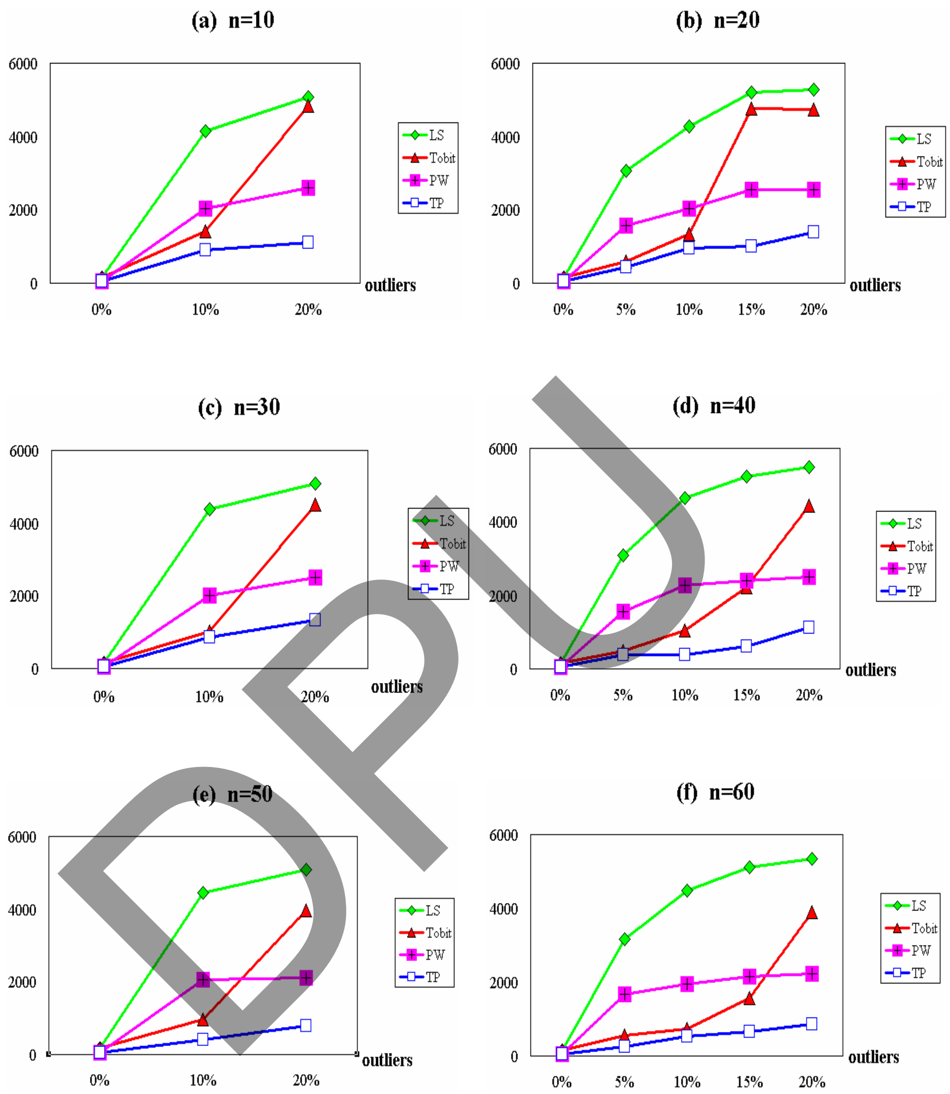

Figure 4.13 ASSR of four different regression models for Levenberg-Marquardt method varied by percentage of outliers when n=10, 20,…, 100 where *xy-direction* outliers exist

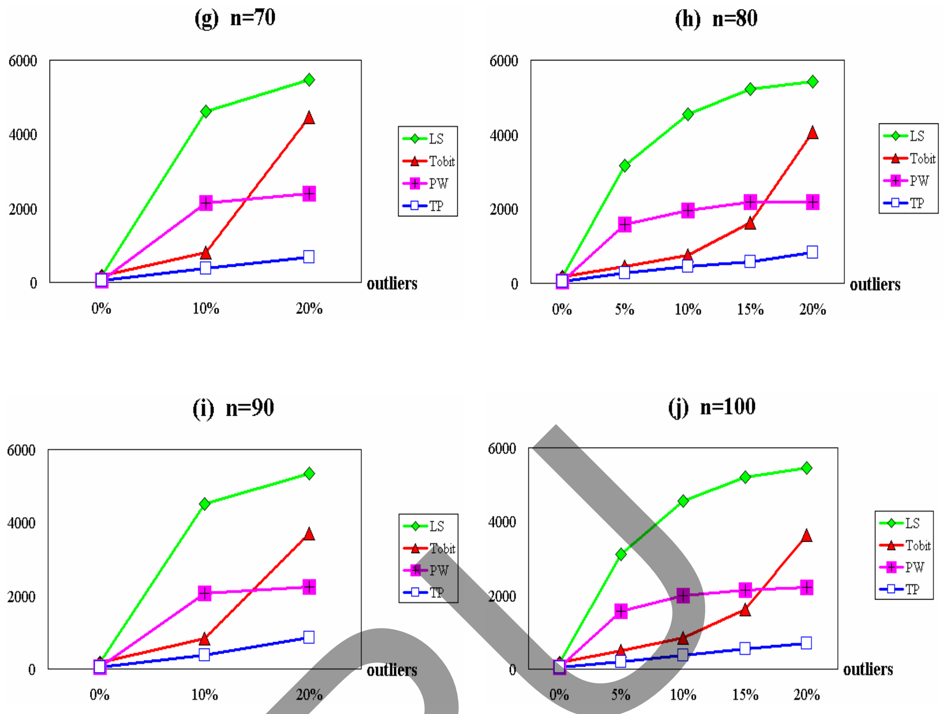

Figure 4.13 ASSR of four different regression models for Levenberg-Marquardt method varied by percentage of outliers when n=10, 20,…, 100 where *xy-direction* outliers exist (continued)

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	% of Y-	ASSR			
Size	**Outliers**	**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

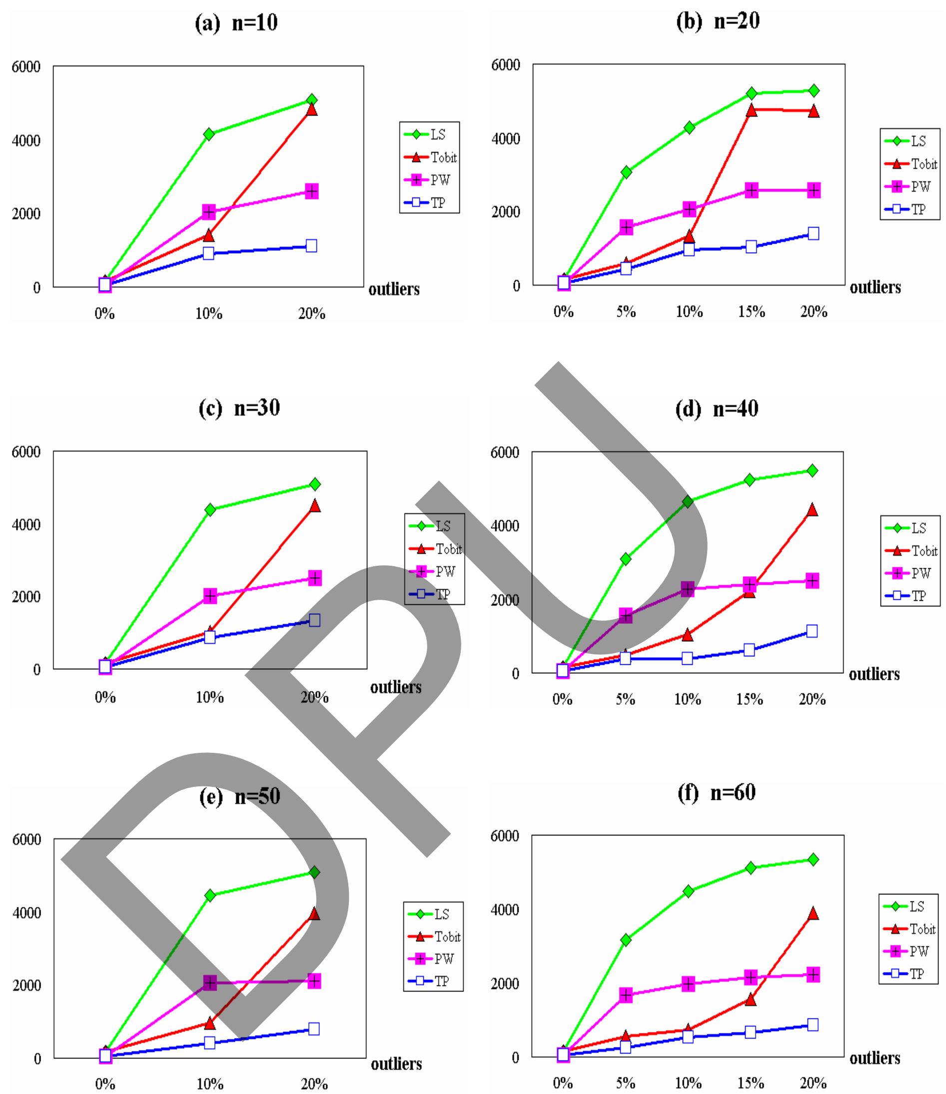

Figure 4.14 ASSR of four different regression models for Quandt's method varied by percentage of outliers when n=10, 20,…, 100 where *xy-direction* outliers exist

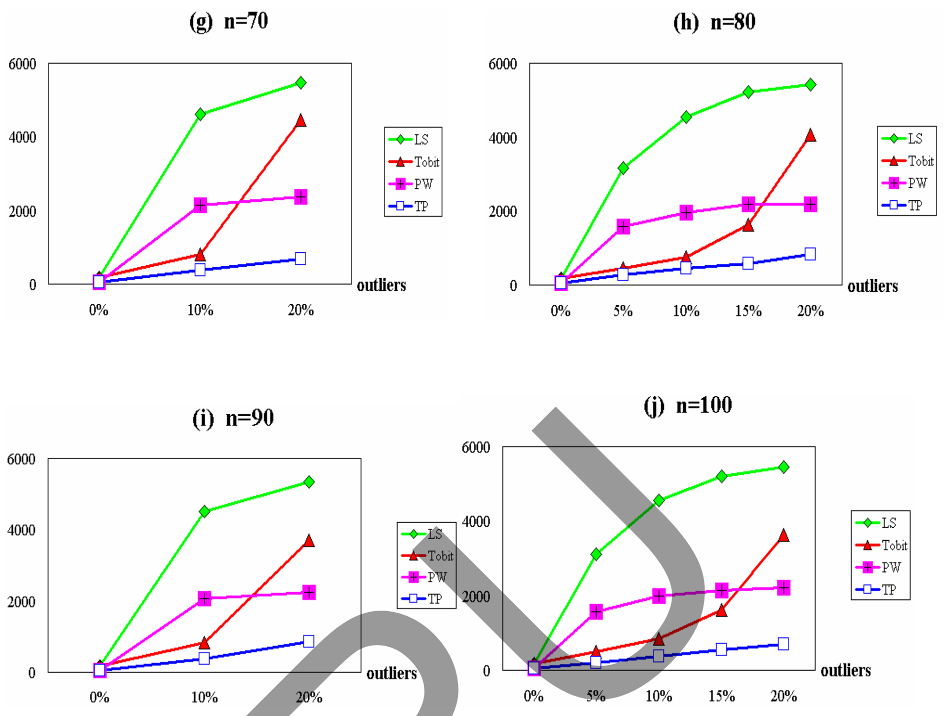

Figure 4.14 ASSR of four different regression models for Quandt's method varied by percentage of outliers when n=10, 20,…, 100 where *xy-direction* outliers exist (continued)