New Approaches in Applied Statistics
Anuška Ferligoj and Andrej Mrvar (Editors)
Metodološki zvezki, 16, Ljubljana: FDV, 2000

Some Aspects of Differences between L_1 and L_2 Criteria in the Linear Switching Regression

Polona Tominc[1] and Lada Bele Tominc[2]

Abstract

The least squares procedure or L_2 criterion is in theory and in practice generally used to estimate the regression coefficients. It is well known that given the assumptions of the classical linear regression model the least squares estimates posses some ideal properties. One of the assumptions underlying the L_2 criterion is that the disturbance terms are normally distributed. But there are many cases where the disturbance terms are not normally distributed. Therefore, the use of some other criteria could be legitimate. As reported in the literature (for example Narula and Korhonen, 1994) the least absolute value or L_1 criterion is less sensitive to outliers than the L_2 criterion.

With the purpose to illustrate some aspects of differences between L_2 and L_1 criteria in the presence of switching regression function with a priori known switch, the Monte Carlo simulation was performed.

The least absolute value criterion has another advantage, especially in the cases, where the switch is not known in advance. Using the least absolute value criterion the estimation problem can be formulated and solved as a linear mixed integer optimisation model. If the switch is known in advance the optimisation model is linear.

1 Introduction

The main idea for this paper arises from the problem of the piecewise linear sales-response function with the disturbance terms that are not normally distributed, and which is to be included into the linear mixed integer optimisation model of the multiphase business process.

[1] University of Maribor, Faculty of Business and Economics Maribor, Razlagova 14, Maribor, Slovenia.

[2] The Secondary School of Textile, Pariške komune 34, Maribor, Slovenia.

The optimisation model is linear mixed integer if the objective function and all the restrictions are linear or piecewise linear functions. The linear mixed integer optimisation model of the multiphase business process offers a very good approximation of almost every important business relationships and it can also be treated by the available computer software very quickly and costless (Mesko, 1994).

In most practical cases the linear mixed integer optimisation model is used with the purpose to establish the optimal purchasing quantities of all inputs and selling quantities of all outputs (products and services) of the business process so that the objective function - the difference between the income and costs (variable and semi-fixed costs) - is maximal. If none of the regressors in the piecewise linear sales-response function is the decision variable in the optimisation model, the sales response function can be estimated separately and the estimated maximal selling quantity of the product is included into the optimisation model as a constant. But the piecewise linear sales-response function and the business optimisation model can not be treated separately, if the maximal selling quantity of a product is not known in advance and it depends to the value of other decision variables of the optimisation model (for example advertising expenditure). Therefore it is important to have the possibility to include the estimation of the regression coefficients and of the switch of the piecewise sales-response function directly into the linear mixed integer optimisation model of the multiphase business process. The problem of the estimation of the regression coefficients and of the switch of the linear piecewise regression function can be formulated as a linear mixed integer optimisation model, if the least absolute value criterion is taken into account (Mesko, 1989). It can be included into the linear mixed integer optimisation model of the multiphase business process.

The least squares procedure or L_2 criterion is in theory and in practice generally used to estimate the regression coefficients of the linear model as well as of the linear switching regression model with deterministic and a priori known switch (Goldfeld and Quandt, 1976). This is because of its analytical tractability, its highly developed theory and widespread literature. Given the assumptions of the classical linear regression model, it is well known that the least squares estimates posses some ideal or optimum properties (Gujarati, 1995: 72): they are best linear unbiased estimates (BLUE). The parameter estimates follow a normal distribution where the mean values are equal to the real values of the regression coefficients. The variance-covariance matrix can be obtained from $\sigma^2(\boldsymbol{X'X})$-1, where σ^2 is the homoscedastic variance of the disturbances (also of the responses) and $\boldsymbol{X}$ is the matrix of K explanatory variables and T observations (Gujarati, 1995: 290). One of the assumptions underlying the L_2 criterion is that the disturbance terms are normally distributed. But there are many cases, and the piecewise linear sales-response function is often among them, where the disturbance terms are not normally distributed.

2 Linear regression function and the L_1 criterion

In the literature, the L_1 criterion in the linear regression is involved especially in the presence of large disturbances (outliers) and when the classical assumption of a normal error distribution is violated due to "contamination " or "heavy tails". Let us mention here some of the facts about the use of L_1 estimators, found in the published literature:

(1) There are several researches about the properties of L_1 estimators if small samples are taken into account. Let us mention a very extensive simulation study, which was performed by Rosenberg and Carlson, who investigated the L_1 estimators in multiple regression models with symmetric error distributions, specifically normal and contaminated normal error distributions (Rosenberg and Carlson, 1977). A contaminated normal distribution is a residual distribution where the majority of the disturbances are taken from a normal distribution with constant variance except for one or more outliers from a normal distribution with a much larger variance (Dielman and Pfaffenberger, 1982). After performing over 100.000 L_1 regressions, they reported the following results:

   a) The L_1 estimates had a significantly smaller standard error than the least-squares estimates for a regression with high-kurtosis disturbances.
   b) The L_1 estimators were almost exactly normally distributed in the presence of high-kurtosis disturbances.
   c) The error (the difference between the estimated and a real value of the regression coefficient) was approximately normally distributed with mean zero and covariance matrix $\lambda^2(\boldsymbol{X'X})^{-1}$, where $\boldsymbol{X}$ is the matrix of K explanatory variables and T observations, and λ^2/T is the variance of the median of a sample of size T from the disturbance distribution.

(2) The most important result is the asymptotic theory for L_1 estimators. The paper of Basset and Koenker (1978) confirms the assumed hypothesis, that for any error distribution for which the median is (asymptotically) superior to the mean as an estimator of location, the L_1 estimator of the regression coefficients are preferred to the L_2 estimators in the sense of having strictly smaller asymptotic confidence ellipsoids for the regression coefficients for a fixed sample size. The asymptotic variance of the median is smaller than of the mean for example for Cauchy, Laplace and Logistic distribution. Basset and Koenker prove the result (c) cited above holds true asymptotically.

(3) Charnes, Copper, and Ferguson (1955) were the first who demonstrated that L_1 estimates could be produced by linear programming. Nowadays it is very easy to obtain the L_1 estimates with the existing computer software. For the estimation of regression coefficients of the piecewise linear regression function it is possible to form the linear mixed integer optimisation model, as mentioned before (Mesko, 1989).

Although the use of L_1 criterion when estimating the piecewise linear regression functions can not be found in the literature it is assumed that the above results could hold true also for the piecewise linear regression function. Especially in the cases, where the switch is known in advance, since the piecewise linear regression function can be written with the linear function by the use of 0-1 variables. With the purpose to illustrate some aspects of differences between the L_1 and L_2 estimators (having in mind the BLUE properties of L_2 estimates under classical conditions) the Monte Carlo simulation was performed. Among several possibilities for error distribution which could appear, the "contaminated" normal distribution for disturbance terms was taken into account.

3 The Monte Carlo simulation

The following Monte Carlo experiment was conducted:

(1) The population piecewise linear function with deterministic and a priori known switch was chosen for the experiment

$$\begin{aligned} y_1^{(k)} &= \alpha_0 + \alpha_1 x^{(k)} + u^{(k)} \qquad k \leq 40 \\ y_1^{(k)} &= \beta_0 + \beta_1 x^{(k)} + u^{(k)} \qquad k > 40 \end{aligned} \tag{3.1}$$

where k items, k=1,2,...,100, are observed, $y_1^{(40)} = y_2^{(40)}$, and with

$$\alpha_0 = 0 \qquad \alpha_1 = 1 \qquad \beta_0 = -40 \qquad \beta_1 = 2$$

The values of the explanatory variable were selected: $x^{(k)} = k$, k=1,2,...,100. The model (3.1) can be written as

$$y^{(k)} = \alpha_0 + \alpha_1 x^{(k)} + \delta_1 (x^{(k)} - 40) U + u^{(k)} \tag{3.2}$$

where U is 0-1 variable

$$\begin{aligned} U = 0 \text{ for} \qquad & x^{(k)} \leq 40 \\ U = 1 \text{ for} \qquad & x^{(k)} > 40 \end{aligned}$$

with the following relationships between the regression coefficients

$$\alpha_0 + \delta_0 = \beta_0 \qquad \alpha_1 + \delta_1 = \beta_1 \qquad \delta_0 = -40\delta_1 \tag{3.3}$$

Therefore

$\delta_1 = 1$ and $\delta_0 = -40$

(2) The disturbances were generated:

a) normal disturbances N(0,1)

b) "contaminated" normal disturbances; the majority of the disturbances were taken from a normal distribution (0, 1) and five randomly selected items (out of k=100 items) – the outliers, were taken from a normal distribution with a much larger variance (0, 100).

(3) Let us by a_0, a_1, b_0, b_1, c_0 and c_1 denote the estimates of the regression coefficients α_0,
α_1, α_1, β_0, β_1, δ_0 and δ_1. Using the equation

$$y^{(k)} = x^{(k)} + (x^{(k)} - 40)U + u^{(k)}$$

n = 100 responses ($y^{(k)}$) were calculated, for each case, (a) in (b), determined in step (2).

The regression coefficients in the model

$$y^{(k)} = a_0 + a_1 x^{(k)} + c_1(x^{(k)} - 40)U \tag{3.4}$$

are to be estimated.

(4) Both L_2 and L_1 estimation techniques were applied to the resulting data.

(5) Steps (1)-(4) of the simulation has been repeated for 10.000 samples with the computer programme TSP (TSP, 1993).

Results of the simulation – the mean values and standard deviations for a_0, a_1 and c_1 are shown in Tables 3.1 and 3.2.

The comparison is based on the comparison of the standard errors. The results show that the least absolute value estimates of the regression coefficients have smaller standard errors than the least squares estimates, when the switching regression model with “contaminated” normal error distribution is estimated. Also the mean values of distributions differ less from the real values of regression coefficients (written in brackets) when L_1 criterion is used.

Table 3.1: Results when the normal disturbances were applied.

	Least squares		
	a_0 (0)	a_1 (1)	c_1 (1)
Mean	0.002725	0.999903	1.00013
Standard deviation	0.293657	0.010227	0.014708
	Least absolute deviations		
	a_0 (0)	a_1 (1)	c_1 (1)
Mean	–0.002883	1.00001	1.00010
Standard deviation	0.73017	0.012871	0.018410

Table 3.2: Results when the "contaminated" normal disturbances were applied.

	Least squares		
	a_0 (0)	a_1 (1)	c_1 (1)
Mean	1.53275	0.977529	1.13344
Standard deviation	8.14543	0.221069	0.269247
	Least absolute deviations		
	a_0 (0)	a_1 (1)	c_1 (1)
Mean	0.014526	0.999884	1.00095
Standard deviation	0.398765	0.013569	0.019303

However, in the case with normal error distribution, as would be expected, the least squares estimates have smaller standard errors. As mentioned before, under classical assumptions of the regression model, the parameter estimates follow a normal distribution where the mean values are equal to the real values of the regression coefficients. The variance-covariance matrix can be obtained from $\sigma^2(\boldsymbol{X'X})$-1, where σ^2 is the homoscedastic variance of the disturbances and $\boldsymbol{X}$ is the matrix of K explanatory variables and T observations, as mentioned before.

4 Piecewise sales-response function

The estimations of the piecewise linear sales-response function is presented in this section – an example of the sales-response function of the washing powder produced by the Henkel-Zlatorog enterprise (Tominc) denoted by the letter A for the time from July 1993 to March 1996 (monthly data). The function relationship

$$y_A = f(p_A, p_D, p_F, I, t)$$

where

$y_A(t)$ - selling quantity of the washing powder A in time t
$p_A(t)$ - price (deflationed) of the washing powder A in time t
$p_D(t)$ - price (deflationed) of the washing powder D in time t
$p_F(t)$ - price (deflationed) of the washing powder F in time t
$I(t)$ - average gross wages (deflationed) in time t
t - time

and the piecewise log-linear sales-response function with two linear pieces is chosen. The data used are in the appendix. The data of the prices (average monthly prices) of washing powders A, D and F were bought from marketing office ITEO by the enterprise Henkel Zlatorog Maribor. The bases of ITEO report are data obtained from the sample of 204 Slovene supermarkets. ITEO office is the most important source of this kind of data in Slovenia for enterprise Henkel Zlatorog.

This example is presented here with the purpose to illustrate some aspects of differences between L_1 and L_2 criteria, therefore the question of choosing appropriate explanatory variables is not discussed here. Also let us assume, that the regression function is correctly specified.

The washing powder F is manufactured by Procter & Gamble enterprise. The washing powder D is also produced by Henkel Zlatorog enterprise and represents the dominant competitor of the washing powder A in Slovenian market.

The log-linear piecewise linear sales-response function

$$y_1 = \ln y_A = \ln a_0 + a_1 \ln r + a_2 \ln r_1 + a_3 \ln I + a_4 t \qquad \ln r_1 \leq 0{,}1271$$

$$y2 = \ln y_A = \ln b_0 + b_1 \ln r + b_2 \ln r_1 + b_3 \ln I + b_4 t \qquad \ln r_1 > 0{,}1271 \tag{4.1}$$

was used, where

r- relative price $p_A(t)/p_F(t)$
r_1- relative price $p_A(t)/p_D(t)$

and $y_1 = y_2$ if $\ln r_1 = 0{,}1271$.

The switch between the two linear pieces is assumed, since the dynamic models, lately used in the literature (Narasimhan and Ghosh, 1994) can prove, that the selling quantity of the product is affected by its ‘relative price’, ‘relative quality’ and ‘relative advertising rates’, where the comparison (without loss of generality) is assumed to be made with a dominant competitor. In our example we assumed, that the value of the proportion between the price of the washing powder A and the price of the washing powder D (denoted by r_1), influences two different log-linear function relationships between the selling quantity of the washing powder A and the explanatory variables.

The price of the washing powder D is, for all units t, t=1,2,...,33 smaller than the price of the washing powder A. But, if the price of the washing powder D is high, so that it is almost as high as the price of the washing powder A the selling quantity of the washing powder A is affected differently as if the prices of the washing powders A and D differ a lot. This is the assumption, which was made together with the marketing managers in Henkel Zlatorog. They assumed the switch to be approximately at r_1 = 1,15. By using the least absolute deviation criterion, the linear mixed integer optimisation model was formed and the switch of the piecewise linear function

$$\ln y_A = c_0 + c_1 \ln r_1 \qquad \ln r_1 \leq \overline{\ln r_1}$$

$$\ln y_A = d_0 + d_1 \ln r_1 \qquad \ln r_1 > \overline{\ln r_1}$$

was found at $\overline{\ln r_1} = 0.1271$ ($r_1 = 1.13$) and it was used as known in advance in the further analysis.

When estimating the regression coefficients of the regression function (4.1), it was first expressed by the dummy variable U

$$\ln y_A = \ln a_0 + a_1 \ln r + a_2 \ln r_1 + a_3 \ln I + a_4 t + a_5 U(\ln r_1 - 0.1271) \qquad (4.2)$$

and the results, which follow in Table 4.1 were calculated, if the least squares method is used.

The Jarque-Bera test for normality of regression disturbances was used (Jarque and Bera, 1987). The test statistic is

$$JB = n[(S/6 + (K-3)^2/24]$$

where

$\sqrt{S}$ - coefficient of skewness
K - coefficient of kurtosis

of the distribution of disturbances. If the disturbances are normally distributed, it can be proved that JB is asymptotically distributed as $\chi^2(2)$.

The assumption that the disturbance terms are normally distributed is rejected, since the JB test statistic is JB = 6,8129 and $P(\chi^2 > JB) = 0{,}033$.

If the 3 outliers (items t=8, t=9 and t=16) are excluded from the sample, the assumption that the disturbance terms are normally distributed is not rejected, since the JB statistics is JB = 2,48085 and $P(\chi^2 > JB) = 0{,}289$.

With the purpose to compare L_2 and L_1 estimates in this case also the least absolute value criterion was used. The results in Table 4.2 were obtained.

Table 4.1: Regression function (4.2) – least squares method.

	Estimate	Standard error	t-statistic	p-value
ln a_0	–19.1308	6.99983	–2.73307	0.011
a_1	–1.03995	0.211136	–4.92551	0.000
a_2	–4.52831	1.00279	–4.51573	0.000
a_3	2.32134	0.639858	3.62790	0.001
a_4	–0.02900	0.006456	–4.49250	0.000
a_5	5.23240	1.33001	3.93411	0.001
	Adj. R^2 = 0.874			

Table 4.2: Regression function (4.2) – least absolute value method.

	Estimate	Standard error	t-statistic	p-value
ln a_0	–19.49792	4.56081	–4.27537	0.000
a_1	–1.06287	0.137568	–7.72614	0.000
a_2	–3.50321	0.653375	–5.36171	0.000
a_3	2.34542	0.416906	5.62578	0.000
a_4	–0.02919	0.004207	–6.93925	0.000
a_5	4.19148	0.866579	4.83682	0.000
	Adj. R^2 = 0.869			

All of the regression coefficients are significant at $\alpha < 0.05$ when using least squares and when using least absolute value criterion but the L_1 estimates have smaller standard errors than the L_2 estimates. Also the adjusted R–squared in both cases is significant. The regression function (4.2) could be used for practical purposes.

5 Conclusion

This paper tries to highlight some aspects of differences between the least squares or L_2 estimates and the least absolute deviation or L_1 estimates in the linear switching regression. The results reported in the literature as well as of the Monte Carlo simulation that was performed suggest, that in the presence of large disturbances (outliers) and when the classical assumption of a normal error distribution is violated due to “contamination “ or “heavy tails”, the use of L_1 criterion could lead to the smaller standard errors of regression coefficients.

The disturbance terms are often not normally distributed, when the sales-response functions are estimated, as well as when other, micro economic and business relationships are estimated. In the example of the piecewise linear sales-response function presented in the paper the assumption of normality of error distribution was rejected. The regression coefficients estimated by L_1 and L_2 criteria were compared. Although the results were statistically significant in both cases, the L_1 estimates had smaller standard errors.

The least absolute value criterion has another advantage, especially in the cases, where the switch of the piecewise linear regression model is not known in advance. Using the least absolute value criterion the estimation problem can be formulated and solved as a linear mixed integer optimisation model (Mesko, 1989). If the switch is known in advance (a priori) the optimisation model is linear.

References

[1] Basset, G. and Koenker, R. (1978): Theory of least absolute error regressions. *Journal of the American Statistical Association*, **73**, 618-622.

[2] Charnes, A., Cooper, W.W., and Ferguson R.O. (1955): Optimal estimation of executive compensation by linear programming. *Management Science*, **1**, 138-151.

[3] Dielman, T. and Pfaffenberger, R. (1982): Least absolute value estimation in linear regression: A Review. *Studies in the Management Science*, 19, 31-52.

[4] Goldfeld, S.M. and Quandt R.E. (1976): Techniques for estimating switching regressions. In *Studies in Nonlinear Estimation*, Balinger, 3-35.

[5] Gujarati, D.N. (1995): *Basic Econometrics*. 3rd edition, Mc Graw Hill company.

[6] Jarque, C.M., Bera, A.K. (1987): A test for normality of observations and regression residuals. *International Statistical Review*, **55**, 163-172.

[7] Mesko, I. (1989): Discrete piecewise linear Ll approximation, *L'analyse Numerique et la Theorie de L'approximation*, **18**, 139-145.

[8] Mesko, I. (1994): Fractional piecewise linear optimisation of the business process including investments. *Decision Support Systems*, **12**, 213-218.

[9] Narasimhan, R. and Ghosh S. (1994): A dynamic model of manufacturing quality's effect on optimal advertising and pricing policies. *European Journal of Operational Research*, **72**, 485-502.

[10] Narula, S.C. and Korhonen P.J. (1994): Multivariate multiple linear regression based on the minimum sum of absolute errors criterion. *European Journal of Operational Research*, **73**, 70-75.

[11] Rosenberg, B. and Carlson, D. (1977): A simple approximation of the sampling distribution of least absolute residuals regression estimates. *Communications in Statistics – Simulation and Computation*, **B6**, 421-438.

[12] Tominc, P. (1996): *The Piecewise Linear Approximations in the Sales-response Functions and Their Use in the Business Optimisation.* Ph.D., Faculty of Business and Economics Maribor, University of Maribor.

[13] Time Series Processor (1993): *Reference Manual*. TSP International, Palo Alto USA.

Appendix

	$y_A(t)$	$p_A(t)$	$p_D(t)$	$p_F(t)$	$I(t)$
1	441.0	252.356	234.952	387.962	54369.101
2	495.0	259.608	249.456	387.962	54946.300
3	417.0	250.901	226.387	385.724	56310.699
4	421.0	259.339	222.701	381.465	56807.500
5	474.0	257.673	228.408	385.438	59154.898
6	471.0	245.626	223.932	381.385	61191.000
7	488.0	263.927	246.518	383.008	58298.101
8	541.0	262.647	252.252	353.430	60375.601
9	882.0	260.481	254.295	343.642	61177.300
10	447.0	269.073	211.171	348.773	61057.898
11	494.0	268.342	237.771	402.173	61962.601
12	470.0	265.993	228.956	404.040	61962.300
13	454.0	231.233	215.147	408.847	52519.398
14	478.0	230.153	214.142	364.909	63759.800
15	872.0	226.526	219.960	490.479	64930.398
16	455.0	253.916	197.780	383.159	67323.796
17	357.0	252.761	190.383	279.402	68074.101
18	360.0	250.967	187.741	265.806	69192.296
19	331.0	252.252	185.328	263.835	68893.796
20	258.0	249.681	183.439	231:847	67978.296
21	307.0	245.448	174.513	247.959	68780.296
22	246.0	244.222	173.641	246.720	67791.398
23	311.0	225.806	171.836	236.972	69052.703
24	275.0	224.969	171.199	236.093	68344.898
25	213.0	226.124	174.368	229.821	68467.000
26	227.0	225.429	173.832	229.115	69345.796
27	251.0	237.050	174.893	226.081	68315.101
28	175.0	236.330	174.362	195.018	69945.601
29	256.0	241.798	169.501	227.217	72402.203
30	265.0	241.358	169.193	226.804	72793.796
31	231.0	201.086	160.628	221.618	72288.000
32	243.0	229.066	159.090	219.497	71718.296
33	241.0	226.190	145.834	228.571	71508.601

Data (monthly).
Source of $y_A(t)$, $p_A(t)$, $p_D(t)$ and $p_F(t)$: ITEO office - Henkel Zlatorog Maribor.
Source of average gross wages: Mesečni statistični pregled, Statistični urad Republike Slovenije.