INFERENCE PROCEDURES FOR THE L_1 REGRESSION

Gabriela Stangenhaus
and
Subhash C. Narula

RELATÓRIO TÉCNICO Nº 49/87

ABSTRACT. It is well known that the L_1 estimators of the parameters of the regression model asymptotically follow a normal distribution. In this paper, using Monte Carlo approach, we determine the sample size at which we can use the normal distribution approximation to construct the confidence intervals and tests of hypothesis on the parameters in the L_1 regression model.

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Ciência da Computação
IMECC – UNICAMP
Caixa Postal 6065
13.081 - Campinas, SP
BRASIL

Novembro – 1987

INFERENCE PROCEDURES FOR THE L_1 REGRESSION

Gabriela Stangenhaus
Univ. Estadual de Campinas
Campinas, SP, Brasil

Subhash C. Narula
Virginia Commonwealth Univ.
Richmond, Virginia, U.S.A.

Keywords and Phrases: Least absolute error; Monte Carlo experiment; convergence; coverage probability.

ABSTRACT

It is well known that the L_1 estimators of the parameters of the regression model asymptotically follow a normal distribution. In this paper, using Monte Carlo approach, we determine the sample size at which we can use the normal distribution approximation to construct the confidence intervals and tests of hypothesis on the parameters in the L_1 regression model.

1. INTRODUCTION

Consider the multiple linear regression model

$$Y = X\beta + \varepsilon \tag{1}$$

where y is an $n \times 1$ vector of observations on a response variable corresponding to X, an $n \times (k+1)$ matrix of observations on the k regressor variables and a column of one's for the intercept term, β is a $(k+1) \times 1$ vector of the unknown parameters and ε is an $n \times 1$ vector of observable random errors.

As an alternative to the popular least squares procedure to determine the unknown parameters of the model (1), one may use the L_1 criterion. The L_1 regression is a robust alternative to the least squares regression and unlike other robust regression procedures, it does not require a tuning constant. It is more resistant to outliers than the least squares regression and as such provides a good starting solution for one-step and iteratively weighted multi-step least squares procedures.

A number of problems, for example, (i) the computational difficulties associated with determining the L_1 estimates of the parameters, (ii) the lack of knowledge of statistical properties of the L_1 estimators of the parameters, and (iii) non-availability of the statistical inference procedures for the parameters have prevented the use of the L_1 regression until recently. At present the computational difficulties no longer pose a problem for the L_1 regression. Charnes, Cooper and Ferguson (1955) formulated and solved the problem as a linear programming problem. Since then a number of efficient algorithms and computer programs have been depeloped. In a comparative study of the available computer programs, Gentle, Narula and Sposito (1987) show that the computer program of Armstrong, Frome and Kung (1979) is the most efficient. Furthermore, the well known statistical packages, e.g., SAS (1983) and IMSL (1980) have programs for the L_1 regression.

Just when the computational difficulties associated with the L_1 regression were resolved, Bassett and Koenker (1978) proved that in a general linear model with independent and identically distributed errors, the L_1 estimator $\hat{\beta}$ of β is unbiased, consistent and asymptotically follows a multinormal distribution with variance-covariance matrix $\lambda^2(X'X)^{-1}$, where λ^2/n is the variance of the median of a sample of size n from the error distribution. An important implication of this result is that the L_1 estimator has a strictly smaller confidence ellipsoid than the least squares estimator for β for any error distribution for which the sample median is a more efficient estimator of location than the sample mean (Rosenberg and Carlson (1977) had hinted at this result based on a Monte Carlo study with sample sizes equal to 31 and 59). Based on these results, Dielman and Pfaffenberger (1982a) gave formulas for confidence intervals and tests of hypothesis on the parameters of the model for "large" sample sizes. In an effort to determine how small is "large", Dielman and Pfaffenberger (1982b) investigated the sampling distribution of the L_1 estimator via a Monte Carlo study. They concluded that the sampling distribution appeared to be normal for sample sizes of 20 and 30 when the errors followed a normal or a contaminated normal distribution, respectively. The Cauchy and the Laplace error distributions required much larger sample sizes before the hypothesis of normally distributed sampling distributions was accepted. For the Cauchy the acceptance at the 5 percent level ocurred for a sample size 100; and for the Laplace, acceptance at the 1 percent level occurred when the sample size was 150.

Although the study by Dielman and Pfaffenberger (1982b) provided some insight into the convergence of the sampling distributions of the L_1 estimators to normality, it did not resolve the problem of statistical inference about the parameters of the model. Our objective in this paper is to determine the smallest sample size for which the normal distribution can be used to draw inferences about the parameters of the model (1). The rest of the paper is organized as follows: In the next section we describe the Monte Carlo study followed by the results and their discussion in Section 3. We conclude the paper with a few remarks in Section 4.

2. METHODOLOGY

In an effort to determine the smallest sample size for which the normal distribution can be used to draw inferences about the parameters of the model (1) we conducted a Monte Carlo study for $k = 1$ (i.e., the simple linear regression) and $k = 2$ regressor variables. In the study the intercept term β_o and the slope terms were assigned a value of 1.0. We decided to study the sampling distribution of the L_1 estimators and the inference for the parameters for sample sizes $n = 10, 15, 20, 30, 40, 50, 75$ and 100 and the following error distributions, viz., the standard normal distribution; a contaminated normal distribution that consisted of random variables drawn from the standard normal distribution with probability 0.85 and from a normal distribution with expected value zero and variance 25 with probability 0.15; the Laplace distribution with mean zero and variance 2; the Cauchy distribution with median zero and scale parameter one.

The values of the regressor variable(s) were generated as independent standard normal variates independent of the errors. The uniform (0,1) random variables were generated using FORTRAN's RAN function from VAX11/785 library; the standard normal variates were generated using the polar method of Marsaglia (1962); and the Laplace and the Cauchy random variates were generated using the inverse transformations.

For each sample size n and the error distribution, 5000 regressions were solved with identical values of the regressor variables but with different pseudo-random realizations from the error distribution. For each Monte Carlo trial, the L_1 estimates were obtained using the computer program of Armstrong, Frome and Kung (1979). The sampling distribution of the L_1 estimator was constructed from these estimates. To test whether the sampling distribution followed a normal distribution, the Kolmogorov D-statistic and the observed significance level were computed by the SAS computer package.

For each sample size and error distribution, besides testing the normality of the error distribution, $(1-\alpha)$ confidence intervals were generated for each parameter using

$$\hat{\beta}_i \pm z_{\alpha/2}\, \sigma_{\hat{\beta}_i} \tag{2}$$

where $\hat{\beta}_i$ is the L_1 estimate of β_i, z_α is the $(1-\alpha)100$ percentile of the standard normal distribution and $\sigma_{\hat{\beta}_i}$ is the standard error of the estimator obtained from the Monte Carlo sampling distribution. The coverage probabilities were calculated for each case for $\alpha = 0.01$, 0.05 and 0.10.

3. RESULTS AND DISCUSSION

The L_1 estimator of the intercept term and the slope terms in the multiple linear regression model $(k = 2)$ performed similar to the corresponding terms in the simple linear regression model, as can be observed in Tables VI, VII, VIII and IX in the Appendix A.

In table I we give the sample size and the observed significance level for the hypothesis that the sampling distribution of the L_1 estimators $\hat{\beta}_o$ and $\hat{\beta}_1$ follow a normal distribution. The sampling distribution of the estimators did not converge to the normal distri ution even for sample size 100 when errors followed aplace distribution. For this reason we increased the sample size until we achieved the convergence. From table I, one can conclude that if the errors follow a normal distribution, the sampling distriburion of the estimators follows a normal distribution for sample size as small as 10 (the smallest sample size

TABLE I

Convergence of the Sampling Distribution of the Intercept and the Slope L_1 Estimator to Normality

error distribution	sample size	observed significance level
Normal	10	>.15
Contaminated Normal	20	>.15
Laplace	200	.02
Cauchy	100	.10

considered in this study); for the contaminated normal distribution at sample size 20; for the Cauchy and Laplace distributions at sample sizes 100 and 200, respectively.

For practical purposes, it is more important to determine the sample size at which we can use the normal distribution approximation for statistical inference, e.g., confidence interval and tests of hypothesis. As noted in methodology, the confidence intervals on β_o and β_1 were constructed using (2) and the coverage percentage calculated for each sample size and error distribution. The results of this part of the study are given in Tables II, III, IV and V, for the normal, the contaminated normal, the Laplace and the Cauchy distributions, respectively.

TABLE II

Coverage Probabilities for the Intercept and Slope Parameters for the Normal Error Distribution

1-α	.90		.95		.99	
n	β_o	β_1	β_o	β_1	β_o	β_1
10	.896	.903	.948	.949	.990	.987
15	.900	.897	.948	.949	.990	.989
20	.893	.902	.943	.953	.987	.990
30	.900	.899	.953	.946	.990	.990
40	.902	.901	.948	.954	.989	.991
50	.901	.898	.949	.950	.991	.989
75	.898	.903	.949	.949	.989	.989
100	.900	.900	.948	.951	.991	.990

TABLE III

Coverage Probabilities for the Intercept and Slope Parameters for the Contaminated Normal Error Distribution

n \ 1-α	.90		.95		.99	
	β_o	β_1	β_o	β_1	β_o	β_1
10	.903	.912	.949	.949	.987	.979
15	.907	.910	.950	.952	.985	.983
20	.900	.900	.949	.950	.990	.987
30	.893	.902	.947	.947	.988	.989
40	.899	.896	.949	.948	.990	.991
50	.895	.900	.947	.951	.991	.989
75	.900	.895	.948	.945	.989	.988
100	.902	.898	.952	.946	.990	.989

TABLE IV

Coverage Probabilities for the Intercept and Slope Parameters for the Laplace Error Distribution

n \ 1-α	.90		.95		.99	
	β_o	β_1	β_o	β_1	β_o	β_1
10	.904	.900	.946	.948	.982	.981
15	.909	.900	.945	.942	.982	.982
20	.896	.901	.943	.947	.986	.985
30	.895	.903	.946	.946	.985	.984
40	.898	.899	.942	.985	.985	.984
50	.898	.899	.946	.942	.985	.984
75	.894	.903	.944	.943	.986	.986
100	.903	.898	.946	.945	.987	.985

The results in Tables II, III, IV and V are very encouraging. For example, the difference between the nominal confidence level and the percentage coverage (i) for the normal error distribution is less then or equal to 0.007, (ii) for the contaminated normal error distribution it is less than or equal to 0.012 for sample size 10 and less then 0.007 for sample sizes greater then or equal to 20, (iii) for the Laplace error distribution, it is less than or equal to 0.009 and (iv) for the Cauchy error distribution, it is less than or equal to 0.015 for sample sizes greater than or equal to 15.

TABLE V

Coverage Probabilities for the Intercept and Slope Parameters for the Cauchy Error Distribution

1-α	.90		.95		.99	
n	β_o	β_1	β_o	β_1	β_o	β_1
10	.918	.927	.951	.950	.976	.971
15	.911	.915	.948	.945	.980	.975
20	.894	.908	.938	.945	.976	.978
30	.903	.901	.949	.943	.985	.980
40	.901	.901	.948	.943	.985	.984
50	.901	.905	.944	.944	.986	.983
75	.901	.904	.950	.947	.986	.984
100	.895	.889	.949	.941	.989	.984

4. CONCLUDING REMARKS

From the results of this study it seems that the sampling distribution of the L_1 estimators of the parameters of model (1) converges to normality for very small sample sizes when the errors either follow a normal or a contaminated normal error distributions. However, the convergence is slow when the errors follow either the Laplace or the Cauchy distributions.

For all the error distributions, the difference between the nominal confidence level and the coverage probabilities is very small. This is a very useful result in that although the sampling distribution may not converge to normality, we can use the normal distribution to construct confidence intervals and test of hypothesis on the parameters of the model.

5. ACKNOWLEDGEMENT

This work was supported by grants from FAPESP and FINEP. We thank Pedro Ferreira Fo. for help with the computations.

APPENDIX A

TABLE V

Coverage Probabilities for β_o, β_1 and β_2 for the Normal Error Distribution

1−α	.90			.95			.99		
n	β_o	β_1	β_2	β_o	β_1	β_2	β_o	β_1	β_2
10	.898	.900	.897	.948	.951	.947	.989	.989	.991
15	.900	.901	.903	.950	.949	.951	.988	.990	.989
20	.900	.905	.904	.949	.954	.951	.990	.988	.989
30	.908	.902	.896	.953	.954	.946	.988	.991	.988
40	.900	.947	.989	.898	.950	.991	.897	.949	.987
50	.904	.906	.898	.957	.951	.949	.992	.989	.990
75	.898	.897	.894	.951	.950	.944	.991	.990	.988
100	.903	.894	.898	.949	.943	.952	.991	.989	.990

TABLE VI

Coverage Probabilities for β_o, β_1 and β_2 for the Contaminated Normal Error Distribution

1-α / n	.90			.95			.99		
	β_o	β_1	β_2	β_o	β_1	β_2	β_o	β_1	β_2
10	.929	.918	.955	.955	.948	.950	.977	.976	.978
15	.900	.918	.914	.948	.952	.954	.985	.981	.984
20	.898	.910	.914	.950	.955	.956	.989	.987	.983
30	.903	.904	.902	.949	.952	.949	.989	.987	.986
40	.898	.897	.903	.943	.949	.950	.988	.989	.989
50	.904	.906	.898	.957	.951	.949	.992	.989	.990
75	.887	.905	.900	.940	.950	.949	.985	.991	.988
100	.896	.891	.901	.946	.946	.950	.986	.986	.988

TABLE VII

Coverage Probabilities for β_o, β_1 and β_2 for the Laplace Error Distribution

1-α / n	.90			.95			.99		
	β_o	β_1	β_2	β_o	β_1	β_2	β_o	β_1	β_2
10	.907	.906	.900	.944	.942	.942	.977	.980	.98
15	.900	.902	.903	.942	.942	.944	.982	.982	.983
20	.895	.895	.906	.938	.941	.950	.980	.982	.984
30	.899	.895	.896	.943	.942	.942	.986	.982	.981
40	.901	.900	.899	.948	.943	.941	.984	.985	.984
50	.892	.894	.903	.941	.941	.951	.984	.982	.986
75	.896	.893	.908	.945	.942	.948	.989	.984	.985
100	.902	.902	.899	.953	.944	.948	.993	.988	.987

TABLE VIII

Coverage Probabilities for β_o, β_1 and β_2 for the Cauchy Error Distribution

$1-\alpha$	.90			.95			.99		
n	β_o	β_1	β_2	β_o	β_1	β_2	β_o	β_1	β_2
10	.952	.945	.945	.966	.962	.963	.982	.981	.978
15	.919	.917	.916	.948	.947	.944	.977	.975	.974
20	.905	.911	.926	.944	.948	.952	.978	.980	.978
30	.899	.908	.909	.941	.944	.943	.985	.979	.980
40	.903	.907	.899	.947	.945	.945	.984	.980	.983
50	.901	.896	.906	.847	.943	.948	.987	.983	.984
75	.900	.893	.898	.945	.938	.946	.988	.983	.988
100	.893	.907	.904	.944	.953	.956	.989	.988	.988

BIBLIOGRAPHY

Armstrong, R.D. Frome, E.L. and Kung, D.S. (1979). A revised simplex algorithm for the absolute deviation curve fitting problem. *Communications in Statistics - Simulation and Computation, B8*, 179-90.

Bassett, G.W. and Koenker, R. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association, 73*, 618-22.

Charnes, A., Cooper, W.W. and Ferguson, R.O. (1955). Optimal estimation of executive compensation by linear programming, *Management Science, 1*, 138-50.

Dielman, T. and Pfaffenberger, R. (1982a). LAV (least asbolute value) estimation in the regression model: A review. *TIMS Studies in the Management Sciences, 19*, 31-52.

Dielman, T. and Pfaffenberger, R. (1982b). On the rate of convergence in distribution to normality of the least absolute value estimator in the linear regression model. Working Paper No. 5, Texas Christian University, Fort Worth, Texas.

Gentle, J.E., Narula, S.C. and Sposito, V.A. (1987). Algorithms for unconstrained L_1 linear regression. *Statistical Data Analysis Based in the L_1-norm and Related Methods*, edited by Yadolah Dodge, North-Holland, 83-94.

International Mathematical and Statistical Libraries, Inc. (1980). *IMSL Library 3, Reference Manual*, Houston, TX.

Marsaglia, G. (1962). Random Variables and computers. *Information Theory Statistical Decision Functions Random Processes: Transactions of the Third Prague Conf.*, edited by J. Kozesnik Czechoslovak Academy of Sciences, Prague, 499-510.

Rosenberg, B. and Carlson, D. (1977). A simple approximation of of the sampling distribution of least absolute residuals regression estimates. *Communications in Statistics - Simulation and Computation, B6*, 421-27.

SAS Institute, Inc. (1983). *SUGI Supplemental Library User's Guide, 1983 Edition*, SAS Institute, Inc., Cary, N.C.