M-Estimators in Regression Models

Muthukrishnan.R
Department of Statistics, Bharathiar University
Coimbatore-641 046, Tamilnadu, India
E-mail: muthukrishnan70@rediffmail.com

Radha.M
Department of Statistics, Bharathiar University
Coimbatore-641 046, Tamilnadu, India
E-mail: radhamyilsamy@gmail.com

Abstract

Regression analysis plays a vital role in many areas of science. Almost all regression analyses rely on the method of least squares for estimation of the parameters in the model. But this method is usually constructed under specific assumptions, such as normality of the error distribution. When outliers are present in the data, this method of estimation, results in parameter estimates that do not provide useful information for the majority of the data. Robust regression analyses have been developed as an improvement to least square estimation in the presence of outliers. The main purpose of robust regression analysis is to fit a model that represents the information of the majority of the data. Many researchers have worked in this field and developed methods for these problems. The most commonly used robust estimators are Huber's M-estimator, Hampel estimator, Tukey's bisquare estimator etc. In this paper, an attempt is made to review such type of estimators and a simulation study of these estimators in regression models is carried out. R code has been written for the purpose and illustrations are provided.

Keywords: Regression Model, Robust estimator, M-estimators, R software

1. Introduction

The theory of robustness developed by Huber and Hampel (1960) laid the foundation for finding practical solutions too many problems, when statistical concepts were vague to serve the purpose. Robust regression analyses have been developed as an improvement to least squares estimation in the presence of outliers and to provide us information about what a valid observation is and whether this should be thrown. The primary purpose of robust regression analysis is to fit a model which represents the information in the majority of the data. Robust regression is an important tool for analyzing data that are contaminated with outliers. It can be used to detect outliers and to provide resistant results in the presence of outliers. Many methods have been developed for these problems. Many researchers have worked in this field and described the methods of robust estimators. The class of robust estimators includes M-, L- and R-estimators. The M-estimators are most flexible ones, and they generalize straightforwardly to multiparameter problems, even though they are not automatically scale invariant and have to be supplemented for practical applications by an auxiliary estimate of scale any estimate. In this paper, an attempt has been made to make an elaborate study of the some of the M-estimators. Section 2 deals with the descriptions of the M-estimators. The redescending M-estimators are presented in the section 3. A simulation study of these estimators providing certain numerical illustrations by using R software is presented in the last section.

2. M-estimator

The class of M-estimator was introduced by P.J.Huber in 1964; subsequently, such estimators have been discussed extensively by several authors, Andrews et al. (1972), Bunke and Bunke (1986), Hampel et al. (1986), Lecoutre and Tassi (1987), Robusseeuw and Leroy (1987), Staudte and Sheather (1990), Rieder (1994), Jureckova and Sen (1996), Antoch et al. (1998), Dodge and Jureckova (2000), Jureckova and Picek (2006) and others. M-estimator T_n is defined as a solution of the minimization problem,

$$\sum_{i=1}^{n} \rho(X_i, \theta) := min, \;\; with\, respect\, to \quad \theta \in \Theta$$

$$E_{P_n}[\rho(X, \theta)] = min, \quad \theta \in \Theta \tag{1}$$

where $\rho(\cdot, \cdot)$ is a properly chosen function. The class of M-estimator covers also the maximal likelihood estimator of parameter θ in the parametric model $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$;if $f(x, \theta)$, is the density function of P_θ, then the MLE is a solution of the minimization

$$\sum_{i=1}^{n}(-logf(X_i,\theta)) = min, \quad \theta \in \Theta$$

If ρ in (1) is differentiable in θ with a continuous derivative $\psi(\cdot,\theta) = \frac{\partial}{\partial\theta}\rho(\cdot,\theta)$ then, T_n is a root or roots of the equation

$$\sum_{i=1}^{n}\psi(X_i,\theta) = 0, \quad \theta \in \Theta$$

hence

$$\frac{1}{n}\sum_{i=1}^{n}\psi(X_i,T_n) = E_{P_n}[(X,T_n)] = 0 \quad T_n \in \Theta \tag{2}$$

From (1) and (2) that the M-functional corresponding to T_n, is defined as a solution of the minimization

$$\int_x \rho(x,T(P))dP(x) = E_P[\rho(X,T(P)] = min, T(P) \in \Theta \tag{3}$$

or as the solution of the equation

$$\int_x \psi(x,T(P))dP(x) = E_P[\psi(X,T(P)] = min, T(P) \in \Theta \tag{4}$$

The function T(P) is Fisher consistent, if the solutions of (3) and (4) are uniquely determined.

2.1 M-estimator of Location parameter

An important special case is the model with the shift parameter θ, where $X_1, X_2, ..., X_n$ are independent observations with the same distribution function $F(x-\theta), \theta \in \Re$; the distribution function F is generally unknown. M-estimator of location parameter T_n is defined as a solution of the minimization

$$\sum_{i=1}^{n}\rho(x_i-\theta) := min \tag{5}$$

and if $\rho(\cdot)$ is differentiable with absolutely derivative $\psi(\cdot)$, then T_n solves the equation

$$\sum_{i=1}^{n}\psi(x_i-\theta) = 0 \tag{6}$$

The corresponding M-functional T(F) is Fisher consistent, provided the minimization

$$\int_x \rho(X-\theta)dP(x) = min \tag{7}$$

have a unique solution $\theta = 0$, i.e., the solution of the equation is,

$$\int_x \psi(X-\theta)dP(x) = 0. \tag{8}$$

2.2 Asymptotic properties of M-estimator

A fairly simple and straightforward theory is possible if $\psi(x,\theta)$ is monotone in θ. Assume that $\psi(x,\theta)$ is measurable in x and decreasing in θ, from strictly positive to strictly negative values. Put

$$T_n^* = Sup\left\{|t|\sum_{i=1}^{n}\psi(x_i;t) > 0\right\},$$

$$T_n^{**} = Inf\left\{|t|\sum_{i=1}^{n}\psi(x_i;t) < 0\right\}, \tag{9}$$

Clearly,$-\infty < T_n^* \leq T_n^{**} < \infty$, and any value T_n satisfying $T_n^* \leq T_n \leq T_n^{**}$ can serve as our estimate. Note that

$$\{T_n^* < t\} \subset \left\{\sum \psi(x_i;t) \leq 0\right\} \subset \{T_n^* \leq t\},$$

$$\{T_n^{**} < t\} \subset \left\{\sum \psi(x_i; t) < 0\right\} \subset \{T_n^{**} \le t\}. \tag{10}$$

hence

$$P\{T_n^* < t\} = P\left\{\sum \psi(x_i; t) \le 0\right\},$$

$$P\{T_n^{**} < t\} = P\left\{\sum \psi(x_i; t) < 0\right\}, \tag{11}$$

at the continuity points t of the left-hand side. The distribution of the customary midpoint estimate $1/2(T_n^* + T_n^{**})$ is somewhat difficult to work out, but the randomized estimate T_n, which selects one of T_n^* or T_n^{**} at random with equal probability, has an explicitly expressible distribution function

$$P\{T_n < t\} = \frac{1}{2}P\left\{\sum \psi(x_i; t) \le 0\right\} + \frac{1}{2}P\left\{\sum \psi(x_i; t) < 0\right\} \tag{12}$$

It follows that the exact distributions of T_n^*, T_n^{**}, and T_n can be calculated from the convolution powers of $G_n = \mathcal{L}(\psi(x_i; t))$. Let,

$$\lambda(t) = \lambda(t, F) = E_F\psi(X, t). \tag{13}$$

If λ exists and is finite for atleast one value of t, then it exists and is monotone for all t. Assume that there is a t_0 such that $\lambda(t) > 0$ for $t < t_0$ and $\lambda(t) < 0$ for $t > t_0$. Then both T_n^* and T_n^{**} converge in probability and almost surely to t_0. Consider the following conditions,

(C1) $\psi(x, t)$ is measurable in x and monotone decreasing in t.

(C2) There is atleast one t_0 for which $\lambda(t_0) = 0$. Let Γ_0 be the set of t-values for which $\lambda(t) = 0$.

(C3) λ is continuous in a neighborhood of Γ_0.

(C4) $\sigma(t)^2 = E_F[\psi(X, t)^2] - \lambda(t, F)^2$ is finite, nonzero, and continuous in a neighborhood of Γ_0. Put $\sigma_0 = \sigma(t_0)$.

Under the above conditions $\sqrt{n}\lambda(T_n)$ is a asymptotically normal $N(0, \sigma_0^2)$.

3. Redescending M-estimator

Redescending M-estimators are very popular Ψ-type M-Estimator which has Ψ functions that are non-decreasing near the origin, but decreasing toward 0 far from the origin. Their Ψ functions can be chosen to redescend smoothly to zero, so that they usually satisfy $\Psi(x) = 0$ for all x with $|X| > k$,where r is referred to as the minimum rejected point. When choosing a redescending Ψ functions we must take care that it does not descend too steeply, which may have a very bad influence on the denominator in the expression for the asymptotic variance

$$\frac{\int \Psi^2 dF}{(\int(\Psi' dF))^2}$$

where F is the mixture model distribution. This effect is particularly harmful when a large negative values of $\Psi'(x)$ combines with a large positive values of$\Psi^2(x)$, and there is a cluster of outliers near x. First we introduce Hampel's three-part M-estimator, it has Ψ functions which are odd functions and defined for any x by:

$$\psi(x) = \left\{\begin{array}{ccc} x, & 0 \le |x| \le a & (central\ segment) \\ asign(x) & a \le |x| \le b & (high\ and\ low\ flat\ segments) \\ \frac{a(k-|x|)}{k-b} sign(x) & b \le |x| \le k & (end\ slopes) \\ 0, & k \le |x| & (left\ and\ right\ tails) \end{array}\right\}.$$

Tukey's biweight or bisquare M-estimator have ψ functions for any positive k, which defined by

$$\psi(x) = \left\{\begin{array}{cc} x[(1 - x/k)^2]^2 & for\ |x| \le k \\ 0 & for\ |x| > k \end{array}\right\}.$$

Huber proposed function in 1964,that is

$$\psi(x) = \left\{\begin{array}{cc} x, & for|x| \le k \\ ksignx, & for|x| > k \end{array}\right\}.$$

For regression analysis, some of the redecending M-estimators can attain the maximum breakdown point. Moreover, some of them are the solutions of the problem of maximizing the efficiency under bounded influence function when the regression coefficient and the scale parameter are estimated simultaneously. Hence redecending M-estimators satisfy several outlier robustness properties.

4. Simulation Results

This section presents the simulation results to check the performance of Huber M-estimator as compared to other well known redescending M-estimators. The simulation study is carried out in three stages. First stage is the normal situation; consider the following linear regression model

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, ..., n.$$

in which $u_i \sim N(0, 1)$ and the explanatory variables are generated as $x_i \sim N(100, 2)$ using R software and then the values of y_i's are evaluated for the specified values of $\beta_0 = 2$ and $\beta_1 = 2$. Then, values β_0 and β_1 are computed under various methods of estimators by using R software. In the second stage, 10% of the y_i observations are replaced by the values generated from N(10,5), which are referred as outliers in y-direction. After that, as usual, computations performed to estimate the values of β_0 and β_1. In the last stage, 20% of the x_i observations are replaced by the values are generated from N(10,5) which are also referred as outliers in x-direction with the same observations available in the second stage. The estimated values of β_0 and β_1 in different stages are summarized in Table 1 for the value of n fixed as 50. The same procedure is repeated for n=100 and n=500, and the results arrived by using R software, are presented in Table 2 and 3. From these tables it is clear that the results of the redescending M-estimator are very similar to that of ordinary least square estimator in normal situation. The redecending estimators are not affected by the outliers in both second and third situations while the ordinary least square estimator is affected in these situations.

5. Conclusion

The performance of robust estimators has been assessed in regression model. Estimators and results are obtained by using R software. It is interesting to note that the class of M-estimators is found to yield essentially the same results as the method of least square estimator in normal situation. When outliers are present in the data; least square estimator does not provide useful information for the majority of the data but not in the case of robust estimators. That is, it is observed that the M-estimators are not affected by outliers. The study establishes the fact that the performance of M-estimators are almost same as the method of least squares in normal situations and also in the presence of outliers. Hence it is concluded that the robust statistical procedures can be considered as modification of the classical procedures and such procedures may not fail when there are small deviations from the assumed conditions.

References

Ali, A., Qadir, M. F. (2005). A Modified M-estimator for the detection of outliers. *Pakistan journal of Statistics and Operations Research*, 1, 49-64.

Andrews, D. F. (1971). Significance test based on residuals. *Biometrika*, 58, 139-148.

Anscobe, F. J. (1960). Rejection of outliers. *Technometrics*, 2, 123-147.

Bunke, H & Bunke, O (eds.). (1986). *Statistical inference in Linear models*. John Wiley & Sons, Chichester, U.K.

Huber, P.J. (1964). Robust estimation of location parameter. *The Annals of Mathematical Statistics*, 35, 73-101.

Huber, P.J. (1973). Robust regression: Asymptotics, Conjectures and Monte Carlo. *The Annals of Statistics*, 1, 799-821.

Huber, P.J. (1981). *Robust Statistics*. John Wiley & Sons, New York.

Insha Ullah, Muhammd, F. Quadir, Asad ali. (2006). Insha's redecending M-estimator for robust regression: A comparative study. *Pakistan Journal of Statistics and Operations research*, Vol. II, No.2, 2006, pp 135-144.

Jureckova, J. (1980). Asymptotic representation of M-estimators of location. Math. Operationsforsch. Und Statistik, *Ser. Statistics*, 11:1, 61-73.

Jureckova, J. & Jan picek. (2006). *Robust statistical methods with R*. Chapman & Hall/CRC.

Lecoutre, J.P and Tassi, P. (1987). Statistique non parametricque et robustesse. *Economica*, Paris.

Rand.R.Wilcox. (2005). *Introduction to Robust Estimation and Hypothesis Testing*. Academic Press.

Rieder. (1994). *Robust Asymptotic Statistics*. Springer, New York.

Rousseeuw and Leory, A.M. (1987). *Robust regression and outlier detection*. John Wiley & Sons, New York.

Staudte, W.A and Sheather, S.J. (1990). *Robust estimation and testing*. John Wiley & Sons, New York.

Table 1. Simulation Results of Regression with Intercept, for n=50

Estimators	**Normal**		**Outliers in y**		**Outliers in x**	
	β_0	β_1	β_0	β_1	β_0	β_1
Least Square Estimator	2.07	1.98	-18.46	0.37	13.87	1.22
Huber M-estimator	2.00	1.99	1.99	1.96	2.02	1.97
Hampel M-estimator	2.07	1.98	2.03	1.93	2.09	1.95
Tukey's M-estimator	2.04	2.00	2.01	1.96	2.06	1.98

Table 2. Simulation Results of Regression with Intercept, for n=100

Estimators	**Normal**		**Outliers in y**		**Outliers in x**	
	β_0	β_1	β_0	β_1	β_0	β_1
Least Square Estimator	2.02	2.03	-4.52	0.26	10.58	1.43
Huber M-estimator	2.00	2.00	1.99	1.99	2.00	2.00
Hampel M-estimator	2.00	2.01	1.97	1.98	2.06	2.03
Tukey's M-estimator	2.01	2.00	1.96	1.95	2.04	2.03

Table 3. Simulation Results of Regression with Intercept, for n=500

Estimators	**Normal**		**Outliers in y**		**Outliers in x**	
	β_0	β_1	β_0	β_1	β_0	β_1
Least Square Estimator	2.01	2.02	-3.58	0.17	8.53	8.53
Huber M-estimator	2.00	2.03	2.00	2.01	2.02	2.05
Hampel M-estimator	1.96	2.00	1.97	1.98	1.98	1.99
Tukey's M-estimator	1.97	1.99	1.95	1.97	1.95	1.95