International Journal of Statistical Distributions and Applications
2021; 7(4): 108-114
http://www.sciencepublishinggroup.com/j/ijsda
doi: 10.11648/j.ijsd.20210704.15
ISSN: 2472-3487 (Print); ISSN: 2472-3509 (Online)

Likelihood-Based Confidence Intervals for the Parameters of a Simple Linear Regression Model with Cauchy Errors

Orawo Luke Akongo

Mathematics Department, Egerton University, Njoro, Kenya

Email address:
luke.orawo@egerton.ac.ke

To cite this article:
Orawo Luke Akongo. Likelihood-Based Confidence Intervals for the Parameters of a Simple Linear Regression Model with Cauchy Errors. *International Journal of Statistical Distributions and Applications.* Vol. 7, No. 4, 2021, pp. 108-114. doi: 10.11648/j.ijsd.20210704.15

Received: October 21, 2021; **Accepted**: November 12, 2021; **Published**: November 24, 2021

Abstract: The estimators of the slope and the intercept of simple linear regression model with normal errors are normally distributed and their exact confidence intervals are constructed using the t-distribution. However, when the normality assumption is not fulfilled, it is not possible to obtain exact confidence intervals. The Wald method of interval estimation is commonly used to provide approximate confidence intervals in such cases, and since it is derived from the central limit theorem it requires large samples in order to provide reliable approximate confidence intervals. This paper considers an alternative method of constructing approximate confidence intervals for the parameters of a simple linear regression model with Cauchy errors which is based the normal approximation to the Cauchy likelihood. The normal approximation to the Cauchy likelihood is obtained by a Tailor series expansion of the Cauchy log-likelihood function about the maximum likelihood estimate of the parameters and ignoring terms of order greater two. The maximized relative log-likelihood function for each parameter is then derived from the normal Cauchy relative log-likelihood function. The approximate confidence intervals for the parameters are constructed from their respective maximized relative log-likelihood functions. These confidence intervals have closed form confidence limits, are short and have coverage probabilities close to the nominal value 0.95.

Keywords: Normal Approximations, Relative Likelihood Function, Maximized Relative Likelihood Function, Likelihood Confidence Intervals

1. Introduction

The usual assumption on the distribution of the error term in a linear regression model is that the errors are identically and independent distributed normal random variables with mean equal to zero and has constant but unknown variance σ^2. However, there are situations where this assumption may not hold in that the distribution of the errors departs significantly from the normal distribution. For instance, in finance the distribution of the errors in the capital asset pricing model is well known to be 'fat-tailed' [22]. Bartolucci and Scaccia [2] showed, through a simulation study, that an effective strategy to deal with these situations is fitting a regression model based on the assumption that the error terms follow a mixture of normal distributions.

The Least Squares method has been widely applied in linear regression analysis as a procedure for estimating the regression coefficients and its popularity is due to the fact that it does not involve heavy computations as well as the only assumption is that the errors have finite mean and variance. Furthermore, according to the Gauss-Markov theorem, the LS estimators will have minimum variance within the class of all linear unbiased estimators. In fact, if the errors are normally distributed, the LS estimators will be equivalent to the Maximum Likelihood estimators. However, when non-normality assumption prevails the distribution of the errors may not have finite mean and variance and the LS estimators will not have minimum variance and not efficient. On the other hand, the ML estimators which are based on the distribution of the errors are well known to be consistent and efficient hence they can be relied on. Some efforts have been put in search for better methods in non-normal regression. Kadiyala and Murthy [13], compared the ML method with the method of minimum sum of absolute errors (MSAE) in a multiple linear regression model with Cauchy errors. The method of Modified Maximum Likelihood (MML) estimation was invoked by Bian and Tiku [3] and Tiku et al

[19] to estimate the parameters of the linear regression model where the errors assumed a distribution from the Student's t family of distributions. The MML method utilizes Taylor series expansion to approximate some terms in the likelihood function so as to make it linear so that closed form estimates can be obtained. The MML method was used by Islam *et al.* [11] to a linear regression model with the errors assumed to be from a family of skewed distributions, Weibull and generalized logistic. They showed that these estimators are efficient and robust. Tiku *et al.* [20] also applied the MML method to a linear regression model with errors from non-normal symmetric distributions. They also found that the estimators are efficient and robust in the presence of outliers. Modified Maximum Likelihood estimators for the multiple regression coefficients in a linear model with the underlying distribution assumed to be one of Student's t family were developed by Wong and Bian [21]. They demonstrated that the MML estimators are more efficient in estimating the parameters in the Capital Asset Pricing Model by comparing its performance with that of least squares estimators on the monthly returns of US portfolios. Akkaya and Tikku [1] considered the estimation of the parameters of a simple autoregressive model where the errors have asymmetric distribution and showed that the OLS estimators are inefficient. Xiang et al [23], carried out a simulation study to investigate the efficiency of the ordinary least squares estimators of the parameters of a simple linear regression with non-normal errors and concluded that for large samples (n>3000) the OLS estimates are efficient. Howard M. [9] used simulation to compare OLS estimators and quantile estimators when the assumptions of normality and homoscedasticity of error distribution are violated and found that quantile estimators perform drastically better than OLS estimators.

A comparison of the ML method with the method MSAE was done by Kadiyala and Murthy [13]. They compared the performance of these two methods in a multiple regression model with Cauchy errors and found that the Maximum Likelihood estimators out-performed the MSAE estimators. However, the MSAE estimators seemed to be a closer competitor of the ML estimators. Bayesian methods have also been used by Geweke [8] and Fernandez and Steel [5]. Geweke [9] employed Bayesian methods to construct a linear model where the errors followed a symmetric t-distribution. Fernandez and Steel [5] used Bayesian Markov Chain Monte Carlo (MCMC) methods to develop a linear model with the errors following a skewed t-distribution. Some pitfalls were revealed by Fernandez and Steel [6] on both Maximum Likelihood and Bayesian methods of inference from a multivariate regression model with independent Student-t errors with unknown degrees of freedom. A class of robust estimators for the parameters of a regression model with the distribution of the error terms coming from a log-gamma distribution was proposed Bianco *et al.* [4].

This paper presents construction of approximate likelihood confidence intervals and regions for the parameters of a simple linear regression model with Cauchy errors by means of normal approximation to the Cauchy likelihood model. The efficiency of the approximate confidence intervals is evaluated by means of coverage probabilities via simulation.

2. The Cauchy Model

2.1. Cauchy Probability Distribution

The Cauchy distribution, also known as the Lorentz distribution, is family of continuous probability distributions similar to the normal distribution family of curves. However, they are more peaked and have heavier tails than the normal distribution. The Cauchy distribution serves as counter example for some well accepted results and concepts in statistics. For instance, Stuart and Ord [18] showed that the sampling distribution of the mean of a random sample of n independent observations from a Cauchy distribution fails to tend to normal for large sample size n and the central limit theorem is inapplicable. A random variable X follows a Cauchy distribution with location parameter θ and scale parameter σ if it has the probability density function

$$f(x) = (\pi\sigma)^{-1}\left[1+\left(\frac{x-\theta}{\sigma}\right)^2\right]^{-1}, -\infty < x < \infty. \quad (1)$$

The location parameter θ is the location of the peak of the distribution (the mode of the distribution), while the scale parameter σ specifies half the width of the density function at half the maximum height. We write $X \sim C(\theta, \sigma)$ to mean that the random variable X has a Cauchy distribution with parameters θ and σ.

The cumulative distribution function is given by

$$F(x) = \frac{1}{2} + \pi^{-1}\tan^{-1}\left(\frac{x-\theta}{\sigma}\right) \quad (2)$$

The Cauchy distribution is symmetrical about $x = \theta$ and is bell-shaped just like the normal distribution curve. It is a classical example of a distribution that has inexistent moments of order greater than or equal to 1. Consequently, it does not have both mean and finite variance. However, expressions for fractional absolute moments of a Cauchy distribution were derived by Goria [7]. The median of the Cauchy distribution is θ while the lower and upper quartiles are $\theta - \sigma$ and $\theta + \sigma$, respectively. As a result of the divergent moments, the law of large numbers and the central limit theorem do not apply. The location parameter θ and the scale parameter σ may be regarded as analogous to the mean and standard deviation respectively (Johnson *et al.,* [12].

The Cauchy distribution arises in resonance and spectroscopy applications in physics, and also serves as an additional probability to model fat tails in computational finance; the Cauchy distribution can be used to model VAR (value at risk) producing much larger probability of extreme risk than the Gaussian distribution. Mahdizadeh and Zamanzade [17] developed goodness of fit tests for the Cauchy distribution and considered its application to financial modeling.

Apart from the maximum likelihood method, an iterative method for estimating the scale and location parameters of a Cauchy distribution using the empirical characteristic function was proposed by Koutrouvelis [15]. An alternative estimator for the location and scale parameters are the median and the semi-interquartile range, respectively (Howlader and Weiss [10].

2.2. The Cauchy Likelihood Model

Suppose there are n pairs of observations ($x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. These observations are assumed to satisfy the simple linear regression model, and so we can write

$$y_i = \alpha + \beta x_i + \varepsilon_i, i = 1,2, \ldots, n \tag{3}$$

where x_i is the predictor variable, y_i is the dependent variable, and ε_i is the random error term. The $\varepsilon_i's$ are assumed to be IID $C(0, \sigma)$ random variables. The distribution of the response variable will be $C(\alpha + \beta x_i, \sigma)$ and the likelihood function will be defined as

$$L(\alpha, \beta, \sigma) = (\pi\sigma)^{-n} \prod_{i=1}^{n} \left[1 + \left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)^2\right] \tag{4}$$

Taking the logarithm of (4) the log-likelihood function is obtained.

$$l(\alpha, \beta, \sigma)$$
$$= -n \log \pi - n \log \sigma - \sum_{i=1}^{n} \log\left(1 + \left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)^2\right) \tag{5}$$

Differentiating partially the above log-likelihood function with respect to each of the three parameters, the score functions are obtained as

$$\frac{\partial l}{\partial \alpha} = 2 \sum_{i=1}^{n} \frac{y_i - \alpha - \beta x_i}{\sigma^2 + (y_i - \alpha - \beta x_i)^2} \tag{6}$$

$$\frac{\partial l}{\partial \beta} = 2 \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)x_i}{\sigma^2 + (y_i - \alpha - \beta x_i)^2} \tag{7}$$

$$\frac{\partial l}{\partial \sigma} = 2 \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2 + (y_i - \alpha - \beta x_i)^2} - \frac{n}{\sigma} \tag{8}$$

The elements of the hessian matrix are the following second order partial derivatives of the log-likelihood function

$$\frac{\partial^2 l}{\partial \alpha^2} = -2 \sum_{i=1}^{n} \frac{\sigma^2 + (y_i - \alpha - \beta x_i)^2}{(\sigma^2 + (y_i - \alpha - \beta x_i)^2)^2}$$

$$\frac{\partial^2 l}{\partial \beta^2} = 4 \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2 {x_i}^2}{(\sigma^2 + (y_i - \alpha - \beta x_i)^2)^2} - 2 \sum_{i=1}^{n} \frac{{x_i}^2}{\sigma^2 + (y_i - \alpha - \beta x_i)^2}$$

$$\frac{\partial^2 l}{\partial \sigma^2} = \frac{n}{\sigma^2} - 4 \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{(\sigma^2 + (y_i - \alpha - \beta x_i)^2)^2} - \frac{2}{\sigma^2} \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2 + (y_i - \alpha - \beta x_i)^2}$$

$$\frac{\partial^2 l}{\partial \alpha \partial \beta} = 4 \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2 {x_i}^2}{(\sigma^2 + (y_i - \alpha - \beta x_i)^2)^2} - 2 \sum_{i=1}^{n} \frac{x_i}{\sigma^2 + (y_i - \alpha - \beta x_i)^2}$$

$$\frac{\partial^2 l}{\partial \alpha \partial \sigma} = -4\sigma \sum_{i=1}^{n} \frac{y_i - \alpha - \beta x_i}{(\sigma^2 + (y_i - \alpha - \beta x_i)^2)^2}$$

$$\frac{\partial^2 l}{\partial \beta \partial \sigma} = -4\sigma \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)x_i}{(\sigma^2 + (y_i - \alpha - \beta x_i)^2)^2}$$

The partial derivatives in $(6), (7)$ and (8) are equated to zero and the resulting equations are then numerically solved simultaneously for α , β and σ to obtain the maximum likelihood estimates of the unknown parameters. The starting points for α and β in the iterative algorithm were the corresponding least squares estimates while that of the scale parameter was half of the inter-quartile range.

3. Likelihood Regions and Intervals

The relative likelihood function (RLF) of $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$, is defined as

$$R(\theta_1, \theta_2, \ldots, \theta_k) = \frac{L(\theta_1, \theta_2, \ldots, \theta_k)}{L(\hat{\theta}_1 \hat{\theta}_2, \ldots, \hat{\theta}_k)} \tag{9}$$

Note that $0 \leq R(\theta) \leq 1$ and$R(\hat{\theta}) = 1$. Consequently, the natural logarithm of R, called the relative log-likelihood function, is defined as

$$r(\theta_1, \theta_2, \ldots, \theta_n) = l(\theta_1, \theta_2, \ldots, \theta_k) - l(\hat{\theta}_1 \hat{\theta}_2, \ldots, \hat{\theta}_k) \tag{10}$$

A major importance of the joint RLF is that it ranks the sets of parameter values according to their plausibility such that a set of parameter values with RLF value close to 1 is more plausible than any other set with RLF value near zero. For $k \leq 2$, it is possible to plot the RLF as well as its logarithmic function. When $k = 2$, the $100p\%$ likelihood region is the set of values (θ_1, θ_2) such that $R(\theta_1, \theta_2) \geq p$ while the curve $R(\theta_1, \theta_2) = p$, which forms the boundary of this region, is called the $100p\%$ likelihood contour. The 14.7% and 3.6% likelihood regions correspond to 95% and 99% confidence regions, respectively.

It is very hard to plot and interpret the RLF for the case of $k > 2$. Therefore, the parameters can be considered one or even two at a time and inference is made just for the selected parameters. This can be achieved using the maximized relative likelihood function. The maximized relative likelihood function of θ_j is obtained by holding θ_j fixed and maximizing the joint RLF over the other $k - 1$ parameters. Thus,

$$R_{\max(\theta_j)} = \max_{\theta_j \, fixed} R(\theta_1, \theta_2, \ldots, \theta_k) \tag{11}$$

If the number of unknown parameters is small in comparison with the number of independent observations, then the maximized RLF has properties similar to those of a one parameter RLF and satisfactory results will be obtained (Kalbfleisch, [14]. Inferences concerning θ_j can be made using the maximized RLF of θ_j. The $100p\%$ likelihood interval for θ_j is the set of parameter values such that $R_{\max(\theta_j)} \geq p$. The 14.7% and 3.6% likelihood regions correspond to 95% and 99% confidence regions, respectively.

4. Normal Approximations to the Cauchy Likelihood Model

4.1. The Normal Log-Relative Likelihood

The normal approximation to the Cauchy likelihood model is derived using the Taylor series expansion for the Cauchy log likelihood function $l(\theta)$ in (5), with $\theta = (\alpha, \beta, \sigma$. The Taylor series expansion for the Cauchy likelihood model at the maximum likelihood estimate $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\sigma})$ is given by

$$l(\theta) = l(\hat{\theta}) + (\theta - \hat{\theta})l'(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})l''(\hat{\theta})(\theta - \hat{\theta})' + \cdots + R_n \quad (12)$$

where R_n is the remainder term. Noting that $l'(\hat{\theta}) = 0$ and ignoring the remainder term and the terms that are of order greater 2, the normal approximation to the Cauchy relative log likelihood function, $r(\theta)$, is obtained as

$$r_N(\theta) = -\tfrac{1}{2}(\theta - \hat{\theta})I(\hat{\theta})(\theta - \hat{\theta})' \quad (13)$$

where $I(\hat{\theta}) = -l''(\hat{\theta})$. The matrix I is the information matrix. Expanding the above equation results to the following normal approximation to the Cauchy likelihood model with $\boldsymbol{\theta} = (\alpha, \beta, \sigma)$

$$r_N(\alpha, \beta, \sigma) = -\frac{1}{2}(\alpha - \hat{\alpha})^2\hat{I}_{11} - \frac{1}{2}(\beta - \hat{\beta})^2\hat{I}_{22} - \frac{1}{2}(\sigma - \hat{\sigma})^2\hat{I}_{33} - (\alpha - \hat{\alpha})(\beta - \hat{\beta})\hat{I}_{12} - (\alpha - \hat{\alpha})(\sigma - \hat{\sigma})\hat{I}_{13} - (\beta - \hat{\beta})(\sigma - \hat{\sigma})\hat{I}_{23} \quad (14)$$

Differentiating equation (14) partially with respect to σ, and equating it to zero and solving for σ as a function of both α and β yields.

$$\hat{\sigma}(\alpha, \beta) = \hat{\sigma} - \frac{(\alpha - \hat{\alpha})\hat{I}_{13} + (\beta - \hat{\beta})\hat{I}_{23}}{\hat{I}_{33}} \quad (15)$$

Substituting (15) into (14), the maximized log-RLF of α and β is obtained as

$$r_{max}(\alpha, \beta) = -\frac{1}{2}(\alpha - \hat{\alpha})^2\left(\hat{I}_{11} - \frac{\hat{I}_{13}{}^2}{\hat{I}_{33}}\right) - \frac{1}{2}(\beta - \hat{\beta})^2\left(\hat{I}_{22} - \frac{\hat{I}_{23}{}^2}{\hat{I}_{33}}\right) - (\alpha - \hat{\alpha})(\beta - \hat{\beta})\hat{I}_{12}\left(\hat{I}_{12} - \frac{\hat{I}_{13}\hat{I}_{23}}{\hat{I}_{33}}\right) \quad (16)$$

The construction of approximate confidence intervals for individual parameters requires $r_{max}(\alpha)$, $r_{max}(\beta)$ and$r_{max}(\sigma)$, the maximized relative log likelihood functions of α, β and σ, respectively. To obtain $r_{max}(\alpha)$, equation (14) is maximized over (β, σ) while holding the parameter α constant. Thus,

$$r_{max}(\alpha) = r_N(\alpha, \hat{\beta}(\alpha), \hat{\sigma}(\alpha)) \quad (17)$$

The functions $\hat{\beta}(\alpha)$ and $\hat{\sigma}(\alpha)$ are solutions to the simultaneous equations $\frac{\partial r_N}{\partial \beta} = 0$ and $\frac{\partial r_N}{\partial \sigma} = 0$. These solutions are as follows

$$\hat{\sigma}(\alpha) = \hat{\sigma} + (\alpha - \hat{\alpha})\left(\frac{\hat{I}_{22}\hat{I}_{13} - \hat{I}_{12}\hat{I}_{23}}{\hat{I}_{23}{}^2 - \hat{I}_{22}\hat{I}_{33}}\right) \quad (18)$$

$$\hat{\beta}(\alpha) = \hat{\beta} + (\beta - \hat{\beta})\left(\frac{\hat{I}_{12}\hat{I}_{33} - \hat{I}_{13}\hat{I}_{23}}{\hat{I}_{23}{}^2 - \hat{I}_{22}\hat{I}_{33}}\right) \quad (19)$$

Substituting these equations into (17), $r_{max}(\alpha)$ is obtained as

$$r_{max}(\alpha) = -\frac{1}{2}(\alpha - \hat{\alpha})^2 \left(\hat{I}_{11} + Q^2\hat{I}_{22} + P^2\hat{I}_{33} + 2Q\hat{I}_{12} + 2P\hat{I}_{13} + 2PQ\hat{I}_{23}\right) \quad (20)$$

where the terms P and Q are given as below.

$$P = \frac{\hat{I}_{22}\hat{I}_{13} - \hat{I}_{12}\hat{I}_{23}}{\hat{I}_{23}{}^2 - \hat{I}_{22}\hat{I}_{33}},$$

$$Q = \frac{\hat{I}_{12}\hat{I}_{33} - \hat{I}_{13}\hat{I}_{23}}{\hat{I}_{23}{}^2 - \hat{I}_{22}\hat{I}_{33}}.$$

Equation (20) can be rewritten as

$$r_{max}(\alpha) = -\tfrac{1}{2}(\alpha - \hat{\alpha})^2 T_\alpha \quad (21)$$

where

$$T_\alpha = \hat{I}_{11} + Q^2\hat{I}_{22} + P^2\hat{I}_{33} + 2Q\hat{I}_{12} + 2P\hat{I}_{13} + 2PQ\hat{I}_{23} \quad (22)$$

The other maximized relative log likelihood functions, $r_{max}(\beta)$ and $r_{max}(\sigma)$, are obtained in a similar manner and are given as

$$r_{max}(\beta) = -\tfrac{1}{2}(\beta - \hat{\beta})^2 T_\beta \quad (23)$$

where

$$T_\beta = S^2\hat{I}_{11} + \hat{I}_{22} + R^2\hat{I}_{33} + 2S\hat{I}_{12} + 2RS\hat{I}_{13} + 2R\hat{I}_{23} \quad (24)$$

$$R = \frac{\hat{I}_{11}\hat{I}_{23} - \hat{I}_{12}\hat{I}_{13}}{\hat{I}_{13}{}^2 - \hat{I}_{11}\hat{I}_{33}}$$

$$S = \frac{\hat{I}_{12}\hat{I}_{33} - \hat{I}_{13}\hat{I}_{23}}{\hat{I}_{13}{}^2 - \hat{I}_{11}\hat{I}_{33}}$$

and

$$r_{max}(\sigma) = -\tfrac{1}{2}(\sigma - \hat{\sigma})^2 T_\sigma \tag{25}$$

where

$$T_\sigma = U^2\hat{I}_{11} + V^2\hat{I}_{22} + \hat{I}_{33} + 2UV\hat{I}_{12} + 2U\hat{I}_{13} + 2V\hat{I}_{23} \tag{26}$$

$$U = \frac{\hat{I}_{22}\hat{I}_{13} - \hat{I}_{12}\hat{I}_{23}}{\hat{I}_{12}^{\ 2} - \hat{I}_{11}\hat{I}_{22}}$$

$$V = \frac{\hat{I}_{11}\hat{I}_{23} - \hat{I}_{12}\hat{I}_{13}}{\hat{I}_{12}^{\ 2} - \hat{I}_{11}\hat{I}_{22}}$$

4.2. Likelihood Confidence Intervals

The maximized log-relative likelihood functions for the individual parameters are quadratic in nature and closed form likelihood interval of each of the parameters can be obtained. The limits of $100p\%$ likelihood interval for α is the solution to the equation

$$r_{max}(\alpha) = -\frac{1}{2}(\alpha - \hat{\alpha})^2 T_\alpha = \log p$$

This solution is

$$\alpha = \hat{\alpha} \pm \sqrt{\frac{-2\log p}{T_\alpha}}$$

Therefore, the $100p\%$ likelihood for α is

$$\hat{\alpha} - \sqrt{\frac{-2\log p}{T_\alpha}} \le \alpha \le \hat{\alpha} + \sqrt{\frac{-2\log p}{T_\alpha}} \tag{27}$$

Similarly, the respective $100p\%$ likelihood for β and σ are

$$\hat{\beta} - \sqrt{\frac{-2\log p}{T_\beta}} \le \beta \le \hat{\beta} + \sqrt{\frac{-2\log p}{T_\beta}} \tag{28}$$

and

$$\hat{\sigma} - \sqrt{\frac{-2\log p}{T_\sigma}} \le \sigma \le \hat{\sigma} + \sqrt{\frac{-2\log p}{T_\sigma}} \tag{29}$$

The quantities $T_\alpha > 0$ $T_\beta > 0$, and $T_\sigma > 0$ are defined as in equations (22), 24 and (26), respectively.

5. Simulation

In this study, data simulated from a simple linear regression model in (3) was used. The parameters α, β and σ were arbitrarily chosen to represent the true values of the population parameters. For the purpose of simulation study the values for the predictor variable x were generated from the uniform distribution. With parameter values fixed at $(\alpha, \beta, \sigma) = (7{,}13{,}2)$, four datasets were simulated for $n = 50, 250, 500, 1000$ and the respective 95% approximate likelihood confidence intervals for the parameters α, β and σ were computed and are shown in table 1 below.

Table 1. *The 95% approximate likelihood intervals for α, β and σ when $n = 50, 250, 500,$ and 1000.*

n	α	β	σ
50	5.3907, 7.9058	9.4551, 15.0463	1.2325, 2.5702
250	6.1561, 7.4558	12.2513, 13.7473	1.4357, 2.0556
500	6.5381, 7.4939	12.0326, 13.7473	1.8371, 2.3561
1000	6.6707, 7.3047	12.2423, 13.3959	1.7448, 2.0790

Further, the graph of $R_{max}(\alpha, \beta)$ was plotted for appropriate values of α and β. As shown in figure 1. Figure 2 shows the 75%, 50% and 10% contour likelihood regions for α and β To show the accuracy in estimating α, β and σ alone, the graphs of $R_{max}(\alpha), R_{max}(\beta)$and $R_{max}(\sigma)$ are plotted as they are in figures 3, 4 and 5, respectively.

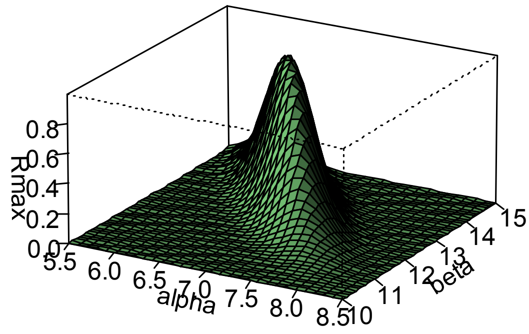

Figure 1. *The graph of the maximized relative likelihood function of α and β.*

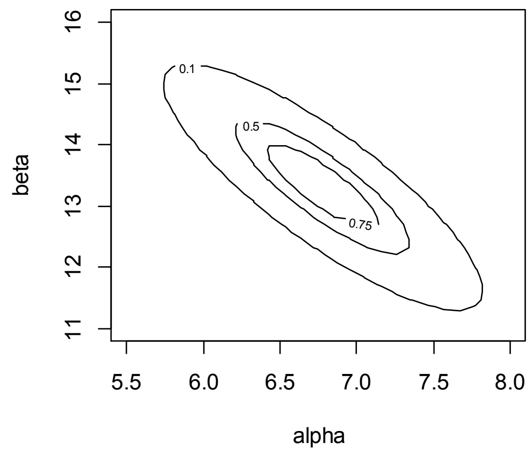

Figure 2. *The 0.1, 0.5 and 0.75 contour likelihood regions for α and β.*

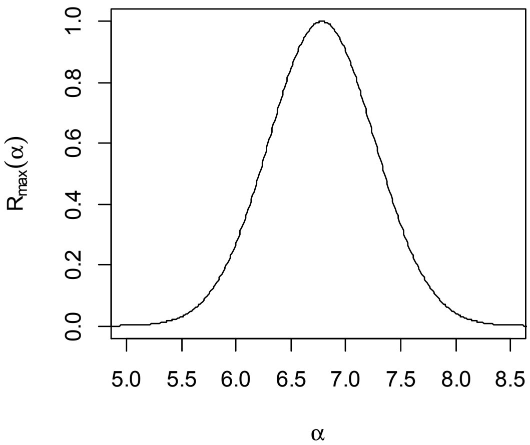

Figure 3. *A plot of $R_{max}(\alpha)$.*

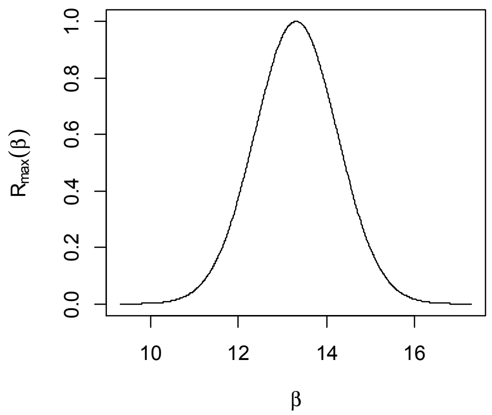

Figure 4. *A plot of* $R_{max}(\beta)$.

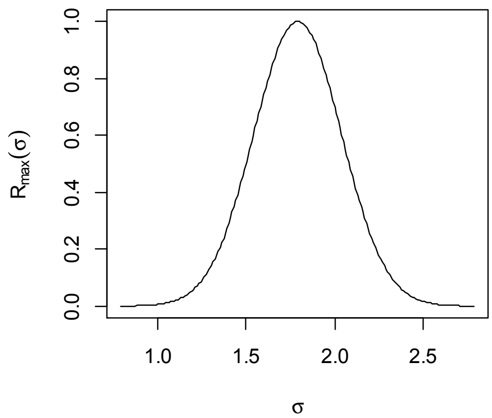

Figure 5. *A plot of* $R_{max}(\sigma)$.

The coverage probability achieved by the approximate 95% confidence intervals for the three parameters were calculated for 500 repeated simulations each of size $n = 250$ and the summary statistics for the coverage probability are presented in table 2.

Table 2. *Summary statistics for coverage probabilities of the approximate confidence intervals.*

Summary statistics	σ	α	β
Min	0.9183	0.9160	0.9060
Mean	0.9447	0.9459	0.9456
Max	0.9673	0.9760	0.9760
Std deviation	0.01076	0.01001	0.01003

6. Discussion

The derivation of the normal approximation to the Cauchy likelihood model using the Taylor series expansion has been presented. The relevance of this approximation is that it is used to express the maximized relative likelihood functions of one or two of the three parameters explicitly as a function of the data. The maximized log-relative likelihood functions for α, β and σ are quadratic in nature and therefore the closed form end points of the approximate likelihood intervals in (27), (28) and (29) are easy to derive and compute.

The effect of the sample size is seen on the accuracy of the approximate 95% confidence intervals for α, β and σ. In table 3 it can be noted that for $n = 50$ the intervals for the interval estimates are wider for all parameters but they are narrower as the sample size increased to $n = 1000$ as is expected. A more detailed evaluation of the accuracy of the approximate 95% confidence intervals is provided in table 2. The minimum, mean and maximum values of the 500 computed coverage probabilities indicate that the actual coverage probabilities for these approximate confidence intervals are very close to 95%. The standard deviation of the computed values of coverage probability for confidence interval for each parameter is low, presenting additional evidence of reasonable accuracy.

Table 3. *The length of approximate likelihood intervals.*

n	α	β	σ
50	2.5151	5.5912	1.3377
250	1.2997	1.496	0.6199
500	0.9558	1.7147	0.519
1000	0.634	1.1536	0.3342

The accuracy in jointly estimating α and β was examined by plotting the graph of the maximized relative function and contour regions for α and β as shown in figures 1 and 2, respectively. The plot of the maximized relative likelihood for α and β is clearly seen to be analogous to a surface sitting on a mountain. A pair of parameters (α, β) that is close to the peak of the likelihood graph is more plausible than those which are far from the peak. The relative likelihood surface is steep and has a sharp peak and as a result produces narrower likelihood regions in figure 2. This indicates good accuracy on the estimates of the regression coefficients α and β. Even within the 10% likelihood region, these values are very close to the true parameters,$(7,13)$. The joint evaluation of the accuracy of the estimates of regression coefficients is not possible with the traditional Wald interval estimation method but is important because the joint likelihood function is not orthogonal in terms of the parameters. This paper has not considered a data-based comparison of the Wald method and the likelihood normal approximation, but the Wald method is expected to perform poorly than the likelihood normal approximation method when the sample size is small. The normality assumption requires very large samples for it to be valid (see [9])

7. Conclusion

It is generally recognized that non-Normal samples occur very frequently in practice. This study has considered the construction of approximate likelihood regions and confidence intervals for the parameters of a simple linear regression model with the errors assumed to follow a Cauchy distribution. These likelihood regions and intervals were constructed by applying Taylor series expansion to obtain the normal approximation to the Cauchy likelihood. An attractive feature about this approximation is that it is possible to obtain the relative likelihood function of one parameter and the two regression coefficients in terms of the data, and hence made it possible to obtain closed form of the limits of the confidence intervals for individual parameters and construct the likelihood confidence region for the regression coefficients. By simulation we have demonstrated that the approximate confidence intervals for all the three parameters have short lengths and that their coverage probabilities are close to 95% (see table 2); indicating that they are accurate. Future research may be to extend the likelihood normal approximation method to a multiple linear regression model with Cauchy errors and construct confidence intervals for the regression coefficients and the standard deviation. The idea of normal approximation can also be applied to other nonnormal error distributions apart from the Cauchy distribution for the purpose of constructing the likelihood intervals.

References

[1] Akkaya, A. D. and Tiku, M. L. (1999), Estimating parameters in autoregressive models in non- normal situations: Asymmetric innovations. Communications in Statistics Theory and Methods. 30. 517-536.

[2] Bartolucci, F. and Scaccia, L. (2005). The Use of Mixtures for Dealing with Non-normal Regression Errors. Computational Statistics and Data Analysis. 48, 821-834.

[3] Bian, G., Tiku, L. M., 1997. Bayesian inference based on robust priors and MML estimators: Part I, symmetric location-scale distributions. Statistics. 29, 317-345.

[4] Bianco, A. M., Gercia, B. M. and Victor, J. (2005). Robust Estimation for Linear Regression with Asymmetric Errors. Canadian Journal of Statistics. 33: 511-528.

[5] Fernandez, C. and Steel, M. F. J. (1998). On Bayesian Modeling of Fat Tails and Skewness. Journal of the American Statistica Association. 93: 359-371.

[6] Fernandez, C. and Steel, M. F. J. (1999). Multivariate Student-t regression models: Pitfalls and inference. Biometrika. 86: 153-167

[7] Goria, M. (1978). Fractional Absolute Moments of the Cauchy Distribution, Quaderni di Statistica e Matematica Applicataalle Scienze Economico-Sociali, University of Trento, Trento. 1: 89-96.

[8] Geweke, J. (1993). Bayesian Treatment of the Independent Student-t Linear Model. Journal of Applied Econometrics. 8: 19-40.

[9] Howard, M.(2018). Comparison of the Performance of Simple Linear Regression and Quantile Regression with Non-Normal Data: A Simulation Study. *(Doctoral dissertation)*.

[10] Howlader, H. A., and Weiss, G. (1988). On Bayesian Estimation of the Cauchy Parameters. The Indian Journal of Statistics, Series B, 50: 350-361.

[11] Islam, M. Q., Tiku, M. L. and Yildirim, F. (2001). Non-normal Regression. I. Skew Distributions, Communications in Statistics - Theory and Methods. 30: 993-1020.

[12] Johnson, N. L., Kotz, S. and Balakrishnan, N. (2004). Continuous Univariate Distributions. Vol. 1, 2nd edition, John Wiley & sons Inc, New York.

[13] Kadiyala, K. R. and Murthy, K. S. R. (1977). Estimation of Regression Equation with Cauchy Disturbances, Canadian Journal of Statistics. 5: 111-120.

[14] Kalbfleisch, L. G. (1985). Probability and Statistical Inference. 2nd edition, Springer-Verag, New York.

[15] Koutrouvelis, I. A. (1981). Estimation of Location and Scale in Cauchy Distributions Using the Empirical Characteristics Function. Biometrika. 69: 205-213.

[16] Lawless, J. F. and Singhal, K. (1978). Effective Screening of Non-normal regression Models. Biometrics. 34: 318-327.

[17] Mahdizadeh M. and Zamanzade (2019). Goodness-of-fit testing for the Cauchy Distribution with Application to Financial Modeling, Journal of King Saud University-Science. 31: 1167-1174.

[18] Stuart, A. and Ord, J. K. (1994). Kendall's Advanced Theory of Statistics. 6th edition, Vol. 1, Distribution Theory, Edward Anorld, London.

[19] Tiku, M. L., Wong W. K., and Bian, G. (1999), Estimating parameters in autoregressive models in non-normal situations. Communications in Statistics - Theory and Methods. 28: 315-341.

[20] Tiku, M. L., Islam, M. Q., and Selçuk, A. S. (2001). Non-normal Regression. II. Skew Distributions. Communications in Statistics - Theory and Methods. 30: 1021-1045.

[21] Wong, K. W. and Bian, G. (2004). Robust Estimation of Multiple Regression Model with Non-normal Error: Symmetric Distribution. Monash Economics Working Papers. 09: 1-17.

[22] Wong, K. W. and Bian, G. (2000). Robust Bayesian Inference in Capital Asset Pricing Model. Journal of Applied Mathematics & Decision Sciences, 4: 65-82.

[23] Xiang L, Wanling W., Ecosse L. L., Tien and Wong (2012). Are Linear Regression Techniques Appropriate for Analysis When the Dependent Variable is not Normally Distributed? Investigative Ophthalmology & Visual Science Journal, 53: 3082-3083.