

## RESEARCH

## Open Access

# Estimation of a linear model with two-parameter symmetric platykurtic distributed errors

Asrat Atsedeweyn Andargie\* and K Srinivasa Rao

\* Correspondence: [asrat07@gmail.com](mailto:asrat07@gmail.com)  
Department of Statistics, Andhra  
University, Visakhapatnam 530003,  
India

## Abstract

**Purpose:** A linear regression model with Gaussian-distributed error terms is the most widely used method to describe the possible relationship between outcome and predictor variables. However, there are some drawbacks of Gaussian errors such as the distribution being mesokurtic. In many practical situations, the variables under study may not be mesokurtic but are platykurtic. Hence, to analyze this sort of platykurtic variables, a multiple regression model with symmetric platykurtic (SP) distributed errors is needed. In this paper, we introduce and develop a multiple linear regression model with symmetric platykurtic distributed errors for the first time.

**Methods:** We used the methods of ordinary least squares (OLS) and maximum likelihood (ML) to estimate the model parameters. The properties of the ML estimators with respect to the symmetric platykurtic distributed errors are discussed. The model selection criteria such as Akaike information criteria (AIC) and Bayesian information criteria (BIC) for the models are used. The utility of the proposed model is demonstrated with both simulation and real-time data.

**Results:** A comparative study of symmetric platykurtic linear regression model with the Gaussian model revealed that the former gives good fit to some data sets. The results also revealed that ML estimators are more efficient than OLS estimators in terms of the relative efficiency of the one-step-ahead forecast mean square error.

**Conclusions:** The study shows that the symmetric platykurtic distribution serves as an alternative to the normal distribution. The developed model is useful for analyzing data sets arising from agricultural experiments, portfolio management, space experiments, and a wide range of other practical problems.

**Keywords:** Maximum likelihood; Multiple linear regression model; Simulation; Symmetric platykurtic distribution

## Introduction

Regression analysis is one of the most commonly used statistical methodologies in many branches of science and engineering used for discovering functional relationships between variables. The most typical example of regression analysis is multiple linear regression modeling, which is used for predicting values of one or more response variables from any factor of interest, the independent variables. It has received applications in almost every area of science, engineering, and medicine. Comprehensive accounts of the theory and applications of the linear regression model are discussed in Seber [1], Montgomery et al. [2], Grob [3], Sengupta and Jammalamadaka [4], Seber and Lee [5], Weisberg [6], and Yan and Su [7]. This technique is usually based on a statistical

model in which the error terms are assumed to be independent and identically distributed random variables, whose distribution is considered to be multivariate normal with a zero mean vector and a positive definite covariance matrix [8]. However, in many disciplines, scientific research based on empirical studies or theoretical reasoning provided support for the presence of skewness or heavy tails in the distribution of the error terms. The departures from normality may be caused also by the presence of outlying values in the responses. Examples can be found, amongst others, in Fama [9] and Sutton [10]. For these reasons, several researchers proposed to perform multivariate regression analysis using a model that assumes a different parametric distribution family for the error terms.

Zeckhauser and Thompson [11] studied on a linear regression model with power distributions. Zellner [12] and Sutradhar and Ali [13] studied on a regression model with a multivariate  $t$  error variable. Tiku et al. [14-16] investigated a linear regression model with symmetric innovations, discussed a first-order autoregressive model with symmetric innovations, and presented a linear model with  $t$  distribution, respectively. Sengupta and Jammalamadaka [17] studied on linear models. Liu and Bozdogan [18] studied on power exponential (PE) multiple regression. Wong and Bian [19,20] studied on multiple regression coefficients in a linear model with errors being Student's  $t$  distribution and a linear regression model with underlying distribution being a generalized logistic distribution, respectively. Liu and Bozdogan [21] studied on multivariate regression models with PE random errors under various assumptions. Soffritti and Galimberti [22] discussed a multivariate linear regression model under the assumption that the error terms follow a finite mixture of normal distributions. Jafari and Hashemi [23] studied on linear regression with the error term of skew-normal distribution. Jahan and Khan [24] investigated the  $g$ -and- $k$  distribution as the underlying assumption for the distribution of error in a simple linear regression model. Bian et al. [25] studied a multiple linear regression model with underlying Student's  $t$  distribution.

No serious attempt is made to develop and analyze multiple regression models with symmetric platykurtic (SP) errors. For this reason, to achieve more flexibility in statistical modeling and model selection, and to robustify many multiple statistical procedures, the purpose of this paper is to introduce and develop a multivariate linear regression model for conditions in which the distribution of error terms is assumed to be independent and identically distributed SP random errors with mean 0 and constant variance  $\sigma^2$ .

### Model description

The multiple regression model assumes a linear (in parameters) relationship between a dependent variable  $Y = (y_1, y_2, \dots, y_n)'$  and a set of independent variables  $X_i' = (x_{i0}, x_{i1}, x_{i2}, \dots, x_{ik})$ , where the first regressor  $x_{i0} = 1$  is a constant and  $i = 1, 2, \dots, n$ . The model of interest is the standard linear regression model of the following form:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (i = 1, 2, \dots, n), \quad (1)$$

where  $y_i$  is an observed dependent variable,  $x_{ik}$  are observed independent variables,  $\beta_0, \beta_1, \dots, \beta_k$  are unknown regression coefficients to be estimated, and  $\varepsilon_i$  are independently

and identically distributed. Using linear algebra notation, model (1) may be written alternatively in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim SP(0, \sigma^2 I). \tag{2}$$

In this model,

$$\mathbf{y}_{n \times 1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X}_{n \times (k+1)} = \begin{pmatrix} x_{10} & x_{11} & \cdot & \cdot & \cdot & x_{1k} \\ x_{20} & x_{21} & \cdot & \cdot & \cdot & x_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{n0} & x_{n1} & \cdot & \cdot & \cdot & x_{nk} \end{pmatrix}, \boldsymbol{\beta}_{(k+1) \times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix}, \boldsymbol{\epsilon}_{n \times 1} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_n \end{pmatrix}, \tag{3}$$

where  $\mathbf{y}$  is a column vector of  $n$  elements,  $\mathbf{X}$  is an  $n \times (k + 1)$  ( $k + 1 < n$ ) nonrandom design matrix of covariates (with its first column having all elements equal to 1, the second column being filled by the observed values of  $x_1$ , the  $(k + 1)$ th column being filled by the observed values of  $x_k$ ),  $\boldsymbol{\beta}$  is a column vector of the  $(k + 1)$  elements,  $\sigma$  is an unknown scale parameter, and  $\boldsymbol{\epsilon}$  is an  $n \times 1$  column vector of error terms with zero mean and constant variance  $\sigma^2 I$ .

**Model assumptions**

In order to complete the description of the model, some assumptions about the nature of the errors are necessary. It is assumed that the errors are independent and identically distributed (i.i.d.) random variables whose distribution is assumed to be two-parameter SP rather than normal, with zero mean and a positive definite variance-covariance matrix  $\sigma^2 I$  of dimension  $n \times n$ , that is

$$\epsilon_i \sim MSP(0, \sigma^2 I). \tag{4}$$

These assumptions are summarized in the matrix vector form as

$$E(\boldsymbol{\epsilon}) = \mathbf{0}, Cov(\boldsymbol{\epsilon}) = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2 I_{n \times n}, \tag{5}$$

where the notation  $E$  stands for the expected value and  $Cov$  represents an  $n \times n$  variance-covariance matrix. The vector  $\mathbf{0}$  is a column vector with  $n$  zero elements, and  $I$  is an identity matrix of order  $n \times n$ . The parameter  $\sigma^2$  is unspecified, along with the vector parameter  $\boldsymbol{\beta}$ . The elements of  $\boldsymbol{\beta}$  are real-valued, while  $\sigma$  is positive. The covariates are either nonrandom or are independent of the errors.  $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ , and  $Cov(\mathbf{y}) = \sigma^2 I$ . We shall use the triplet  $(\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2 I)$  for linear model (2).

**Article headings**

The manuscript is organized in eight sections. The ‘Introduction’ section frames the objective of the paper and reviews related literatures. In the ‘Properties of the two parameter symmetric platykurtic distribution’ section, we introduced the two-parameter SP distribution, in notation  $SP(\mu, \sigma)$ . We derived the maximum likelihood (ML) estimators in the ‘Maximum likelihood estimation of the model parameters’ section. In order to obtain numerical solutions to the ML estimate problem, the Newton–Raphson (NR) iterative method has been used. In the ‘Properties of the estimators through simulation study’ section, we show the asymptotic properties of the estimators. In the ‘Least

squares estimation of the model parameters' section, OLS estimation for the model parameters is studied. Comparison of MLE with OLS estimators and that of the proposed model with the Gaussian model are done in the 'Comparative study of the model' section. The 'Application of the model' section demonstrates the usefulness of the present model on real data. Finally, the 'Summary and conclusions' section concludes the paper.

### Properties of the two-parameter symmetric platykurtic distribution

Even though the sample frequency curve has a symmetric and bell shape, it often has fat tails than normal and the almost universally used Gaussian distribution may badly fit the fat tails. The family of symmetric platykurtic distributions can best model these features. The origin and genesis of this distribution is given by Srinivasa Rao et al. [26] for analyzing statistical data that arise from biological, sociological, agricultural, and environmental experiments. It became popular thereafter and has received considerable attention in the present time in modeling image segmentation and economic and financial data as a generalization of normal distribution (e.g., [27]). The probability density function (pdf) of such a family of distributions is

$$f_r(y/\mu, \sigma, \gamma) = \frac{\left[2\gamma + \left(\frac{y-\mu}{\sigma}\right)^2\right]^\gamma e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}}{\sum_{j=0}^{\gamma} \binom{\gamma}{j} (2\gamma)^{\gamma-j} 2^{j+\frac{1}{2}} \Gamma\left(j + \frac{1}{2}\right) \sigma} \quad (6)$$

The distribution depends on three parameters  $\mu$ ,  $\sigma$ , and  $\gamma$ . These parameters can be interpreted as follows:

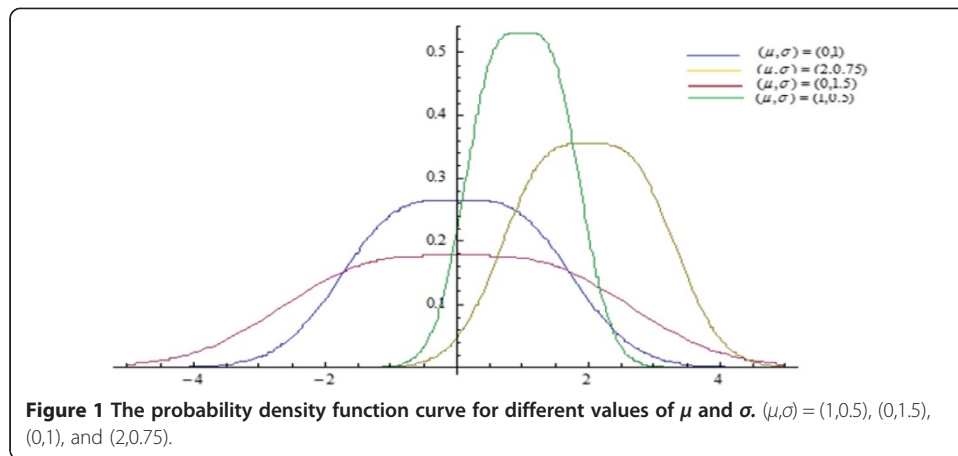
- $\mu$  is a real number and may be thought of as a location measure.
- $\sigma$  is positive and measures the dispersion or the scale of the distribution.
- $\gamma$  is a kurtosis parameter which determines the shape of the distribution, taking values  $\gamma = 0, 1, 2, \dots, n$ . If  $\gamma = 0$ , we retrieve the normal distribution  $N(\mu, \sigma^2)$ . If we take  $\gamma = 1$ , we get a two-parameter symmetric platykurtic distribution.

In this section, we briefly present a two-parameter symmetric platykurtic distribution which is appropriate for data having kurtosis  $\beta = 2.52$ . A random variable  $Y$  is said to follow a two-parameter SP distribution if the density function of  $Y$  is

$$f(y/\mu, \sigma) = \frac{\left[2 + \left(\frac{y-\mu}{\sigma}\right)^2\right] e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}}{3\sigma\sqrt{2\pi}}, y \in \mathfrak{R}, \quad (7)$$

where  $-\infty < \mu < \infty$  and  $\sigma > 0$  are location and scale parameters, respectively. We denote this by  $Y \sim \text{SP}(\mu, \sigma^2)$ . The various shapes of the frequency curves are shown in Figure 1.

Some important properties of the random variable  $Y$  in a univariate context and those which are needed for the simulation of the variate are as follows:



1. The distribution function of the random variable  $Y$  specified by the probability density function (7) is given by

$$F_Y(y) = \int_{-\infty}^y \frac{\left[2 + \left(\frac{t-\mu}{\sigma}\right)^2\right] e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}}{3\sigma\sqrt{2\pi}} dt$$

which, on simplification, reduces to

$$F_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt - \frac{(y-\mu)e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}}{3\sigma\sqrt{2\pi}} = F_0(y; \mu, \sigma) - F_1(y; \mu, \sigma), \quad (8)$$

where  $F_0(y; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$ ,  $y \in \mathfrak{R}$ , is the distribution function of the normal random variable with mean  $\mu$  and variance  $\sigma$ , while  $F_1(y) = \frac{(y-\mu)}{3\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$ ,  $y \in \mathfrak{R}$  is the nondistribution function (cannot be a cumulative density function) since it is negative for  $Y > \mu$ .

2. Numerical approximations for the two-parameter symmetric platykurtic cumulative distribution function (CDF): Following Marsaglia's [28] approximation for standard normal distribution who suggested a simple algorithm based on the Taylor series expansion, we have also approximated the values  $F(y; \mu, \sigma)$  as follows. For standard normal distribution with arbitrary precision,  $\Phi(y) = \frac{1}{2} + \phi(y) \left( y + \frac{y^3}{3} + \frac{y^5}{3.5} + \frac{y^7}{3.5.7} + \dots \right)$  where  $\phi$  and  $\Phi$  is the pdf and CDF of the normal distribution with mean  $\mu$  and variance  $\sigma$ . Accordingly, after little algebra, the standard symmetric platykurtic CDF,  $F(y)$ , is approximated by

$$F(y) = \frac{1}{2} + f(y) \left( \frac{-y}{3} + \sum_{i=1}^n \frac{y^i}{i!!} \right) \quad (9)$$

where  $n = 1, 3, 5, \dots, n$  and  $n!!$  denotes the double factorial that is the product of every odd number from 1 to  $n$ .

3. The cumulant-generating function is the logarithm of the moment-generating function:

$$g(t; \mu, \sigma^2) = \mu t + \frac{1}{2} \sigma^2 t^2 + \ln \left[ 1 + \frac{(\sigma t)^2}{3} \right]. \quad (10)$$

The cumulants  $k_n$  are extracted from the cumulant-generating function via differentiation (at zero) of  $g(t)$ . That is, the cumulants appear as the coefficients in the Maclaurin series of  $g(t)$ :

$$k_1 = g'(0) = \mu, k_2 = g''(0) = \frac{5\sigma^2}{3}, k_3 = g'''(0) = \dots = k_n = g^{(n)}(0) = 0 \quad (11)$$

That is, the first two cumulants are equal to the mean  $\mu$  and the variance  $\frac{5\sigma^2}{3}$  of the two-parameter symmetric platykurtic distribution, respectively, whereas all higher-order cumulants are equal to zero.

4. Hazard rate function of the distribution: The hazard function  $h(y; \mu, \sigma)$  of the two-parameter symmetric platykurtic distribution used in this paper is utilized to characterize life phenomena and can be written as

$$h(y; \mu, \sigma) = \frac{\left( 2 + \left( \frac{y-\mu}{\sigma} \right)^2 \right) \phi \left( \frac{y-\mu}{\sigma} \right)}{3\sigma Q(y) + (y-\mu) \phi \left( \frac{y-\mu}{\sigma} \right)}, \quad (12)$$

where  $\phi$  is the pdf of the normal distribution with mean  $\mu$  and variance  $\sigma$ , whereas the  $Q$ -function  $Q(y)$  is the complement of the standard normal CDF,  $Q(y) = 1 - \Phi(y)$ .

Recently, it was observed by Gupta and Gupta [29] that the reversed hazard function plays an important role in the reliability analysis. The reversed hazard function of the two-parameter SP( $\mu, \sigma$ ) is

$$r(y; \mu, \sigma) = \frac{f(y; \mu, \sigma)}{SP(y; \mu, \sigma)} = \frac{\left( 2 + \left( \frac{y-\mu}{\sigma} \right)^2 \right) \phi \left( \frac{y-\mu}{\sigma} \right)}{3\sigma \Phi \left( \frac{y-\mu}{\sigma} \right) - (y-\mu) \phi \left( \frac{y-\mu}{\sigma} \right)}. \quad (13)$$

It is well known that the hazard function or the reversed hazard function uniquely determines the corresponding probability density function.

5. Entropy: The entropy for a two-parameter symmetric platykurtic distribution random variable  $y$  with probability density function  $f(y)$  on the real line is defined by

$$h(y) = \frac{-529}{1,500} + \ln(3\sigma\sqrt{2\pi}). \quad (14)$$

It can be recalled that the entropy for normal distribution is  $\frac{1}{2} \ln(2\pi e \sigma^2)$ . If  $f$  and  $g$  are the probability distributions of symmetric platykurtic and

normal distributions, respectively, then the relative entropy  $D(f||g)$  from  $f$  to  $g$  is

$$D(f||g) = \int_{-\infty}^{\infty} f(y) \log \frac{f(y)}{g(y)} dy = 0.7088. \tag{15}$$

This gives us a measure of something like the distance between the two probability distributions, in the sense that the relative entropy is always positive, is zero if and only if the two distributions are the same, and increases as the distributions diverge. Some of the more important properties of the SP distribution are summarized in the Appendix.

**Maximum likelihood estimation of the model parameters**

In this section, we consider the homoscedastic regression model  $(Y, X\beta, \sigma^2 I)$  with  $\varepsilon \sim SP(0, \sigma^2 I)$ . The unknown parameters of this model are the coefficient vector  $\beta$  and the error variance  $\sigma^2$ . We deal with the problem of ML estimation of these parameters, which requires some fairly intricate mathematics, from the observables  $Y$  and  $X$ . The MLEs of  $\beta$  and  $\sigma^2$  are the parameter values that maximize the likelihood function:

$$\begin{aligned} L(y; \beta, \sigma) &= \prod_{i=1}^n f(y_i/X\beta, \sigma^2 I) = \prod_{i=1}^n \frac{\left[2 + \left(\frac{y_i - x_i' \beta}{\sigma}\right)^2\right] e^{-\frac{1}{2} \left(\frac{y_i - x_i' \beta}{\sigma}\right)^2}}{3\sigma\sqrt{2\pi}} \\ &= (3\sigma\sqrt{2\pi})^{-n} \prod_{i=1}^n \left[ \left(2 + \left(\frac{y_i - x_i' \beta}{\sigma}\right)^2\right) e^{-\frac{1}{2} \left(\frac{y_i - x_i' \beta}{\sigma}\right)^2} \right]. \end{aligned} \tag{16}$$

The log-likelihood function in the i.i.d. case, ignoring the additive constants, equals

$$\begin{aligned} \ell &= \ln L(y; \beta, \sigma) = -\frac{n}{2} \ln \sigma^2 + \sum_{i=1}^n \ln h\left(\frac{y_i - x_i' \beta}{\sigma}\right) \\ \text{where } h\left(\frac{y_i - x_i' \beta}{\sigma}\right) &= \left[2 + \left(\frac{y_i - x_i' \beta}{\sigma}\right)^2\right] e^{-\frac{1}{2} \left(\frac{y_i - x_i' \beta}{\sigma}\right)^2}. \end{aligned} \tag{17}$$

The unknown parameters of this model are the coefficient vector  $\beta$  and the error variance  $\sigma^2$ . The MLE is that  $\hat{\theta}_{MLE} = (\hat{\beta}, \hat{\sigma}^2)$  which maximizes the log-likelihood. Taking the partial derivatives of the log of the likelihood with respect to the  $(k+1) \times 1$  vector  $\beta$  and  $n \times n$  matrix of  $\sigma^2 I$  and setting the result equal to zero will produce (18) and (19). The MLEs are the solutions of the equations:

$$\frac{\partial \ell}{\partial \beta} = -\frac{1}{\sigma} \sum_{i=1}^n x_i \left( \frac{\partial \ln h(u)}{\partial u} \right) \Bigg|_{u = \frac{y_i - x_i' \beta}{\sigma}} = 0 = \frac{1}{\sigma} \sum_{i=1}^n x_i \left( \frac{u^3}{2 + u^2} \right) \Bigg|_{u = \frac{y_i - x_i' \beta}{\sigma}} = 0 \tag{18}$$

and

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^2} \sum_{i=1}^n \left( u \frac{\partial \ln h(u)}{\partial u} \right) \Bigg|_{u = \frac{y_i - x_i' \beta}{\sigma}} = 0 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{i=1}^n \left( \frac{u^4}{2 + u^2} \right) \Bigg|_{u = \frac{y_i - x_i' \beta}{\sigma}} = 0. \tag{19}$$

Since there are no closed-form solutions to the likelihood equations, numerical methods such as the Fisher scoring or Newton–Raphson iterative method can be used to obtain the MLEs. The usual or standard procedure for implementing this solution is to use the Newton–Raphson iteration method given by

$$\theta^{(n+1)} = \theta^{(n)} - \left( H^{(n)} \right)^{-1} S^{(n)}, \theta = (\beta, \sigma^2)', \tag{20}$$

where  $H$  is the matrix of the second derivative and  $S$  is the vector of the first derivative of the log-likelihood function both evaluated at the current values of the parameter vector  $\theta$ .

Here we begin with some starting value, say  $\theta^{(0)}$ , and improve it by finding some better approximation  $\theta^{(1)}$  to the required root. This procedure can be iterated to go from a current approximation  $\theta^{(n)}$  to a better approximation  $\theta^{(n+1)}$ .

**Variance-covariance matrix**

To obtain the asymptotic variances and covariance of the estimates, it is required to construct the Hessian matrix of the log-likelihood. The negative expected values of the second-order partial derivatives of the log-likelihood equations (18) and (19) can be used to estimate the asymptotic covariance matrix of parameter estimates and can be found as follows. Each summand in the right-hand side of (19) has zero expectation. Therefore,

$$\begin{aligned} E \left[ \left( \frac{\partial \ell}{\partial \beta} \right) \left( \frac{\partial \ell}{\partial \beta} \right)' \right] &= \frac{1}{\sigma^2} \sum_{i=1}^n x_i x_i' \int_{-\infty}^{\infty} \left( \frac{\partial \ln h(u)}{\partial u} \right)^2 \Bigg|_{u = \frac{y_i - x_i' \beta}{\sigma}} \frac{1}{\sigma} h \left( \frac{y_i - x_i' \beta}{\sigma} \right) dy_i \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n x_i x_i' \int_{-\infty}^{\infty} \left( \frac{\partial \ln h(u)}{\partial u} \right)^2 h(u) du \\ &= I_{\mu} X' X, \end{aligned} \tag{21}$$

$$\text{where } I_{\mu} = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \left( \frac{\partial \ln h(u)}{\partial u} \right)^2 h(u) du = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \frac{u^6}{u^2 + 2} e^{-\frac{1}{2}u^2} du = \frac{4.934}{\sigma^2}. \tag{22}$$

Further,

$$\begin{aligned} E \left[ \left( \frac{\partial \ell}{\partial \beta} \right) \left( \frac{\partial \ell}{\partial \sigma^2} \right) \right] &= \frac{1}{2\sigma^3} \sum_{i=1}^n x_i \int_{-\infty}^{\infty} u \left( \frac{\partial \ln h(u)}{\partial u} \right)^2 \Bigg|_{u = \frac{y_i - x_i' \beta}{\sigma}} \frac{1}{\sigma} h \left( \frac{y_i - x_i' \beta}{\sigma} \right) dy_i \\ &= \frac{1}{2\sigma^3} \sum_{i=1}^n x_i \int_{-\infty}^{\infty} u \left( \frac{\partial \ln h(u)}{\partial u} \right)^2 h(u) du = 0 \end{aligned} \tag{23}$$

as the integrand in the last expression is an odd function of  $u$ , that is, making use of the fact that the integrand of the off-diagonal term is an odd function of  $u$ .



Finally,

$$\begin{aligned}
 E \left[ \left( \frac{\partial \ell}{\partial \sigma^2} \right)^2 \right] &= \frac{n^2}{4\sigma^4} + \frac{n}{2\sigma^4} \sum_{i=1}^n E \left[ \left( u \frac{\partial \ln h(u)}{\partial u} \right) \Big|_{u=\frac{y_i - x_i' \beta}{\sigma}} \right] \\
 &\quad + \frac{1}{4\sigma^4} \sum_{i=1}^n \sum_{j=1}^n E \left[ \left( u \frac{\partial \ln h(u)}{\partial u} \right) \Big|_{u=\frac{y_i - x_i' \beta}{\sigma}} \cdot \left( u \frac{\partial \ln h(u)}{\partial u} \right) \Big|_{u=\frac{y_j - x_j' \beta}{\sigma}} \right] \\
 &= \frac{n^2}{4\sigma^4} + \frac{n^2}{2\sigma^4} \int_{-\infty}^{\infty} u \frac{\partial h(u)}{\partial u} du + \frac{n^2 - n}{4\sigma^4} \left( \int_{-\infty}^{\infty} u \frac{\partial h(u)}{\partial u} du \right)^2 \\
 &\quad + \frac{n}{4\sigma^4} \int_{-\infty}^{\infty} \left( u \frac{\partial \ln h(u)}{\partial u} \right)^2 \cdot h(u) du.
 \end{aligned}$$

Simplification of the expression of this term is aided by the identity

$$\int_{-\infty}^{\infty} u \frac{\partial h(u)}{\partial u} du = -1 = \frac{n^2}{4\sigma^4} + \frac{n^2}{2\sigma^4} (-1) + \frac{n^2 - n}{4\sigma^4} (-1)^2 + \frac{n}{4\sigma^4} \int_{-\infty}^{\infty} \left( u \frac{\partial \ln h(u)}{\partial u} \right)^2 \cdot h(u) du = nI_{\sigma^2} \tag{24}$$

$$\begin{aligned}
 \text{where } I_{\sigma^2} &= \frac{1}{4\sigma^4} \left[ \int_{-\infty}^{\infty} \left( u \frac{\partial \ln h(u)}{\partial u} \right)^2 \cdot h(u) du - 1 \right] = \frac{1}{4\sigma^4} \left[ \int_{-\infty}^{\infty} \frac{u^8}{u^2 + 2} e^{-\frac{1}{2}u^2} du - 1 \right] \\
 &= \frac{1}{4\sigma^4} [27.731 - 1] = \frac{26.731}{4\sigma^4} = \frac{6.68275}{\sigma^4}.
 \end{aligned} \tag{25}$$

Therefore, the information matrix for  $\theta = (\beta, \sigma^2)'$  is

$$I(\theta) = \begin{pmatrix} I_{\mu} X'X & 0 \\ 0 & nI_{\sigma^2} \end{pmatrix}. \tag{26}$$

The information for  $\beta$  is  $I_{\mu} X'X$ . Therefore, the design issues can be addressed with reference to the matrix  $X'X$  - just as in the normal case. The scalar  $I_{\mu}$  is equal to  $\frac{4.934}{\sigma^2}$  (greater than  $\sigma^{-2}$  when the components of  $\mathbf{y}$  are normally distributed), whereas the scalar  $I_{\sigma^2}$  is equal to  $\frac{6.68275}{\sigma^4}$  (greater than  $\sigma^{-4}$  when the components of  $\mathbf{y}$  are normally distributed). The large sample variances and covariance of the estimates can be approximated by inverting the usual symmetric information matrix (26):

$$\text{Var}(\theta) = [I(\theta)]^{-1} = \begin{pmatrix} (I_{\mu} X'X)^{-1} & 0 \\ 0 & (nI_{\sigma^2})^{-1} \end{pmatrix} \tag{27}$$

Thus, the square root of the elements on the diagonal of this matrix will give us the standard errors associated with the coefficients.

### Simulation and results

The proposed approach was evaluated through Monte Carlo experiments in which artificial data sets were generated from model (2) using Wolfram Mathematica 9. To facilitate exposition of the method of estimation, a multiple data set with two independent variables and one dependent variable are simulated from a model with prespecified parameters for various sample sizes  $n = 100, 1,000, 3,000, 5,000, 10,000$ . (The sample size

in this case, 10,000, is relatively large, and so finite sample bias is less of an issue.) The dependent variable ( $Y$ ) is simulated from the symmetric platykurtic distribution with mean 0 and variance 1 while the two predictor  $x_1$  and  $x_2$  variables are generated from normal and lognormal distributions, respectively, with the prespecified mean and variance using the following simulation protocol:

$$\begin{aligned} x_1 &= \text{RandomReal}[\text{NormalDistribution}[10, 5]]; \\ x_2 &= \text{RandomReal}[\text{LogNormalDistribution}[0, 3]] \end{aligned} \tag{28}$$

and use as explanatory variables for the regression model. Without loss of generality, we considered the values of the model parameters as

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix} \text{ and } \begin{pmatrix} \mu \\ \sigma \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{29}$$

Summary statistics of estimations of the regression model with symmetric platykurtic distributed errors using the ML procedure are presented in Table 1.

The numerical results in Table 1 suggest that as the sample size increases, the estimates of the parameters become more precise. The ML method provides good estimates of the underlying model not only of the regression coefficients but also of the

**Table 1 Summary of ML estimation of the regression model for the simulated data**

Sample size ( $n$ )	Parameter	Estimate	Standard error	Wald 95% confidence limits		Chi-square	Pr > chi-square
				Lower	Upper		
100	$\beta_0$	4.0747	0.2515	3.5817	4.5677	262.46	<.0001
	$\beta_1$	1.9914	0.0216	1.9491	2.0337	8,510.03	<.0001
	$\beta_2$	2.9998	0.0010	2.9978	3.0019	8,263,598	<.0001
	$\sigma$	1.0668	0.0754	0.9288	1.2254		
1,000	$\beta_0$	3.8958	0.0696	3.7593	4.0323	3,129.59	<.0001
	$\beta_1$	2.0062	0.0065	1.9935	2.0189	96,198.0	<.0001
	$\beta_2$	2.9999	0.0001	2.9997	3.0000	1.408E9	<.0001
	$\sigma$	1.0242	0.0229	0.9803	1.0701		
3,000	$\beta_0$	4.0504	0.0407	3.9707	4.1301	9,916.51	<.0001
	$\beta_1$	1.9938	0.0036	1.9867	2.0009	303,055	<.0001
	$\beta_2$	2.9999	0.0000	2.9999	3.0000	3.731E9	<.0001
	$\sigma$	1.0037	0.0130	0.9786	1.0294		
5,000	$\beta_0$	4.0012	0.0315	3.9394	4.0630	16,095.1	<.0001
	$\beta_1$	2.0013	0.0028	1.9957	2.0068	506,637	<.0001
	$\beta_2$	3.0000	0.0000	3.0000	3.0000	6.34E10	<.0001
	$\sigma$	0.9988	0.0100	0.9794	1.0186		
10,000	$\beta_0$	4.0052	0.0220	3.9620	4.0483	33,060.4	<.0001
	$\beta_1$	1.9981	0.0020	1.9942	2.0019	1,031,236	<.0001
	$\beta_2$	3.0000	0.0000	3.0000	3.0000	9.55E11	<.0001
	$\sigma$	0.9850	0.0070	0.9714	0.9987		

correlation matrix. The fitted linear regression model with symmetric platykurtic distributed error terms to the simulated data, based on the sample of size 10,000, is

$$\hat{Y} = 4.0052 + 1.9981X_1 + 3.000X_2. \quad (30)$$

Standard errors of the estimates were estimated by the square root of the diagonal elements of the inverse of the Hessian of the log-likelihood function. Thus, the estimated standard errors are

$$\text{s.e.}(\hat{\beta}_0) = 0.0220, \text{s.e.}(\hat{\beta}_1) = 0.0020, \text{s.e.}(\hat{\beta}_2) = 0.000, \text{and } \text{s.e.}(\hat{\sigma}) = 0.0070. \quad (31)$$

### Properties of the estimators through simulation study

If certain regularity conditions of the density are met, the MLEs are most attractive because they possess many asymptotic or large sample properties. Derivations of the asymptotic properties require some fairly intricate mathematics. The three properties of the regular densities (moments of the derivatives of the log-likelihood) are used in establishing the properties of MLEs. The properties of the ML estimators are as follows.

#### Consistency

One of the basic properties of a good estimator is that it differs from a true value by a very small amount as  $n$  becomes large. This implies that we can reach the exact value of  $\theta$  by indefinitely increasing the sample size. Mathematically, it is expressed as

$$p \lim \hat{\theta} = \theta. \quad (32)$$

More specifically, a consistent estimator should not only be unbiased, but it should also have a variance which is as small as possible. This leads to two definitions:

$$\begin{aligned} E(\hat{\theta}) &\rightarrow^{n \rightarrow \infty} \theta \text{ and} \\ V(\hat{\theta}) &\rightarrow^{n \rightarrow \infty} 0 \end{aligned} \quad (33)$$

From (27), it is clear that the variance tends to zero as  $n \rightarrow \infty$  in each case, so we conclude that the estimators are consistent since they are composed of i.i.d. observations.

#### Asymptotic normality

Greene's derivation of the asymptotic normality of the MLE applies here. The first derivative of the log-likelihood evaluated at the MLE equals zero. So

$$S(\hat{\theta}) = 0. \quad (34)$$

Expand this set of equations in a Taylor series around the true parameters  $\theta_0$  using the mean value theorem to truncate the Taylor series at the second term:

$$S(\hat{\theta}) = S(\theta_0) + H(\bar{\theta})(\hat{\theta} - \theta_0) = 0. \quad (35)$$

The Hessian is evaluated at a point  $\bar{\theta}$  that is between  $\hat{\theta}$  and  $\theta_0$  [ $\bar{\theta} = w\hat{\theta} + (1-w)\theta_0$  for some  $0 < w < 1$ ]. We then rearrange this function and multiply the result by  $\sqrt{n}$  to obtain

$$\sqrt{n}(\hat{\theta} - \theta_0) = [-H(\bar{\theta})]^{-1}[\sqrt{n}S(\theta_0)] \quad (36)$$

Using (32), the probability limits of  $\sqrt{n}(\hat{\theta} - \theta)$  and  $\sqrt{n}(\hat{\theta} - \bar{\theta})$  go to zero because of the consistency of  $\hat{\theta}$  (i.e.,  $p \lim(\hat{\theta} - \theta) = 0$  and  $p \lim(\hat{\theta} - \bar{\theta}) = 0$ ). The second derivatives are continuous functions. Therefore, if the limiting distribution exists, then

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} [-H(\theta_0)]^{-1}[\sqrt{n}S(\theta_0)]. \quad (37)$$

By dividing  $H(\theta_0)$  and  $S(\theta_0)$  by  $n$ , we obtain

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \left[-\frac{1}{n}H(\theta_0)\right]^{-1}[\sqrt{n}\bar{S}(\theta_0)] \quad (38)$$

We may apply the Lindeberg-Levy central limit theorem to  $[\sqrt{n}\bar{S}(\theta_0)]$  because it is  $\sqrt{n}$  times the mean of a random sample. By virtue of  $V[S_i(\theta)] = -E\left[\frac{1}{n}H(\theta)\right]$ , the limiting variance of  $[\sqrt{n}\bar{S}(\theta)]$  is  $-E\left[\frac{1}{n}H(\theta)\right]$ , so

$$\sqrt{n}\bar{S}(\theta) \xrightarrow{d} N\left\{0, -E\left[\frac{1}{n}H(\theta)\right]\right\} \quad (39)$$

By virtue of  $E[S_i(\theta)] = 0, p \lim\left[-\frac{1}{n}H(\theta)\right] = -E\left[\frac{1}{n}H(\theta)\right]$ . This result is a constant matrix, so we can combine results to obtain

$$\begin{aligned} \left[-\frac{1}{n}H(\theta)\right]^{-1} \sqrt{n}\bar{S}(\theta) &\xrightarrow{d} N\left[0, \left\{-E\left[\frac{1}{n}H(\theta)\right]\right\}^{-1} \left\{-E\left[\frac{1}{n}H(\theta)\right]\right\} \left\{-E\left[\frac{1}{n}H(\theta)\right]\right\}^{-1}\right] \text{ or} \\ \sqrt{n}(\hat{\theta} - \theta) &\xrightarrow{d} N\left[0, \left\{-E\left[\frac{1}{n}H(\theta)\right]\right\}^{-1}\right]. \end{aligned} \quad (40)$$

It follows that the MLEs are asymptotically normal with asymptotic distribution:

$$\hat{\theta} \sim N[\theta, \{I(\theta)\}^{-1}]. \quad (41)$$

### Asymptotic efficiency

The information matrix forms a tool of interest to verify efficiency, *viz.* the attainment of the information limit to the variance of the estimator. An estimator whose variance is as small as the Cramer-Rao lower bound when the sample size tends to infinity is

**Table 2 Nonlinear OLS summary of residual errors**

Equation	DF model	DF error	SSE	MSE	Root MSE	R-square	Adjusted R-square
y	3	9996	16650.3	1.6657	1.2906	1.0000	1.0000

DF degrees of freedom, SSE sum of the squared errors.

called asymptotically efficient. This means that an estimator which reaches 100% efficiency only in the  $n \rightarrow \infty$  limit is called asymptotically efficient. It can be shown that the Cramer-Rao lower bound for  $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)'$  and  $\hat{\sigma}^2$ , respectively, are

$$\begin{aligned} \text{Var}(\hat{\theta}) &\geq \frac{1}{-nE\left(\frac{\partial^2 \ln f(y)}{\partial \theta^2}\right)} = (I_{\mu} X'X)^{-1} \text{ and} \\ \text{Var}(\hat{\sigma}^2) &\geq \frac{1}{-nE\left(\frac{\partial^2 \ln f}{(\partial \sigma^2)^2}\right)} = (nI_{\sigma^2})^{-1}, \end{aligned} \tag{42}$$

where  $I_{\mu}$  and  $I_{\sigma^2}$  are as defined in (22) and (25), respectively.

This means that any unbiased estimator that achieves this lower bound is efficient and no better unbiased estimator is possible. Now look back at the variance-covariance matrix (27). It is interesting to note that the variances of the estimators in the variance-covariance matrix do asymptotically coincide with the Cramer-Rao lower bound (42). This means that our MLEs are 100% asymptotically efficient. The asymptotic variance of the MLE is, in fact, equal to the Cramer-Rao lower bound for the variance of a consistent and asymptotically normally distributed estimator [30].

**Invariance**

Last, the invariance property is a mathematical result of the method of computing MLEs; it is not a statistical result as such. If it is desired to analyze a continuous and continuously differentiable function of an MLE, then the function of  $\hat{\theta}$  will, itself, be the MLE since the MLE is invariant to one-to-one transformations of  $\theta$ .

These four properties explain the prevalence of the ML technique. The second is a particularly powerful result. The third greatly facilitates hypothesis testing and the construction of interval estimates. The MLE has the minimum variance achievable by a consistent and asymptotically normally distributed estimator.

**Least squares estimation of the model parameters**

The most widely used technique for estimating the unknown regression coefficients in a standard linear regression model is undeniably the method of ordinary least squares (OLS). The least squares estimates of  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$  are the values which minimize

**Table 3 Nonlinear OLS parameter estimates**

Parameter	Estimate	Approximate standard error	t value	Approximate Pr >  t
$\beta_0$	3.9598	0.0287	138.16	<.0001
$\beta_1$	2.003412	0.00256	782.76	<.0001
$\beta_2$	2.99998	0.000015	202,020	<.0001

Number of observations: used = 10,000, missing = 0; statistics for the system: objective = 1.6652; objective  $\times N$  = 1665.2.

**Table 4 Comparison of the OLS with ML estimators of the SP regression model**

		OLS	MLE
$\hat{\beta}_0$	Bias	-0.04016	0.0109
	MSE	0.002436516	0.000914
$\hat{\beta}_1$	Bias	0.003412	-0.0022
	MSE	1.81953E - 05	1.11E - 05
$\hat{\beta}_2$	Bias	-1.7E - 05	0.0001
	MSE	5.14E - 10	2E - 08

$$SS = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2. \tag{43}$$

This leads to a closed-form expression for the estimated value of the unknown parametric vector  $\beta$ :

$$\hat{\beta} = (X^T X)^{-1} X^T y = \left( \frac{1}{n} \sum_{i=1}^n x_i x_i^T \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i y_i \right). \tag{44}$$

Tables 2 and 3 summarize the results of the OLS estimation for a sample of size 10,000 using the same simulated observations obtained in the ‘Simulation and results’ section.

From Table 3, it is observed that the OLS estimates differ significantly from the ML estimates and the ML estimators are closer to the true values of the parameters compared to the OLS estimators.

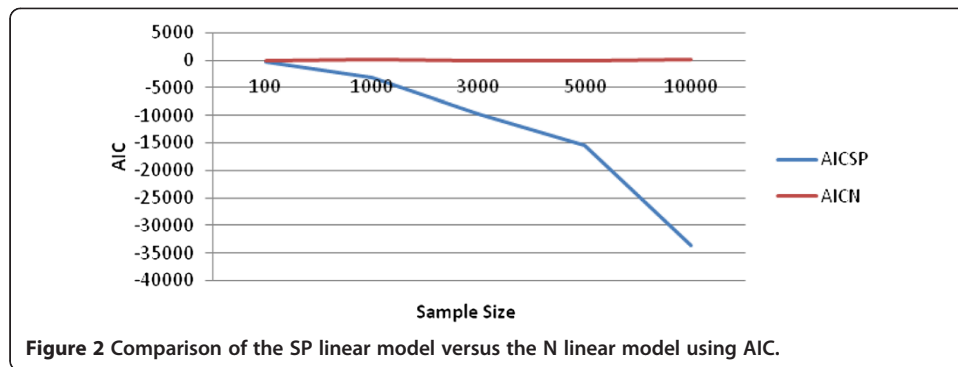
### Comparative study of the model

#### Comparison of estimators of the linear regression model

In this section, ML and OLS estimators are compared in fitting the multivariate linear model with two-parameter symmetric platykurtic error terms. One-step-ahead forecasting is commonly used to compare the performance of different models [31,32]. For

**Table 5 Information criteria and model diagnostics of ML estimators of the parameters**

Model	Sample size (n)	Information criterion		Model diagnostics
		AIC	BIC	RMSE
Symmetric platykurtic	100	-362.2424	-360.0588	0.16106
	1,000	-3,097.2222	-3,095.2041	0.21222
	3,000	-9,716.3881	-9,714.3821	0.19792
	5,000	-15,524.364	-15,522.361	0.21167
	10,000	-33,554.338	-33,552.336	0.18674
Normal	100	-3.9620	-1.7784	0.96601
	1,000	17.3923	19.4103	1.00722
	3,000	-58.9903	-56.9843	0.98972
	5,000	-73.1305	-71.1269	0.99242
	10,000	20.9714	22.9732	1.00090



each estimation technique, bias and mean square error (MSE) are calculated for the sample size of 10,000 where

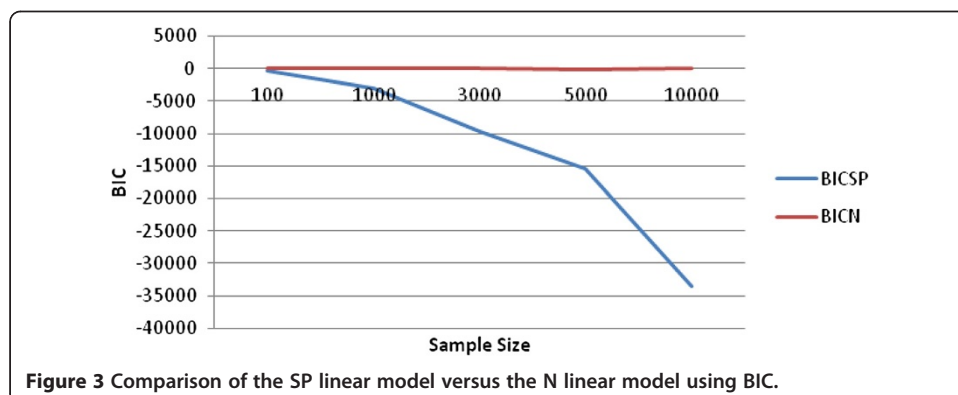
$$\text{MSE}(\hat{\beta}) = \text{var}(\hat{\beta}) + [\text{bias}(\hat{\beta})]^2. \tag{45}$$

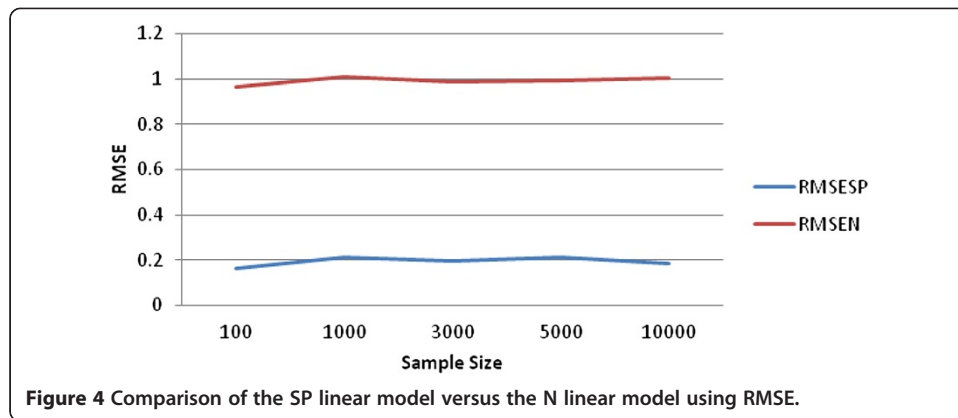
The computational result is presented in Table 4.

As we expect, the results reported in Table 4 show that ML estimators have both smaller one-step-ahead forecast bias and less MSE than OLS estimators. This reveals that ML estimators exhibit superior performance to OLS estimators. This confirms the fact that deviations from normality cause OLS estimators to be poor estimators.

#### Comparison of the SP-LRM with the N-LRM

The linear regression model with symmetric platykurtic errors (SP-LRM) and linear regression model with normal errors (N-LRM) were applied to the simulated data sets. This produced parameter estimates  $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma})$  for the model with SP errors and parameter estimates  $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma})$  for the model with  $N$  errors. The simulated multivariate data are used to compare the performance of the linear regression model with symmetric platykurtic error terms with that of the linear regression model with normal error terms. In order to determine the best linear model among the fitted ones, we computed the Akaike information criteria (AIC) and Bayesian information criteria (BIC) with model diagnostics root-mean-square error (RMSE). Needless to say, the proposed model would be chosen as the best model according to the minimum of AIC, BIC, and RMSE. The output of the simulation study using various sample sizes is presented in Table 5.





Models with small values for the criterion are potential candidate models where it can be seen that the SP distribution provides the best fit to the data. That is, both information criteria and model diagnostics indicate that the linear model with symmetric platykurtic distributed error terms consistently performed best across all the sample sizes of the simulation. The fact that SP-LRM is superior to N-LRM is also consistently noticed from Figures 2, 3, and 4.

#### Application of the model

As an illustration of the proposed methodology, we considered a real data set concerning 202 athletes collected at the Australian Institute of Sport, courtesy of Richard Telford and Ross Cunningham [33]. It is also available within the package *sn* in *R*. The variables examined are body mass index (BMI), red cell count (RCC), white cell count (WCC), and plasma ferritin concentration (PFC). The first is a biometrical variable, while the remaining three concern blood composition. They are summarized in Table 6. We studied the linear dependence of the biometrical variable on the blood composition variables.

We compute the skewness, kurtosis, and Jarque-Bera statistic to test the normality hypothesis for the body mass index. The results are shown in Table 7. Several other statistics could be used to test normality, such as the modified Shapiro-Wilk statistic, Anderson-Darling test, and Kolmogorov-Smirnov test. However, as the Jarque-Bera statistic is one of the most powerful tests of normality, and the results of the other statistics are similar, we report only the results of the Jarque-Bera statistic and its corresponding skewness and kurtosis. The 1% level of significance shown in the table leads to rejection of the null hypothesis of normality for the body mass index.

Table 8 shows the results of fitting multiple regression models with symmetric platykurtic distributed errors and Gaussian errors (SP-LRM and N-LRM) to the Australian Institute of Sport data set using the maximum likelihood method of estimation. The standard errors are the asymptotic standard errors based on the observed information matrix given in (26).

**Table 6 Variables of the Australian Institute of Sport data frame with 202 observations**

	Variables
Response variables	Y = BMI
Independent variables	X1 = RCC
	X2 = WCC
	X3 = PFC



**Table 7 Tests for departure from normality of the response variable (BMI)**

Response variable	Skewness	Kurtosis	Jarque-Bera statistic
BMI	1.174474881	2.65370385	36.85416*

\* $p < 1\%$ .

From Table 8, it is observed that the estimates differ slightly between the two models. Following Lachos et al. [34], we propose selecting the best fit between N-LRM and SP-LRM by inspection of information criteria such as AIC and BIC (the preferred model is the one with the smallest value of the criterion). The AIC and BIC values shown at the bottom of Table 8 indicate that the SP-LRM outperforms the N-LRM. Therefore, the fitted model for BMI is

$$\hat{Y} = 4.7799 + 3.1472X_1 + 0.0899X_2 + 0.1988X_3 \tag{46}$$

with the estimated standard errors

$$\begin{aligned} \text{s.e.}(\hat{\beta}_0) &= 2.3617, \text{s.e.}(\hat{\beta}_1) = 0.4575, \text{s.e.}(\hat{\beta}_2) = 0.1018, \\ \text{s.e.}(\hat{\beta}_3) &= 0.0337, \text{and s.e.}(\hat{\sigma}) = 0.1918. \end{aligned} \tag{47}$$

**Summary and conclusions**

A multiple linear regression model generalizes the simple linear regression model by allowing the response variable to depend on more than one explanatory variable. In this paper, we have explored the idea of using a symmetric platykurtic distribution for analyzing nonnormal errors in the multivariate linear regression model. The symmetric platykurtic distribution serves as an alternative to the normal distribution with platykurtic nature. The maximum likelihood estimators of the model parameters are derived and we found them feasible. Through simulation studies, the properties of these estimators are studied. Traditional OLS estimation is carried out in parallel and the results are compared. The simulated results reveal that the ML estimators are more efficient than the OLS estimators in terms of the relative efficiency of one-step-ahead forecast mean square error. A comparative study of the developed regression model with the Gaussian model revealed that this model gives good fit to some data sets. The asymptotic properties of the maximum likelihood estimators are studied, and the large sample theory with respect to regression coefficients is also presented. The utility of the proposed model is demonstrated with real-time data. This regression model is much more useful for analyzing data sets arising from agricultural experiments, portfolio management, space experiments, and a wide range other practical problems. The calculations

**Table 8 ML estimates of model parameters calculated from the real data set**

Parameter	N-LRM		SP-LRM	
	Estimate	SE	Estimate	SE
$\hat{\beta}_0$	4.7799	2.3617	4.7800	2.3382
$\hat{\beta}_1$	3.1472	0.4575	3.1472	0.4530
$\hat{\beta}_2$	0.08992	0.1018	0.0899	0.1008
$\hat{\beta}_3$	0.1988	0.0337	0.1989	0.0333
$\hat{\sigma}$	2.5133	0.1918	2.4883	0.1238
AIC (smaller is better)	407.0530		376.2918	
BIC (smaller is better)	408.2227		378.4526	

in this paper make considerable use of a combination of three popular statistical packages: Mathematica 9.0, Matlab R2012b, and SAS 9.0.

## Appendix

### Summary of properties of two-parameter symmetric platykurtic distribution

1. Notation	$NS(\mu, \sigma^2)$
2. Parameters	$\mu \in \mathcal{R}$ –Mean (location) $\sigma^2 > 0$ –Variance (squared scale)
3. Support	$y \in \mathcal{R}$
4. PDF	$\frac{\left[2 + \left(\frac{y-\mu}{\sigma}\right)^2\right] e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}}{3\sigma\sqrt{2\pi}}$
5. CDF	$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt - \frac{(y-\mu)e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}}{3\sigma\sqrt{2\pi}}$
6. Mean	$\mu$
7. Median	$\mu$
8. Mode	$\mu$
9. Variance	$\frac{5}{3}\sigma^2$
10. Skewness	0
11. Kurtosis	2.52
12. Entropy	$\frac{-529}{1500} + \ln(3\sigma\sqrt{2\pi})$
13. MGF	$e^{\mu t + \frac{t^2\sigma^2}{2}} \left[1 + \frac{(\sigma t)^2}{3}\right]$
14. CF	$e^{\mu it - \frac{1}{2}\sigma^2 t^2} \left\{1 + \frac{(\sigma it)^2}{3}\right\}$
15. CGF	$\mu t + \frac{1}{2}\sigma^2 t^2 + \ln\left[1 + \frac{(\sigma t)^2}{3}\right]$
16. Central moments	$\mu_{2n} = \left[ \frac{\left(n + \frac{3}{2}\right)\Gamma\left(n + \frac{1}{2}\right)}{3\sqrt{\pi}} \right] 2^{n+1}\sigma^{2n}, \text{ even central moments}$ $\mu_{2n+1} = 0, \text{ odd central moments}$
17. Fisher information	$\begin{pmatrix} I_{\mu} X'X & 0 \\ 0 & nI_{\sigma^2} \end{pmatrix}; I_{\mu} = \frac{4.934}{\sigma^2}, I_{\sigma^2} = \frac{6.68275}{\sigma^4}$

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

The authors are grateful to the Editor of JUAA, two anonymous referees, and SpringerOpen Copyediting Management for the helpful comments and suggestions on the earlier version of this article. The present version of the paper owes much to their precise and kind remarks.

Received: 21 July 2013 Accepted: 6 November 2013

Published: 26 November 2013

#### References

- Seber, GAF: Linear Regression Analysis. Wiley, New York (1977)
- Montgomery, DC, Peck, EA, Vining, GG: Introduction to Linear Regression Analysis, 3rd edn. Wiley, New York (2001)

3. Grob, J: Linear Regression. In: *Lecture Notes in Statistics*, vol. 175. Springer, Berlin (2003)
4. Sengupta, D, Jammalamadaka, SR: Estimation in the linear model. In: *Linear Models: An Integrated Approach*, pp. 93–131. World Scientific, River Edge (2003)
5. Seber, GAF, Lee, AJ: *Linear Regression Analysis*, 2nd edn. Wiley, New York (2003)
6. Weisberg, S: *Applied Linear Regression*, 3rd edn. Wiley, New York (2005)
7. Yan, X, Su, XG: *Linear Regression Analysis: Theory and Computing*. World Scientific, Hackensack (2009)
8. Srivastava, MS: *Methods of Multivariate Statistics*. Wiley, New York (2002)
9. Fama, EF: The behaviour of stock market prices. *J. Bus.* **38**, 34–105 (1965)
10. Sutton, J: Gibrat's legacy. *J. Econ. Lit.* **35**, 40–59 (1997)
11. Zeckhauser, R, Thompson, M: Linear regression with non-normal error terms. *Rev. Econ. Stat.* **52**(3), 280–286 (1970)
12. Zellner, A: Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms. *J. Am. Stat. Assoc.* **71**(354), 400–405 (1976)
13. Sutradhar, BC, Ali, MM: Estimation of the parameters of a regression model with a multivariate t error variable. *Commun. Stat. Theory* **15**, 429–450 (1986)
14. Tiku, ML, Wong, WK, Bian, G: Estimating parameters in autoregressive models in non-normal situations: symmetric innovations. *Commun. Stat. Theory Methods* **28**(2), 315–341
15. Tiku, ML, Wong, WK, Vaughan, DC, Bian, G: Time series models with non-normal situations: symmetric innovations. *J. Time Ser. Anal.* **2**(5), 571–596 (2000)
16. Tiku, ML, Islam, MQ, Selcuk, AS: Non-normal regression, II: symmetric distributions. *Commun. Stat. Theory Methods* **30**(6), 1021–1045 (2001)
17. Sengupta, D, Jammalamadaka, SR: The symmetric non-normal case. In: *Linear Models: An Integrated Approach*, pp. 131–133. World Scientific, River Edge (2003)
18. Liu, M, Bozdogan, H: Power exponential multiple regression model selection with ICOMP and genetic algorithms. Working paper, Springer, Tokyo (2004)
19. Wong, WK, Bian, G: Estimation of parameters in autoregressive models with asymmetric innovations. *Stat. Prob. Lett.* **71**(1), 61–70 (2005)
20. Wong, WK, Bian, G: Robust estimation of multiple regression model with asymmetric innovations and its applicability on asset pricing model. *Euras. Rev. Econ. Financ.* **1**(4), 7 (2005)
21. Liu, M, Bozdogan, H: Multivariate regression models with power exponential random errors and subset selection using genetic algorithms with information complexity. *Eur. J. Pure Appl. Math.* **1**(1), 4–37 (2008)
22. Soffritti, G, Galimberti, G: Multivariate linear regression with non-normal errors: a solution based on mixture models. *Stat. Comput.* **21**(4), 523–536 (2011)
23. Jafari, H, Hashemi, R: Optimal designs in a simple linear regression with skew-normal distribution for error term. *J. Appl. Math.* **1**(2), 65–68 (2011)
24. Jahan, S, Khan, A: Power of t-test for simple linear regression model with non-normal error distribution: a quantile function distribution approach. *J. Sci. Res.* **4**(3), 609–622 (2012)
25. Bian, G, McAleer, M, Wong, WK: Robust estimation and forecasting of the capital asset pricing model. *Ann. Financ. Econ.* (2013). in press
26. Srinivasa Rao, K, Vijay Kumar, CVSR, Lakshmi Narayana, J: On a new symmetrical distribution. *J. Indian Soc. Agric. Stat.* **50**(1), 95–102 (1997)
27. Seshashayee, M, Srinivas Rao, K, Satyanarayana, CH, Srinivasa Rao, P: Image segmentation based on a finite generalized symmetric platykurtic mixture model with K-means. *Int. J. Comput. Sci. Issu.* **8**(3), 2 (2011)
28. Marsaglia, G: Evaluating the normal distribution. *J. Stat. Softw.* **11**(4), 1–7 (2004)
29. Gupta, RC, Gupta, RD: Proportional reversed hazard rate model and its applications. *J. Stat. Plann. Inference* **137**(11), 3525–3536 (2007)
30. Greene, W: *Econometric Analysis*, 5th edn. Prentice-Hall, Upper Saddle River (2003)
31. Clements, MP, Hendry, DF: An empirical study of seasonal unit roots in forecasting. *Int. J. Forecast.* **13**(3), 341–355 (1997)
32. Chiang, TC, Qiao, Z, Wong, WK: New evidence on the relation between return volatility and trading volume. *J. Forecast.* **29**(5), 502–515 (2010)
33. Ferreira, JTAS, Steel, MF: Bayesian multivariate regression analysis with a new class of skewed distributions. Statistics Research Report, 419. Department of Statistics, University of Warwick (2004)
34. Lachos, VH, Bolfarine, H, Arellano-Valle, RB, Montenegro, LC: Likelihood based inference for multivariate skew-normal regression models. *Commun. Stat. Theory Methods* **36**, 1769–1786 (2007)

doi:10.1186/2195-5468-1-13

**Cite this article as:** Andargie and Rao: Estimation of a linear model with two-parameter symmetric platykurtic distributed errors. *Journal of Uncertainty Analysis and Applications* 2013 **1**:13.