INFEREN CE FOR MULTIPLE LINEAR REGRESSION MODEL
WITH EXTENDED SKEW NORMAL ERRORS

A.A. Alhamide1, K. Ibrahim1, and M.T. Alodat2
1 Statistics Program, School of Mathematical Sciences, Faculty of
Science and Technology, Universiti Kebangsaan Malaysia, Malaysia
Email: obideeali@gmail.com, kamarulz@ukm.edu.my
2 Department of Mathematics, Statistics and Physics,
Qatar University, Qatar. Email: alodatmts@qu.edu.qa

ABSTRACT

This paper presents the estimation of the parameters of the multiple linear regression
model when errors are assumed to follow the independent extended skew normal
distribution. The estimators of the regression parameters are determined using the
maximum likelihood and least squares methods. In addition, the asymptotic distributions
of the estimators are studied. The properties of the estimators under both approaches are
compared based on a simulation study and a real data set is applied for illustration.

KEYWORDS

Extended skew normal distribution; Maximum likelihood estimates; Least squares
estimator; Asymptotic distribution; Simulation; Multiple linear regression model; Berry-
Esseen theorem.

1. INTRODUCTION

The multiple linear regression (MLR) model is a common tool which is usually used
for analyzing data, especially in the applied areas such as agriculture, environment,
biometrics and social science, see for example, Arellano-Valle et al. (2005). However,
there are many cases where the assumption of normality is not feasible, may be due to the
presence of outliers or skewness in the data. If the observed data is skewed, it is difficult
to apply the assumption of normality because it may lead to unrealistic results. Many
researchers have used the transformation method to obtain normality by transforming the
data to near normal. Instead of applying the transformation on the data, some symmetric
distributions such as student-t, logistic, power exponential and contaminated normal are
adopted to deal with the lack of fit to the normal distribution, see Cordeiro et al. (2000),
for example. However, in this study, the extended skew normal distribution is assumed
when applying the multiple linear regression model on the data.

Azzalini (1985) has introduced a new class of distribution known as the skew normal
distribution by including the skewness parameter in addition to the scale and location
parameters to model the skewed data. One may refer to Liseo and Loperfido (2003),
Genton et al. (2001) and Capitanio et al. (2003) for more detailed works on skew normal
distribution and its applications. Azzalini and Capitanio (1999) have conducted a study
on the applications of skew normal distributions and some aspects of statistical inference.
Recently, Cancho et al. (2010) have provided the statistical inference for non-linear regression model based on the skew normal error model as suggested by Sahu et al. (2003). In addition to the skew normal distribution, the extended skew normal (ESN) distribution has also been considered in modeling the data with the presence of skewness. This ESN distribution has been introduced by Adcock (2010). As explained in his work, if a random variable $Y$ is said to follow an independent extended skew normal distribution with location parameter $\mu$, scale parameter $\sigma^2$, skewness parameter $\lambda$ and extra parameter $\tau$, then the probability density function of $Y$ is given by

$$f_Y(y; \Psi) = \phi(y; \mu, \sigma^2, \lambda, \tau) \Phi \left( \frac{\tau + \lambda \sigma^{-2}(y - \mu)}{\sqrt{1 + \lambda^2 \sigma^{-2}}} \right) / \Phi(\tau), \ y \in \mathbb{R}^n,$$

where $\Psi = (\mu, \sigma^2, \lambda, \tau)$ represents the parameters, $\phi(y; \mu, \sigma^2)$ and $\Phi(y; \mu, \sigma^2)$ denote the probability density function (pdf) and cumulative distribution function (CDF) of the normal distribution respectively. The notation $y \sim ESN(\mu, \sigma^2, \lambda, \tau)$ can be used to describe the density of $Y$. The main purpose of this paper is to determine the estimators for the parameters of the multiple linear regression model where the errors follow the independent extended skew normal distribution. In addition to determining the estimators, the properties of these estimators are also studied. In Section 2, we present the multiple linear regression model under ESN errors. In Section 3, we show two methods of parameter estimation, i.e., maximum likelihood and least squares methods. In Section 4, we derive the asymptotic distributions for the estimators of the multiple linear regression model denoted as $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$. In Section 5, we conduct a simulation study for computing the maximum likelihood and the least square estimates of the parameters. In Section 6, we apply the findings to the Scottish hills races data. Finally, we conclude the article with a discussion of the results found.

2. MULTIPLE LINEAR REGRESSION MODEL WITH ESN ERRORS

In this section, we consider the multiple linear regression model when the errors are independent and identically distributed following the extended skew normal distribution given by

$$Y_i = \mathbf{x}_i^T \mathbf{\beta} + \varepsilon_i, \ for \ i = 1, 2, ..., n,$$

where

$$\mathbf{x}_i^T = (1, x_{i1}, x_{i2}, x_{i3}, ..., x_{ip}), \mathbf{\beta} = (\beta_0, ..., \beta_p)^T$$

and

$$\varepsilon_1, ..., \varepsilon_n \sim iid ESN(- (\tau + h_1(\tau)) \lambda, \sigma^2, \lambda, \tau),$$

where $h_i(x) = \partial^i \log \Phi(x) / \partial x^i, i = 1, 2$.

Hence, the density function of $Y_i$ is given by

$$Y_i \sim iid ESN(\mathbf{x}_i^T \mathbf{\beta} - (\tau + h_1(\tau)) \lambda, \sigma^2, \lambda, \tau), \ for \ i = 1, 2, ..., n.$$

Let $\mathbf{\theta} = (\mathbf{\beta}, \sigma^2, \lambda, \tau)^T$. Then, the density function of $Y_i$ is given by

$$f_{Y_i}(y_i; \mathbf{\theta}) = \phi(y_i; \mathbf{x}_i^T \mathbf{\beta} - (\tau + h_1(\tau)) \lambda, \sigma^2, \lambda, \tau) \times \Phi \left( \frac{\tau + \lambda \sigma^{-2}(y_i - \mathbf{x}_i^T \mathbf{\beta} + (\tau + h_1(\tau)) \lambda)}{\sqrt{1 + \lambda^2 \sigma^{-2}}} \right) / \Phi(\tau), \ y_i \in \mathbb{R}^n.$$

(2.1)
In order to simplify the mathematical derivation, we write the pdf of \( Y_i \) as follows:

\[
f_{Y_i}(y_i; \theta) = \frac{1}{(2\pi\sigma^2)^{1/2} \Phi(\tau)} \exp \left( -\frac{1}{2} D_i(\theta) \right) \Phi \left( K_i(\theta) \right),
\]

(2.2)

where

\[
D_i(\theta) = \frac{(y_i - x_i^T \beta + (\tau + h_1(\tau)) \lambda)^2}{\sigma^2},
\]

and

\[
K_i(\theta) = \frac{\lambda \sigma^{-2}(y_i - x_i^T \beta + (\tau + h_1(\tau)) \lambda)}{\sqrt{1 + \lambda^2 \sigma^{-2}}}.
\]

Then the joint pdf of \( Y \) is given by

\[
f_Y(y; \theta) = \frac{1}{(2\pi\sigma^2)^{n/2} \Phi(\tau)^n} \exp \left( -\frac{1}{2} \sum_{i=1}^n D_i(\theta) \right) \prod_{i=1}^n \Phi \left( K_i(\theta) \right).
\]

We can easily show that the expectation and variance-covariance of \( Y \) as follows:

\[
E(\epsilon) = 0, E(Y) = X \beta_*, \text{ and Cov}(Y) = \sigma_e^2 I_n, \text{ where } \beta_* = (\beta_0 + \xi, \beta_1 + \xi, \ldots, \beta_p + \xi)^T, \xi = -(1 + h_1(\tau)) \lambda \text{ and } \sigma_e^2 = (\sigma^2 + (1 + h_2(\tau)) \lambda^2).
\]

So, the log-likelihood function of \( Y \) is

\[
\log f_Y(y; \theta) = -\frac{n \log 2\pi}{2} - \frac{n \log \sigma^2}{2} - n \log \Phi(\tau) - \frac{1}{2} \sum_{i=1}^n D_i(\theta) + \sum_{i=1}^n \log \Phi \left( K_i(\theta) \right).
\]

### 3. METHODS OF ESTIMATION

#### Maximum Likelihood Estimation

We can find the maximum likelihood estimation (MLE) for each parameter in vector \( \theta \) by taking the derivative of \( \log f_Y(y; \theta) \) with respect to each parameter and setting the derivatives to zero as follows:

1. Derivative with respect to \( \beta \):

\[
\frac{1}{\sigma^2} \sum_{i=1}^n x_i^T (y_i - x_i^T \beta + (\tau + h_1(\tau)) \lambda) - \frac{\lambda \sigma^{-2}}{\sqrt{1 + \lambda^2 \sigma^{-2}}} \sum_{i=1}^n x_i^T \frac{\Phi(K_i(\theta))}{\Phi(K_i(\theta))} = 0.
\]

(3.1)

2. Derivative with respect to \( \lambda \):

\[
\frac{(\tau + h_1(\tau))}{\sigma^2} \sum_{i=1}^n (y_i - x_i^T \beta + (\tau + h_1(\tau)) \lambda) - \sum_{i=1}^n \left( \frac{\lambda \sigma^{-2}}{\sqrt{1 + \lambda^2 \sigma^{-2}}} \right) \frac{\Phi(K_i(\theta))}{\Phi(K_i(\theta))} = 0.
\]

(3.2)

3. Derivative with respect to \( \sigma^2 \):

\[
\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - x_i^T \beta + (\tau + h_1(\tau)) \lambda)^2 - \frac{n}{2\sigma^2} - \sum_{i=1}^n \left( \frac{\lambda \sigma^{-2}}{\sqrt{1 + \lambda^2 \sigma^{-2}}} \right) \frac{\Phi(K_i(\theta))}{\Phi(K_i(\theta))} = 0.
\]

(3.3)
4. Derivative with respect to $\tau$:
\[
\frac{n\phi(\tau)}{\Phi(\tau)} + \frac{1}{\sigma^2} \sum_{i=1}^{n} \lambda \left( 1 + h_1(\tau) \right) (y_i - x_i^T \beta + (\tau + h_1(\tau)) \lambda) \\
- \frac{1}{\sqrt{1 + \lambda^2 \sigma^{-2}}} \sum_{i=1}^{n} \frac{\phi(k_i(\theta))}{\Phi(k_i(\theta))} \left( 1 + \lambda^2 \sigma^{-2} (1 + h_1'(\tau)) \right) = 0. \tag{3.4}
\]

We used the numerical technique to solve the equations above.

**Least Square Estimation of the Parameter vector $\beta$**

We seek estimators that minimize the sum of squares of the deviation of the $n$ observed $y_i$ from their predicted values $\hat{y}_i$. It can be easily shown that $\hat{\beta} = (X^T X)^{-1} X^T Y$ and the variance $s^2 = \frac{1}{n-p} (Y^T Y - \hat{\beta}^T X^T Y)$. Then, the properties of the least squares estimator can be described by the following theorem:

**Theorem 1:**

i) $E (\hat{\beta}) = E ((X^T X)^{-1} X^T Y) = \beta$, i.e., $\hat{\beta}$ is an unbiased estimator for $\beta$.

ii) $\text{Cov}(\hat{\beta}) = \text{Cov}((X^T X)^{-1} X^T Y) = (\sigma^2 + (1 + h_2(\tau)) \lambda^2) (X^T X)^{-1}$.

iii) $E (s^2) = \sigma^2 + (1 + h_2(\tau)) \lambda^2$.

This theorem can easily be proven by referring to Rencher and Schaalje (2008)

### 4. Asymptotic Distributions for the Estimators of the Multiple Linear Regression Model

Consider the following multiple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i, \quad i = 1, 2, ..., n,$$

where $\beta_0, \beta_1$, and $\beta_2$ are unknown parameters and $\epsilon_i$ is the $i$th error term. The least squares estimator for the parameters are:

$$
\begin{align*}
\bar{\beta}_0 &= \bar{Y} - \bar{\beta}_1 \bar{x}_1 - \bar{\beta}_2 \bar{x}_2, \\
\bar{\beta}_1 &= \frac{S_{x_1y} - \bar{\beta}_2 S_{x_1x_2}}{S_{x_1x_1}}, \\
\bar{\beta}_2 &= \frac{S_{x_1x_1} S_{x_2y} - S_{x_1x_2} S_{x_1y}}{S_{x_1x_1} S_{x_2x_2} - (S_{x_1x_2})^2}, \\
\end{align*}
\tag{4.1}
$$

where

$$S_{x_1y} = \sum_{i=1}^{n} x_{i1} y_i - \left( \sum_{i=1}^{n} x_{i1} \right) \left( \sum_{i=1}^{n} y_i \right),$$

$$S_{x_2y} = \sum_{i=1}^{n} x_{i2} y_i - \left( \sum_{i=1}^{n} x_{i2} \right) \left( \sum_{i=1}^{n} y_i \right),$$

$$S_{x_1x_2} = \sum_{i=1}^{n} x_{i1} x_{i2} - \left( \sum_{i=1}^{n} x_{i1} \right) \left( \sum_{i=1}^{n} x_{i2} \right),$$

$$S_{x_1x_1} = \sum_{i=1}^{n} x_{i1}^2 - \frac{(\sum_{i=1}^{n} x_{i1})^2}{n},$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} y_i.$$
Using equation (4.1) and with some straightforward simplification, we have
\[
\bar{\beta}_0 - \beta_0 = \sum_{i=1}^{n} a_i e_i,
\]
where
\[
a_i = \frac{1}{n} + \frac{U(x_{i1} - \bar{x}_1) + V(x_{i2} - \bar{x}_2)}{RZ},
\]
and
\[
Z = S_{x_1 x_2} (s_{x_1 x_1})^2 - S_{x_1 x_2} (s_{x_1 x_2})^2, U = \bar{x}_2 S_{x_1 x_2} Z - \bar{x}_1 S_{x_1 x_1} S_{x_2 x_2} R,
\]
and
\[
V = \bar{x}_1 S_{x_1 x_2} S_{x_1 x_1} R - \bar{x}_2 S_{x_1 x_1} Z.
\]

Similarly, we can rewrite \(\hat{\beta}_1\) and \(\hat{\beta}_2\) given in (4.1) to obtain
\[
\bar{\beta}_1 - \beta_1 = \sum_{i=1}^{n} b_i e_i,
\]
where
\[
b_i = \frac{S_{x_1 x_1} S_{x_2 x_2} (x_{i1} - \bar{x}_1) - S_{x_1 x_2} S_{x_1 x_1} (x_{i2} - \bar{x}_2)}{Z},
\]
and
\[
\bar{\beta}_2 - \beta_2 = \sum_{i=1}^{n} c_i e_i,
\]
where
\[
c_i = \frac{S_{x_1 x_1} (x_{i2} - \bar{x}_2) - S_{x_1 x_2} (x_{i1} - \bar{x}_1)}{R}.
\]

It is easy to show that the estimators \(\hat{\beta}_0, \hat{\beta}_1\) and \(\hat{\beta}_2\) are unbiased for \(\beta_0, \beta_1\) and \(\beta_2\) respectively and the variances of these estimators as the following:
\[
\text{Var}(\bar{\beta}_0) = \left(\sigma^2 + (1 + h_2(\tau))\lambda^2\right) \left\{ \frac{1}{n} + \frac{\bar{x}_1^2 S_{x_2 x_2} + \bar{x}_2^2 S_{x_1 x_1}}{S_{x_1 x_1} S_{x_2 x_2} - (S_{x_1 x_2})^2} \right\},
\]
\[
\text{Var}(\bar{\beta}_1) = \left(\sigma^2 + (1 + h_2(\tau))\lambda^2\right) \frac{S_{x_1 x_2}}{S_{x_1 x_1} S_{x_2 x_2} - (S_{x_1 x_2})^2},
\]
and
\[
\text{Var}(\bar{\beta}_2) = \left(\sigma^2 + (1 + h_2(\tau))\lambda^2\right) \frac{S_{x_1 x_1}}{S_{x_1 x_1} S_{x_2 x_2} - (S_{x_1 x_2})^2}.
\]

In order to derive the asymptotic distributions of \(\hat{\beta}_0, \hat{\beta}_1\) and \(\hat{\beta}_2\), we need to use Berry-Esseen theorem and Slutsky’s theorem. For more details (see Chow and Teicher, 1978; Alodat et al., 2010).

**Theorem 2. (Berry-Esseen)**

If \(\{X_n, n \geq 1\}\) are independent random variables such that, \(E(X_n) = 0, E(X_n^2) = \sigma_n^2 > 0, S_n^2 = \sum_{i=1}^{n} \sigma_i^2 > 0, \Gamma_n^{2+\delta} = \sum_{i=1}^{n} E|X_i|^{2+\delta} < \infty, n \geq 1, \) for some
\( \delta \in (0,1) \) and \( W_n = \sum_{i=1}^{n} X_i \), then there is exist a universal constant \( C_\delta \) such that

\[
\sup_{-\infty < x < \infty} \left| P(W_n \leq xS_n) - \Phi(x) \right| \leq C_\delta \left( \frac{T_n}{S_n} \right)^{2+\delta}.
\]

**Theorem 3.**

By using Slutsky theorem and Berry-Esseen theorem, we have

\[
\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} \xrightarrow{d} N(0,1),
\]

\[
\frac{\hat{\beta}_2 - \beta_2}{\sqrt{\text{Var}(\hat{\beta}_2)}} \xrightarrow{d} N(0,1),
\]

and

\[
\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\text{Var}(\hat{\beta}_0)}} \xrightarrow{d} N(0,1),
\]

where

\[
\text{Var}(\hat{\beta}_0) = \frac{SSE}{(n-3)} \left\{ \frac{1}{n} + \frac{\bar{x}_1^2 S_{x_2 x_2} + \bar{x}_2^2 S_{x_1 x_1}}{S_{x_1 x_1} S_{x_2 x_2} - (S_{x_1 x_2})^2} \right\},
\]

\[
\text{Var}(\hat{\beta}_1) = \frac{SSE}{(n-3)} \left( S_{x_1 x_2} S_{x_2 x_2} - (S_{x_1 x_2})^2 \right),
\]

and

\[
\text{Var}(\hat{\beta}_2) = \frac{SSE S_{x_1 x_1}}{(n-3) \left( S_{x_1 x_1} S_{x_2 x_2} - (S_{x_1 x_2})^2 \right)}.\]

For the proofs of Theorem 3, see the Appendix.

**5. SIMULATION**

**The Maximum Likelihood Estimation (MLE) Method**

In order to estimate the parameters of the multiple linear regression model under extended skew normal errors (ESN-MLR) and normally distributed errors (N-MLR) using the maximum likelihood method, we conduct a simulation study for sample sizes \( n = 35 \) and 10000 iterations. Then, we compare the bias and mean square errors (MSE) for the estimators in both cases. The simulation results are shown in Table (1).
Table 1
Maximum Likelihood Estimates, Bias and Mean Square Errors
Assuming Certain Values for Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>N-MLR</th>
<th>ESN-MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Bias</td>
</tr>
<tr>
<td>$\beta_0(2)$</td>
<td>2.1268</td>
<td>0.0282</td>
</tr>
<tr>
<td>$\beta_1(3)$</td>
<td>2.9902</td>
<td>-0.0042</td>
</tr>
<tr>
<td>$\beta_2(5)$</td>
<td>4.9850</td>
<td>-0.0522</td>
</tr>
<tr>
<td>$\lambda(0.3)$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma^2(1)$</td>
<td>0.8783</td>
<td>-0.0457</td>
</tr>
<tr>
<td>$\tau(0.2)$</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Based on the Table (1), we have parameter estimates for N-MLR and ESN-MLR, in addition to the estimates of bias and MSE for the respective estimators which have been found based on simulation. In general, we note from Table (1) that the estimates of bias and MSE for the ESN-MLR are smaller than those in the N-MLR.

The Least Squares Estimation (LSE) Method
The LSE is a common method which can also be used for estimating the parameters $\beta_0, \beta_1, \beta_2$ and $\sigma^2$ for both MLR-N and MLR-ESN. So, we conduct a simulation study to compare the standard errors for the estimators obtained. The results are shown in Table (2).

Table 2
Least Square Estimates and Standard Errors
Assuming Certain Values of the Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>N-MLR</th>
<th>ESN-MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>S.E</td>
</tr>
<tr>
<td>$\beta_0(2)$</td>
<td>2.1979</td>
<td>0.1733</td>
</tr>
<tr>
<td>$\beta_1(3)$</td>
<td>3.0797</td>
<td>0.0381</td>
</tr>
<tr>
<td>$\beta_2(5)$</td>
<td>4.9973</td>
<td>0.0439</td>
</tr>
<tr>
<td>$\lambda(0.3)$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma^2(1)$</td>
<td>1.1012</td>
<td>1.0494</td>
</tr>
<tr>
<td>$\tau(0.2)$</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

From Table (2), we can notice that the standard errors of the parameter estimates are found to be smaller for ESN-MLR as compared to N-MLR, indicating that the estimation is more precise under ESN-MLR errors.
6. AN APPLICATION: THE SCOTTISH HILLS RACES DATA

The Scottish hills races data, which has also been applied by Chatterjee and Hadi (2012), consist of $n = 33$ observations where the response variable $y = \text{time (in seconds)}$ is related to two other predictor variables, namely, $x_1 = \text{distance (in mile)}$ and $x_2 = \text{climb (in feet)}$. This set of data is also available in the R package call (MASS). To investigate the presence of skewness in the data, as shown in Figure 1, normal and skew normal probability density functions are fitted to the histogram of the residual found after fitting the skew normal error model. As given in Figure 2, the data are further plotted on the normal Q-Q plot. Several points fall away from the straight line on the normal Q-Q plot, indicating the presence of outliers. It will be further shown in the analysis that the extended skew normal model can nicely account for the presence of outliers in the data.

![The Scottish Hills Races Data](image)

Figure 1: The histogram of the residuals of the Scottish hills races data and the fitted normal and skew normal model
Figure 2: The normal Q-Q plots of the Scottish hills races data

Both N-MLR and ESN-MLR are fitted by using the maximum likelihood and least square methods to the Scottish hills races data and the results found are given in Table (3) and (4) respectively. Also, note that the Akaike Information Criterion (AIC) values shown in Table (3) indicate that ESN-MLR outperforms N-MLR since the smaller value is obtained when extended skew normal errors are assumed.

Table 3
Results of Fitting N-MLR and ESN-MLR to the Scottish Hills Races Data Involving Maximum Likelihood Estimates and Standard Errors

<table>
<thead>
<tr>
<th>Parameters</th>
<th>N-MLR</th>
<th>ESN-MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>S.E</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-10.2728</td>
<td>1.8421</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>6.72021</td>
<td>0.2469</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.00787</td>
<td>0.0010</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>34.5378</td>
<td>5.8769</td>
</tr>
<tr>
<td>$\tau$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-104.6752</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>217.3504</td>
<td></td>
</tr>
</tbody>
</table>
Table 4
Results of Fitting N-MLR and ESN-MLR to the Scottish Hills Races Data Involving Least Squares Estimates and Standard Errors

<table>
<thead>
<tr>
<th>Parameters</th>
<th>N-MLR</th>
<th></th>
<th>ESN-MLR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>S.E</td>
<td>Estimate</td>
<td>S.E</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-10.3616</td>
<td>1.8976</td>
<td>-10.3616</td>
<td>1.7031</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>6.6921</td>
<td>0.2543</td>
<td>6.6921</td>
<td>0.2531</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.0080</td>
<td>0.0011</td>
<td>0.0080</td>
<td>0.0011</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>36.6489</td>
<td>6.054</td>
<td>36.6489</td>
<td>6.051</td>
</tr>
</tbody>
</table>

CONCLUSION

In this paper, we study the statistical inference and estimation for the parameters using the maximum likelihood and least squares methods for the multiple linear regression model under normal and extended skew normal errors. Also, we have derived the asymptotic distributions of the estimators for the parameters of the multiple linear regression model under extended skew normal errors. From the comparison of the parameter estimates found based on the simulation study for the regression model under normal and extended skew normal errors using the maximum likelihood method, the fitted model is better for the latter case. The results are further supported by the model fitting of real data since the response variable in the data exhibits some skewness properties due to the presence of outlying observation. There are more potential applications which can be investigated for this proposed regression model in addition to the given example.

ACKNOWLEDGMENT

The authors are thankful to the editor and referees for their helpful comments and suggestions on the earlier version of this paper. Also, we thank Faculty of Science and Technology of Universiti Kebangsaan Malaysia for providing the partial support for this research under the grant DPP-2015-FST.

REFERENCES


APPENDIX

Proofs for the asymptotic distribution of the estimators $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_0$

**Proof 1.**

Using the equation (4.3), we notice that $E (b_t \epsilon_t) = 0$ and

$$E (b_t^2 \epsilon_t^2) = \left( S_{x_1 \times x_1} S_{x_2 \times x_2}(x_i - \bar{x}_1) - S_{x_1 \times x_2} S_{x_1 \times x_2}(x_i - \bar{x}_2) \right)^2 \left( \sigma^2 + (1 + h_2(\tau))\lambda^2 \right),$$

then

$$\Gamma_n^{2+\delta} = A_\delta \sum_{i=1}^n \left| S_{x_2 \times x_2} (x_i - \bar{x}_1) - S_{x_1 \times x_2} (x_i - \bar{x}_2) \right|^{2+\delta},$$

where $A_\delta = E | \epsilon_i |^{2+\delta} < \infty$ is independent of $i$. Also,

$$S_n^2 = \frac{(\sigma^2 + (1 + h_2(\tau))\lambda^2)Q}{Z^*},$$

where

$$Q = S_{x_1 \times x_1} (S_{x_2 \times x_2})^2 - 2 S_{x_2 \times x_2} (S_{x_1 \times x_2})^2 + S_{x_1 \times x_2} (S_{x_1 \times x_2})^2,$$

and

$$Z^* = S_{x_1 \times x_1} S_{x_2 \times x_2} - (S_{x_1 \times x_2})^2.$$

Hence, we have

$$\left( \frac{\Gamma_n}{S_n} \right)^{2+\delta} = A_\delta \sum_{i=1}^n \left| S_{x_2 \times x_2} (x_i - \bar{x}_1) - S_{x_1 \times x_2} (x_i - \bar{x}_2) \right|^{2+\delta} \frac{(\sigma^2 + (1 + h_2(\tau))\lambda^2)Q}{Z^*}^{2+\delta} \frac{Q^{-2}}{2}. $$

By applying Berry-Esseen theorem, we obtain

$$\sup_{-\infty < x < \infty} \left| P \left( \sum_{i=1}^n b_t \epsilon_i \leq y \sqrt{\frac{(\sigma^2 + (1 + h_2(\tau))\lambda^2)Q}{Z^*}} \right) - \Phi(y) \right| \leq \psi_{\delta, \lambda} \sum_{i=1}^n \left| S_{x_2 \times x_2} (x_i - \bar{x}_1) - S_{x_1 \times x_2} (x_i - \bar{x}_2) \right|^{2+\delta} \frac{Q^{-2}}{2},$$

where

$$\psi_{\delta, \lambda} = \frac{C_\delta A_\delta}{(\sigma^2 + (1 + h_2(\tau))\lambda^2)^{2+\delta}}.$$

The following inequality, as given in Bhattacharya and Rao (1976), is used for proving the theorem:

$$\sum_{i=1}^n \frac{a_i}{n} \leq \left( \sum_{i=1}^n a_i^p \right)^{1/p} \left( \sum_{i=1}^n \frac{a_i}{n} \right)^{1-p}.$$

By using this inequality and assume that $a_i = \sum_{i=1}^n \left| S_{x_2 \times x_2} (x_i - \bar{x}_1) - S_{x_1 \times x_2} (x_i - \bar{x}_2) \right|^2$, $p = 2/(2 + \delta)$ and for $\nu > 2$, we have
\[
\left( \sum_{i=1}^{n} \left| S_{x_2 x_2} (x_{t_1} - \bar{x}_1) - S_{x_1 x_2} (x_{t_2} - \bar{x}_2) \right|^2 \right)^{1+\delta} \\
\geq n^{\delta} \sum_{i=1}^{n} \left| S_{x_2 x_2} (x_{t_1} - \bar{x}_1) - S_{x_1 x_2} (x_{t_2} - \bar{x}_2) \right|^{2+\delta}.
\]

Then, based on a straightforward simplification and setting \( p = 2/(2 + \delta) \), we get
\[
\frac{n^{\delta}}{n^v} \sum_{i=1}^{n} \left| S_{x_2 x_2} (x_{t_1} - \bar{x}_1) - S_{x_1 x_2} (x_{t_2} - \bar{x}_2) \right|^{2+\delta}
\]
\[
= \frac{n^{\delta}}{n^v + 1 - \frac{1}{p}} \left( \sum_{i=1}^{n} \left| S_{x_2 x_2} (x_{t_1} - \bar{x}_1) - S_{x_1 x_2} (x_{t_2} - \bar{x}_2) \right|^{(2+\delta)p} \right)^{1/p}
\]
\[
\leq n^{\delta} \left( \sum_{i=1}^{n} \left| S_{x_2 x_2} (x_{t_1} - \bar{x}_1) - S_{x_1 x_2} (x_{t_2} - \bar{x}_2) \right|^2 \right)^{1+\frac{\delta}{2}}.
\]

This concludes that \( \frac{\delta}{2v} (2 - v) \leq 1 \), for all \( n \geq 1 \) and \( v \geq 2 \). Consequently, we may conclude that
\[
\sup_{-\infty < \chi < \infty} \left\{ P \left( \sum_{i=1}^{n} b_i \epsilon_i \leq y \sqrt{\frac{(\sigma^2 + (1 + h_2(\tau)) \lambda^2) Q}{Z^2}} - \Phi(y) \right) \right\} \leq \psi_{\delta, \lambda} \frac{1}{n^v}
\]

Eventually, this implies that
\[
\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} = \frac{\sum_{i=1}^{n} b_i \epsilon_i}{\sqrt{\left( \frac{(\sigma^2 + (1 + h_2(\tau)) \lambda^2) Q}{Z^2} \right)}} \xrightarrow{d} N(0,1),
\]

where
\[
\text{Var}(\hat{\beta}_1) = \frac{\text{SSE} (S_{x_2 x_2})}{(n-3) \left( S_{x_1 x_1} S_{x_2 x_2} - (S_{x_1 x_2})^2 \right)} \xrightarrow{p} \text{Var}(\hat{\beta}_1).
\]

Then, by Slutsky’s theorem, we conclude that
\[
\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} \xrightarrow{d} N(0,1).
\]

**Proof 2.**

Consider equation (4.4) and we note that \( E(\epsilon_i \epsilon_i) = 0 \) for every \( i = 1, \ldots, n \) and \( \sigma_i^2 = E(\epsilon_i^2 \epsilon_i^2) = \left( \frac{S_{x_1 x_1} (x_{t_2} - \bar{x}_2) - S_{x_1 x_2} (x_{t_1} - \bar{x}_1)}{R} \right)^2 \left( \sigma^2 + (1 + h_2(\tau)) \lambda^2 \right) \). On the other hand
\[
\Gamma_n^{2+\delta} = A_\delta \sum_{i=1}^{n} \left| \frac{S_{x_1x_1}(x_{i2} - \bar{x}_2) - S_{x_1x_2}(x_{i1} - \bar{x}_1)}{R} \right|^{2+\delta},
\]
where \( A_\delta = E|\epsilon_i|^{2+\delta} < \infty \) is independent of \( i \). Also,
\[
S_n^2 = \frac{(\sigma^2 + (1 + h_2(\tau))\lambda^2)W}{R^2},
\]
where
\[
W = S_{x_2x_2}(S_{x_1x_1})^2 - 2S_{x_1x_1}(S_{x_1x_2})^2 + S_{x_1x_1}(S_{x_1x_2})^2.
\]
Hence, we have
\[
\left( \frac{\Gamma_n}{S_n} \right)^{2+\delta} = \frac{A_\delta \sum_{i=1}^{n} \left| S_{x_1x_1}(x_{i2} - \bar{x}_2) - S_{x_1x_2}(x_{i1} - \bar{x}_1) \right|^{2+\delta}}{(\sigma^2 + (1 + h_2(\tau))\lambda^2)^{2+\delta}W^{2+\delta}}.
\]
After applying Berry-Esseen theorem, we get
\[
\sup_{-\infty < x < \infty} \left| P \left( \sum_{i=1}^{n} c_i \epsilon_i \leq y \sqrt{\frac{\left( \sigma^2 + (1 + h_2(\tau))\lambda^2 \right)W}{R^2}} - \Phi(y) \right) \right|
\leq \psi_{\delta,\lambda} \frac{\sum_{i=1}^{n} \left| S_{x_1x_1}(x_{i2} - \bar{x}_2) - S_{x_1x_2}(x_{i1} - \bar{x}_1) \right|^{2+\delta}}{W^{2+\delta}},
\]
where
\[
\psi_{\delta,\lambda} = \frac{C_\delta A_\delta}{(\sigma^2 + (1 + h_2(\tau))\lambda^2)^{2+\delta}}.
\]
Similarly, based on the inequality by Bhattacharya and Rao (1976) as given earlier and assuming \( a_i = \sum_{i=1}^{n} \left| S_{x_1x_1}(x_{i2} - \bar{x}_2) - S_{x_1x_2}(x_{i1} - \bar{x}_1) \right|^2 \) and \( p = 2/(2 + \delta) \) for \( v > 2 \), we have
\[
\left( \sum_{i=1}^{n} \left| S_{x_1x_1}(x_{i2} - \bar{x}_2) - S_{x_1x_2}(x_{i1} - \bar{x}_1) \right|^2 \right)^{1+\frac{\delta}{2}}
\geq n^{\frac{\delta}{2}} \sum_{i=1}^{n} \left| S_{x_1x_1}(x_{i2} - \bar{x}_2) - S_{x_1x_2}(x_{i1} - \bar{x}_1) \right|^{2+\delta},
\]
and with more simplification, we obtain
\[
\frac{n^\delta}{\sqrt{\text{Var}(\hat{\beta}_2)}} \sum_{i=1}^{n} \left| S_{x_1 x_1}(x_{i2} - \bar{x}_2) - S_{x_1 x_2}(x_{i1} - \bar{x}_1) \right|^{2+\delta} \\
= n^{\delta+1-rac{1}{p}} \left( \sum_{i=1}^{n} \left| S_{x_1 x_1}(x_{i2} - \bar{x}_2) - S_{x_1 x_2}(x_{i1} - \bar{x}_1) \right|^{(2+\delta)p} \right)^{1/p} \\
\leq n^{\frac{\delta}{2p}(2-v)} \left( \sum_{i=1}^{n} \left| S_{x_1 x_1}(x_{i2} - \bar{x}_2) - S_{x_1 x_2}(x_{i1} - \bar{x}_1) \right|^2 \right)^{1+\frac{\delta}{2}}.
\]

This concludes that \( \frac{\delta}{2v}(2-v) \leq 1 \), for all \( n \geq 1 \) and \( v \geq 2 \). Consequently, we may conclude that

\[
\sup_{-\infty < x < \infty} \left\{ P \left( \frac{\sum_{i=1}^{n} c_i \epsilon_i}{\sqrt{\text{Var}(\hat{\beta}_2)}} \leq y \sqrt{\frac{\sigma^2 + (1 + h_2(\tau)) \lambda^2}{R^2}} - \Phi(y) \right) \right\} \leq \psi_{\delta, \lambda} \frac{1}{n^{\delta/2}}.
\]

Thus, Berry-Esseen theorem is satisfied. Then, we have

\[
\frac{\hat{\beta}_2 - \beta_2}{\sqrt{\text{Var}(\hat{\beta}_2)}} = \frac{\sum_{i=1}^{n} c_i \epsilon_i}{\sqrt{\left( \frac{\sigma^2 + (1 + h_2(\tau)) \lambda^2}{R^2} \right)}} \xrightarrow{d} N(0,1),
\]

where

\[
\text{Var}(\hat{\beta}_2) = \frac{\text{SSE}(S_{x_1 x_1})}{(n-3) \left( S_{x_1 x_1} S_{x_2 x_2} - (S_{x_1 x_2})^2 \right)} \xrightarrow{p} \text{Var}(\hat{\beta}_2).
\]

Then, by Slutsky’s theorem, we conclude that

\[
\frac{(\hat{\beta}_2 - \beta_2)}{\sqrt{\text{Var}(\hat{\beta}_2)}} \xrightarrow{d} N(0,1).
\]

**Proof 3.**

Using equation (4.2) and applying Berry-Esseen theorem, we can prove the following:

\[ E(a_i \epsilon_i) = 0 \text{ for every } i = 1, 2, \ldots, n. \]

Also, from Berry-Esseen theorem, we have

\[
\sigma_i^2 = E(a_i^2 \epsilon_i^2) = \left( \frac{RZ + Un(x_{i1} - \bar{x}_1) + Vn(x_{i2} - \bar{x}_2)}{nRZ} \right)^2 \left( \sigma^2 + (1 + h_2(\tau)) \lambda^2 \right),
\]

\[
S_n^2 = \frac{\left( \sigma^2 + (1 + h_2(\tau)) \lambda^2 \right) G}{(nRZ)^2},
\]

where
\( G = R^2 Z^2 + n^2 U^2 S_{x_1 x_1} + 2n^2 UV S_{x_1 x_2} + n^2 V^2 S_{x_2 x_2} + 2nR U(x_{i_1} - \bar{x}_1)Z \\
+ 2nR V(x_{i_2} - \bar{x}_2)Z. \)

Now, we compute the term \( \Gamma_n^{2+\delta} \) as follows:

\[
\Gamma_n^{2+\delta} = A_\delta \sum_{i=1}^n \left| \frac{RZ + nU(x_{i_1} - \bar{x}_1) + nV(x_{i_2} - \bar{x}_2)}{nRZ} \right|^{2+\delta},
\]

where \( A_\delta = E|\epsilon_i|^{2+\delta}, i = 1, 2, ..., n. \)

By combining the two expressions of \( \Gamma_n \) and \( S_n \), we get

\[
\left( \frac{\Gamma_n}{S_n} \right)^{2+\delta} = A_\delta \sum_{i=1}^n |RZ + nU(x_{i_1} - \bar{x}_1) + nV(x_{i_2} - \bar{x}_2)|^{2+\delta} \cdot \frac{(\sigma^2 + (1 + h_2(\tau))\lambda^2)^{\frac{2+\delta}{2}}}{G^{\frac{2+\delta}{2}}}
\]

Thus, we get

\[
\sup_{-\infty < x < \infty} \left| P\left( \sum_{i=1}^n a_i \epsilon_i \leq y \sqrt{\frac{(\sigma^2 + (1 + h_2(\tau))\lambda^2)G}{(nRZ)^2}} - \Phi(y) \right) \leq \frac{C_\delta A_\delta \sum_{i=1}^n |RZ + nU(x_{i_1} - \bar{x}_1) + nV(x_{i_2} - \bar{x}_2)|^{2+\delta}}{(\sigma^2 + (1 + h_2(\tau))\lambda^2)^{\frac{2+\delta}{2}} G^{\frac{2+\delta}{2}}}
\]

which leads to the following conclusion:

\[
\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\text{Var}(\hat{\beta}_0)}} = \frac{\sum_{i=1}^n a_i \epsilon_i}{\sqrt{\frac{(\sigma^2 + (1 + h_2(\tau))\lambda^2)G}{(nRZ)^2}}} \xrightarrow{d} N(0,1),
\]

where

\[
\text{Var}(\hat{\beta}_0) = \frac{\text{SSE}}{(n-3)} \left\{ \frac{1}{n} + \frac{\bar{x}_{i_1}^2 S_{x_2 x_2} + \bar{x}_{i_2}^2 S_{x_1 x_1}}{S_{x_1 x_1} S_{x_2 x_2} - (S_{x_1 x_2})^2} \right\} \xrightarrow{p} \text{Var}(\hat{\beta}_0).
\]

Then, by Slutsky’s theorem, we conclude that

\[
\frac{(\hat{\beta}_0 - \beta_0)}{\sqrt{\text{Var}(\hat{\beta}_0)}} \xrightarrow{d} N(0,1).
\]