Pak. J. Statist.
2016 Vol. 32(2), 81-96

INFERENCE FOR MULTIPLE LINEAR REGRESSION MODEL WITH EXTENDED SKEW NORMAL ERRORS

A.A. Alhamide[1], **K. Ibrahim**[1§] and **M.T. Alodat**[2]

[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.ed.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 μ, scale parameter σ^2, skewness parameter λ and extra parameter τ, then the probability density function of Y is given by

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

where $\boldsymbol{\Psi} = (\mu,\sigma^2,\lambda,\tau)$ represents the parameters, $\varphi(y;\mu,\sigma^2)$ 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

$$Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_i, for\ i = 1,2,\dots,n,$$

where

$$\boldsymbol{x}_i^T = \left(1, x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip}\right), \boldsymbol{\beta} = \left(\beta_0,\dots,\beta_p\right)^T$$

and

$$\epsilon_1,\dots,\epsilon_n \sim^{\boldsymbol{iid}} ESN\left(-\left(\tau + h_1(\tau)\right)\lambda, \sigma^2, \lambda, \tau\right),$$

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

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

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

$$f_{Y_i}(y_i;\boldsymbol{\theta}) = \varphi\left(y_i; \boldsymbol{x}_i^T\boldsymbol{\beta} - \left(\tau + h_1(\tau)\right)\lambda, \sigma^2, \lambda, \tau\right) \times \Phi\left(\frac{\tau+\lambda\sigma^{-2}\left(y_i - x_i^T\boldsymbol{\beta} + \left(\tau + h_1(\tau)\right)\lambda\right)}{\sqrt{1+\lambda^2\sigma^{-2}}}\right)/\Phi(\tau), y_i \in \mathbb{R}^n. \quad (2.1)$$

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

$$f_{Y_i}(y_i;\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{1/2}\Phi(\tau)} \exp\left(-\frac{1}{2}D_i(\boldsymbol{\theta})\right)\Phi\big(K_i(\boldsymbol{\theta})\big), \tag{2.2}$$

where

$$D_i(\boldsymbol{\theta}) = \frac{\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} + (\tau + h_1(\tau))\lambda\right)^2}{\sigma^2},$$

and

$$K_i(\boldsymbol{\theta}) = \frac{\tau + \lambda\sigma^{-2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} + (\tau + h_1(\tau))\lambda\right)}{\sqrt{1+\lambda^2\sigma^{-2}}}.$$

Then the joint pdf of $\boldsymbol{Y}$ is given by

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{n/2}\big(\Phi(\tau)\big)^n} \exp\left(-\frac{1}{2}\sum_{i=1}^n D_i(\boldsymbol{\theta})\right)\prod_{i=1}^n \Phi\big(K_i(\boldsymbol{\theta})\big).$$

We can easily show that the expectation and variance-covariance of $\boldsymbol{Y}$ as follows:

$E(\epsilon) = 0, E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}_*$ and $\mathrm{Cov}(\boldsymbol{Y}) = \sigma_\epsilon^2\boldsymbol{I}_n$, where $\boldsymbol{\beta}_* = (\beta_0 + \xi, \beta_1 + \xi, \ldots, \beta_P + \xi)^T$, $\xi = -\big(1 + h_1(\tau)\big)\lambda$ and $\sigma_\epsilon^2 = \left(\sigma^2 + \big(1 + h_2(\tau)\big)\lambda^2\right)$.

So, the log-likelihood function of $\boldsymbol{Y}$ is

$$\log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = -\frac{n\log 2\pi}{2} - \frac{n\log\sigma^2}{2} - n\log\Phi(\tau) - \frac{1}{2}\sum_{i=1}^n D_i(\boldsymbol{\theta}) + \sum_{i=1}^n \log\Phi\big(K_i(\boldsymbol{\theta})\big).$$

3. METHODS OF ESTIMATION

Maximum Likelihood Estimation

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

1. Derivative with respect to $\boldsymbol{\beta}$:

$$\frac{1}{\sigma^2}\sum_{i=1}^n \boldsymbol{x}_i^T\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} + \big(\tau + h_1(\tau)\big)\lambda\right) - \frac{\lambda\sigma^{-2}}{\sqrt{1+\lambda^2\sigma^{-2}}}\sum_{i=1}^n \boldsymbol{x}_i^T\,\frac{\varphi\big(K_i(\boldsymbol{\theta})\big)}{\Phi\big(K_i(\boldsymbol{\theta})\big)} = 0. \tag{3.1}$$

2. Derivative with respect to λ:

$$\frac{\big(\tau + h_1(\tau)\big)}{\sigma^2}\sum_{i=1}^n\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} + \big(\tau + h_1(\tau)\big)\lambda\right) - \sum_{i=1}^n\left(\frac{\lambda^3\big(\tau + h_1(\tau)\big) + \sigma^2\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} + \lambda\big(\tau + 2h_1(\tau)\big)\right)}{\sigma^2(\lambda^2+\sigma^2)\sqrt{1+\lambda^2\sigma^{-2}}}\right)\frac{\varphi\big(K_i(\boldsymbol{\theta})\big)}{\Phi\big(K_i(\boldsymbol{\theta})\big)} = 0. \tag{3.2}$$

3. Derivative with respect to σ^2:

$$\frac{1}{2\sigma^4}\sum_{i=1}^n\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} + \big(\tau + h_1(\tau)\big)\lambda\right)^2 - \frac{n}{2\sigma^2} - \sum_{i=1}^n\left(\frac{\lambda^2\tau(\lambda^2+\sigma^2) + \lambda(\lambda^2+2\sigma^2)\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} + \lambda h_1(\tau)\right)}{2\sigma^4(\lambda^2+\sigma^2)\sqrt{1+\lambda^2\sigma^{-2}}}\right)\frac{\varphi\big(K_i(\boldsymbol{\theta})\big)}{\Phi\big(K_i(\boldsymbol{\theta})\big)} = 0. \tag{3.3}$$

4. Derivative with respect to τ:

$$\frac{n\varphi(\tau)}{\Phi(\tau)}+\frac{1}{\sigma^2}\sum_{i=1}^{n}\lambda\left(1+h_1'(\tau)\right)\left(y_i-\boldsymbol{x}_i^T\boldsymbol{\beta}+\left(\tau+h_1(\tau)\right)\lambda\right)$$
$$-\frac{1}{\sqrt{1+\lambda^2\sigma^{-2}}}\sum_{i=1}^{n}\frac{\varphi\left(K_i(\boldsymbol{\theta})\right)}{\Phi\left(K_i(\boldsymbol{\theta})\right)}\left(1+\lambda^2\sigma^{-2}\left(1+h_1'(\tau)\right)\right)=0. \tag{3.4}$$

We used the numerical technique to solve the equations above.

Least Square Estimation of the Parameter vector $\boldsymbol{\beta}$

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

Theorem 1:

i) $E\left(\widehat{\boldsymbol{\beta}}\right)=E((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y})=\boldsymbol{\beta}$, i.e., $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$.
ii) $Cov\left(\widehat{\boldsymbol{\beta}}\right)=Cov((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y})=\left(\sigma^2+\left(1+h_2(\tau)\right)\lambda^{\mathbf{2}}\right)(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.
iii) $E(s^2)=\sigma^2+\left(1+h_2(\tau)\right)\lambda^{\mathbf{2}}$.

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:

$$Y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\epsilon_i,\, i=1,2,\dots,n,$$

where β_0,β_1 and β_2 are unknown parameters and ϵ_i is the ith error term. The least squares estimator for the parameters are:

$$\begin{cases}\widehat{\beta_0}=\bar{Y}-\widehat{\beta_1}\bar{x}_1-\widehat{\beta_2}\bar{x}_2,\\ \widehat{\beta_1}=\dfrac{S_{x_1y}-\widehat{\beta_2}S_{x_1x_2}}{S_{x_1x_1}},\\ \widehat{\beta_2}=\dfrac{S_{x_1x_1}S_{x_2y}-S_{x_1x_2}S_{x_1y}}{S_{x_1x_1}S_{x_2x_2}-\left(S_{x_1x_2}\right)^2},\end{cases} \tag{4.1}$$

where

$$S_{x_1y}=\sum_{i=1}^{n}x_{i1}y_i-\frac{(\sum_{i=1}^{n}x_{i1})(\sum_{i=1}^{n}y_i)}{n},\, S_{x_2y}=\sum_{i=1}^{n}x_{i2}y_i-\frac{(\sum_{i=1}^{n}x_{i2})(\sum_{i=1}^{n}y_i)}{n},$$
$$S_{x_1x_2}=\sum_{i=1}^{n}x_{i1}x_{i2}-\frac{(\sum_{i=1}^{n}x_{i1})(\sum_{i=1}^{n}x_{i2})}{n},\, S_{x_1x_1}=\sum_{i=1}^{n}x_{i1}^2-\frac{\left(\sum_{i=1}^{n}x_{i1}\right)^2}{n},$$
$$S_{x_2x_2}=\sum_{i=1}^{n}x_{i2}^2-\frac{\left(\sum_{i=1}^{n}x_{i2}\right)^2}{n},\, \bar{x}_1=\frac{1}{n}\sum_{i=1}^{n}x_{i1},\, \bar{x}_2=\frac{1}{n}\sum_{i=1}^{n}x_{i2},$$

and

$$\bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i.$$

Using equation (4.1) and with some straightforward simplification, we have

$$\widehat{\beta_0} - \beta_0 = \sum_{i=1}^{n} a_i \epsilon_i, \tag{4.2}$$

where

$$a_i = \frac{1}{n} + \frac{U(x_{i1} - \bar{x}_1)}{RZ} + \frac{V(x_{i2} - \bar{x}_2)}{RZ}, R = S_{x_1x_1}S_{x_2x_2} - \left(S_{x_1x_2}\right)^2,$$

$$Z = S_{x_2x_2}\left(S_{x_1x_1}\right)^2 - S_{x_1x_1}\left(S_{x_1x_2}\right)^2, U = \bar{x}_2 S_{x_1x_2} Z - \bar{x}_1 S_{x_1x_1} S_{x_2x_2} R,$$

and

$$V = \bar{x}_1 S_{x_1x_2} S_{x_1x_1} R - \bar{x}_2 S_{x_1x_1} Z.$$

Similarly, we can rewrite $\hat{\beta}_1$ and $\hat{\beta}_2$ given in (4.1) to obtain

$$\widehat{\beta_1} - \beta_1 = \sum_{i=1}^{n} b_i\, \epsilon_i, \tag{4.3}$$

where

$$b_i = \frac{S_{x_1x_1}S_{x_2x_2}(x_{i1} - \bar{x}_1) - S_{x_1x_2}S_{x_1x_1}(x_{i2} - \bar{x}_2)}{Z},$$

and

$$\widehat{\beta_2} - \beta_2 = \textstyle\sum_{i=1}^{n} c_i \epsilon_i, \tag{4.4}$$

where

$$c_i = \frac{S_{x_1x_1}(x_{i2} - \bar{x}_2) - S_{x_1x_2}(x_{i1} - \bar{x}_1)}{R}.$$

It is easy to show that the estimators $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$ are unbiased for β_0, β_1 and β_2 respectively and the variances of these estimators as the following:

$$Var\left(\widehat{\beta_0}\right) = \left(\sigma^2 + \left(1 + h_2(\tau)\right)\lambda^2\right)\left\{\frac{1}{n} + \frac{\bar{x}_1^2 S_{x_2x_2} + \bar{x}_2^2 S_{x_1x_1}}{S_{x_1x_1}S_{x_2x_2} - \left(S_{x_1x_2}\right)^2}\right\},$$

$$Var\left(\widehat{\beta_1}\right) = \left(\sigma^2 + \left(1 + h_2(\tau)\right)\lambda^2\right)\frac{S_{x_2x_2}}{S_{x_1x_1}S_{x_2x_2} - \left(S_{x_1x_2}\right)^2},$$

and

$$Var\left(\widehat{\beta_2}\right) = \left(\sigma^2 + \left(1 + h_2(\tau)\right)\lambda^2\right)\frac{S_{x_1x_1}}{S_{x_1x_1}S_{x_2x_2} - \left(S_{x_1x_2}\right)^2}.$$

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, $E(X_n) = 0$, $E(X_n^2) = \sigma_n^2 > 0, \mathrm{S}_n^2 = \sum_{i=1}^{n} {\sigma_i}^2 > 0, \Gamma_n^{2+\delta} = \sum_{i=1}^{n} E|X_i|^{2+\delta} < \infty, n \geq 1,$ for some

$\delta \in (0,1]$ and $W_n = \sum_{i=1}^{n} X_i$, then there is exist a universal constant C_δ such that $\sup_{-\infty<x<\infty}\left|P(W_n \le xS_n) - \phi(x)\right| \le C_\delta \left(\frac{\Gamma_n}{S_n}\right)^{2+\delta}$.

Theorem 3.

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

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{Var}(\hat{\beta}_1)}} \xrightarrow{D} \mathrm{N}(0,1),$$

$$\frac{\hat{\beta}_2 - \beta_2}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} \xrightarrow{D} \mathrm{N}(0,1),$$

and

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\widehat{Var}(\hat{\beta}_0)}} \xrightarrow{D} \mathrm{N}(0,1),$$

where

$$\widehat{Var}(\widehat{\beta_0}) = \frac{SSE}{(n-3)}\left\{\frac{1}{n} + \frac{\bar{x}_1^2 S_{x_2x_2} + \bar{x}_2^2 S_{x_1x_1}}{S_{x_1x_1}S_{x_2x_2} - \left(S_{x_1x_2}\right)^2}\right\}, \widehat{Var}(\widehat{\beta_1}) = \frac{SSE\ S_{x_2x_2}}{(n-3)\left(S_{x_1x_1}S_{x_2x_2} - \left(S_{x_1x_2}\right)^2\right)},$$

and

$$\widehat{Var}(\widehat{\beta_2}) = \frac{SSE\ S_{x_1x_1}}{(n-3)\left(S_{x_1x_1}S_{x_2x_2} - \left(S_{x_1x_2}\right)^2\right)}.$$

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

Parameters	N-MLR			ESN-MLR		
	Estimate	Bias	Mse	Estimate	Bias	Mse
$\beta_0(2)$	2.1268	0.0282	0.0355	2.0485	0.0071	0.0107
$\beta_1(3)$	2.9902	-0.0042	0.0014	3.0156	0.0003	0.0006
$\beta_2(5)$	4.9850	-0.0522	0.1537	4.9676	0.0031	0.0005
$\lambda(0.3)$	-	-	-	0.3184	-0.1418	0.0657
$\sigma^2(1)$	0.8783	-0.0457	0.0085	0.9971	-0.5969	0.5254
$\tau(0.2)$	-	-	-	0.4872	0.09759	0.0363

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 σ^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

Parameters	N-MLR		ESN-MLR	
	Estimate	S.E	Estimate	S.E
$\beta_0(2)$	2.1979	0.1733	2.1117	0.1635
$\beta_1(3)$	3.0797	0.0381	2.9819	0.0295
$\beta_2(5)$	4.9973	0.0439	4.9908	0.0338
$\lambda(0.3)$	-	-	-	-
$\sigma^2(1)$	1.1012	1.0494	0.9020	0.9498
$\tau(0.2)$	-	-	-	-

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 $y =$ time (in seconds) is related to two other predictor variables, namely, $x_1 =$ distance (in mile) and $x_2 =$ 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.

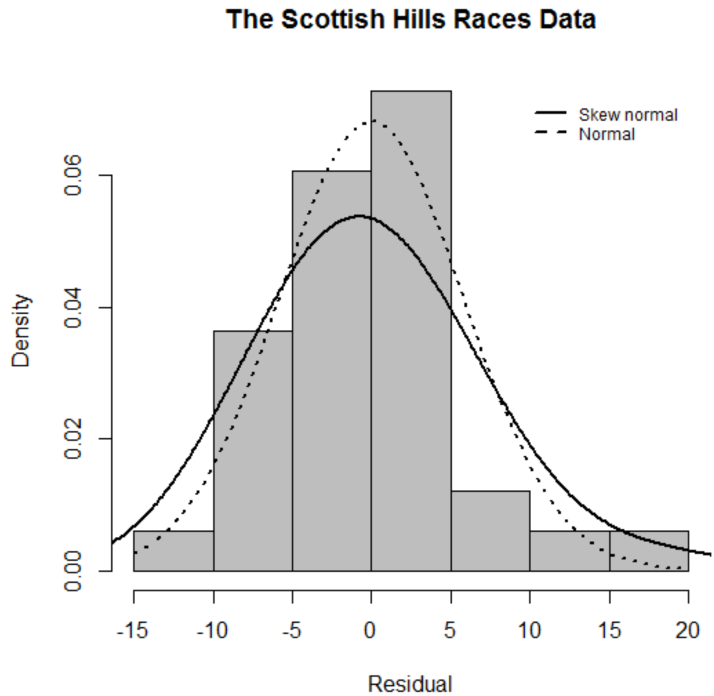

Figure 1: The histogram of the residuals of the Scottish hills races data and the fitted normal and skew normal model

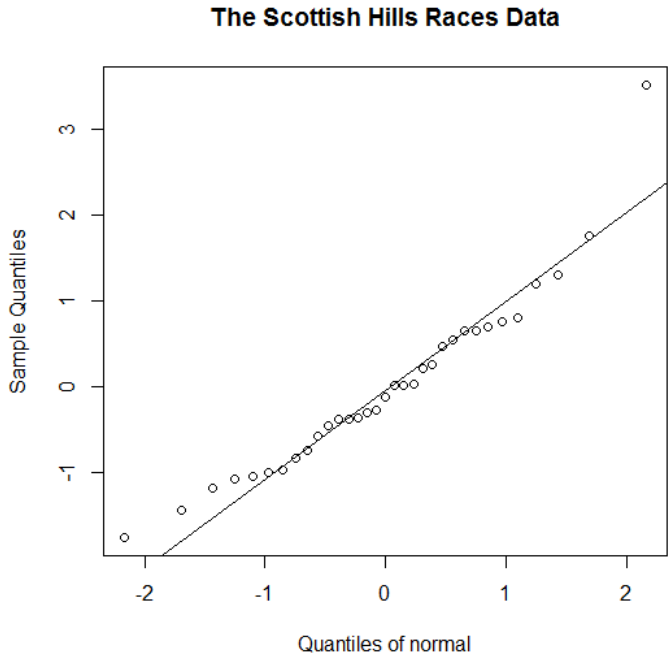

Figure 2: The normal Q-Q plots of the Scottish hills races data

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

Parameters	**N-MLR**		**ESN-MLR**	
	Estimate	**S.E**	**Estimate**	**S.E**
β_0	-10.2728	1.8421	-10.4404	2.1872
β_1	6.72021	0.2469	6.71300	0.2332
β_2	0.00787	0.0010	0.0079	0.009
λ	-	-	0.3439	-
σ^2	34.5378	5.8769	48.4769	6.9625
τ	-	-	1.776735e-09	-
Log-likelihood	-104.6752		-80.2972	
AIC	217.3504		172.5944	

Table 4
Results of Fitting N-MLR and ESN-MLR to the Scottish Hills Races Data Involving Least Squares Estimates and Standard Errors

Parameters	N-MLR		ESN-MLR	
	Estimate	S.E	Estimate	S.E
β_0	-10.3616	1.8976	-10.3616	1.7031
β_1	6.6921	0.2543	6.6921	0.2531
β_2	0.0080	0.0011	0.0080	0.0011
λ	-	-	-	-
σ^2	36.6489	6.054	36.6489	6.051

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

1. Adcock, C.J. (2010). Asset pricing and portfolio selection based on the multivariate extended skew-student–t distribution. *Annals of Operations Research*, 176(1), 221-234.
2. Alodat, M.T., Al-Rawwash, M.Y. and Nawajah, I.M. (2010). Inference about the regression parameters using median-ranked set sampling. *Communications in Statistics–Theory and Methods*, 39(14), 2604-2616.
3. Arellano-Valle, R., Bolfarine, H. and Lachos, V. (2005). Skew normal linear mixed models. *Journal of Data Science*, 3, 415-438.
4. Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.

5. Azzalini, A., and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B*, 61(3), 579-602.
6. Bhattacharya, R.N. and Rao, R. (1976). *Normal Approximation Asymptotic Expansions*. Wiley, New York.
7. Cancho, V.G., Ortega, E.M.M. and Paula, G.A. (2010). On estimation and influence diagnostics for log-Birnbaum-Saunders regression models. *Journal of Statistical Planning and Inference,* 140, 2486-2496.
8. Capitanio, A., Azzalini, A. and Stanghellini, E. (2003). Graphical models for skew normal variates. *Scandinavian Journal of Statistics*, 30, 129-144.
9. Chatterjee, S. and Hadi, A.S. (2012). *Regression analysis by example*. 5^{th} Edition, Wiley, New Jersey.
10. Chow, Y.S. and Teicher, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales*. Springer, New York.
11. Cordeiro, G.M., Ferrari, S.L., Uribe-Opazo, M.A. and Vasconcellos, K.L. (2000). Corrected maximum likelihood estimation in a class of symmetric nonlinear regression models. *Statistics and Probability Letters,* 46, 317-328.
12. Genton, M.G., He, L. and Liu, X. (2001). Moments of skew normal random vectors and their quadratic forms. *Statistics and Probability Letters,* 4, 319-325.
13. Liseo, B. and Loperfido, N. (2003). A Bayesian interpretation of the multivariate skew normal distribution. *Statistics and Probability Letters*, 61, 395-401.
14. Rencher, A. C. and Schaalje, G. B. (2008). *Linear Models in Statistics*. Wiley, New Jersey.
15. Sahu, S.K., Dey, D.K. and Branco, M. (2003). A new class of multivariate skew distributions with application to Bayesian regression models. *Canadian Journal of Statistics,* 31, 129-150.

APPENDIX

Proofs for the asymptotic distribution of the estimators $\widehat{\boldsymbol{\beta}}_{\mathbf{1}}, \widehat{\boldsymbol{\beta}}_{\mathbf{2}}$ and $\widehat{\boldsymbol{\beta}}_{\mathbf{0}}$

Proof 1.

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

$$E(b_i^2\epsilon_i^2) = \left(\frac{S_{x_1x_1}S_{x_2x_2}(x_{i1}-\bar{x}_1)-S_{x_1x_1}S_{x_1x_2}(x_{i2}-\bar{x}_2)}{Z}\right)^2 \left(\sigma^2 + \left(1+h_2(\tau)\right)\lambda^{\mathbf{2}}\right),$$

then

$$\Gamma_n^{2+\delta} = A_\delta \sum_{i=1}^{n} \left|\frac{S_{x_2x_2}(x_{i1}-\bar{x}_1) - S_{x_1x_2}(x_{i2}-\bar{x}_2)}{Z^*}\right|^{2+\delta},$$

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

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

where

$$Q = S_{x_1x_1}\left(S_{x_2x_2}\right)^2 - 2S_{x_2x_2}\left(S_{x_1x_2}\right)^2 + S_{x_2x_2}\left(S_{x_1x_2}\right)^2,$$

and

$$Z^* = S_{x_1x_1}S_{x_2x_2} - \left(S_{x_1x_2}\right)^2.$$

Hence, we have

$$\left(\frac{\Gamma_n}{S_n}\right)^{2+\delta} = \frac{A_\delta \sum_{i=1}^{n}\left|S_{x_2x_2}(x_{i1}-\bar{x}_1) - S_{x_1x_2}(x_{i2}-\bar{x}_2)\,\right|^{2+\delta}}{\left(\sigma^2 + \left(1+h_2(\tau)\right)\lambda^{\mathbf{2}}\right)^{\frac{2+\delta}{2}} Q^{\frac{2+\delta}{2}}}.$$

By applying Berry-Esseen theorem, we obtain

$$\sup_{-\infty<x<\infty}\left|P\left(\sum_{i=1}^{n} b_i\epsilon_i \le y\sqrt{\frac{\left(\sigma^2 + \left(1+h_2(\tau)\right)\lambda^{\mathbf{2}}\right)Q}{{Z^*}^2}}\right) - \Phi(y)\right|$$
$$\le \psi_{\delta,\lambda}\frac{\sum_{i=1}^{n}\left|S_{x_2x_2}(x_{i1}-\bar{x}_1) - S_{x_1x_2}(x_{i2}-\bar{x}_2)\,\right|^{2+\delta}}{Q^{\frac{2+\delta}{2}}},$$

where

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

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

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

By using this inequality and assume that $a_i = \sum_{i=1}^{n}\left|S_{x_2x_2}(x_{i1}-\bar{x}_1) - S_{x_1x_2}(x_{i2}-\bar{x}_2)\,\right|^2$, $p = 2/(2+\delta)$ and for $v > 2$, we have

$$\left(\sum_{i=1}^{n}\left|S_{x_2x_2}(x_{i1}-\bar{x}_1)-S_{x_1x_2}(x_{i2}-\bar{x}_2)\right|^2\right)^{1+\frac{\delta}{2}} \geq n^{\frac{\delta}{v}}\sum_{i=1}^{n}\left|S_{x_2x_2}(x_{i1}-\bar{x}_1)-S_{x_1x_2}(x_{i2}-\bar{x}_2)\right|^{2+\delta}.$$

Then, based on a straightforward simplification and setting $p=2/(2+\delta)$, we get

$$\begin{aligned} n^{\frac{\delta}{v}}\sum_{i=1}^{n}&\left|S_{x_2x_2}(x_{i1}-\bar{x}_1)-S_{x_1x_2}(x_{i2}-\bar{x}_2)\right|^{2+\delta} \\ &= n^{\frac{\delta}{v}+1-\frac{1}{p}}\left(\sum_{i=1}^{n}\left|S_{x_2x_2}(x_{i1}-\bar{x}_1)-S_{x_1x_2}(x_{i2}-\bar{x}_2)\right|^{(2+\delta)p}\right)^{1/p} \\ &\leq n^{\frac{\delta}{2v}(2-v)}\left(\sum_{i=1}^{n}\left|S_{x_2x_2}(x_{i1}-\bar{x}_1)-S_{x_1x_2}(x_{i2}-\bar{x}_2)\right|^2\right)^{1+\frac{\delta}{2}}. \end{aligned}$$

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

$$\sup_{-\infty<x<\infty}\left\{\left|P\left(\sum_{i=1}^{n}b_i\epsilon_i\leq y\sqrt{\frac{\left(\sigma^2+\left(1+h_2(\tau)\right)\lambda^2\right)Q}{{Z^*}^2}}\right)-\Phi(y)\right|\right\}\leq \psi_{\delta,\lambda}\frac{1}{n^{\frac{\delta}{v}}}.$$

Eventually, this implies that

$$\frac{\hat{\beta}_1-\beta_1}{\sqrt{Var(\hat{\beta}_1)}}=\frac{\sum_{i=1}^{n}b_i\epsilon_i}{\sqrt{\frac{\left(\sigma^2+\left(1+h_2(\tau)\right)\lambda^2\right)Q}{{Z^*}^2}}}\xrightarrow{D}N(0,1),$$

where

$$\widehat{Var}(\widehat{\beta_1})=\frac{SSE\ (S_{x_2x_2})}{(n-3)\left(S_{x_1x_1}S_{x_2x_2}-\left(S_{x_1x_2}\right)^2\right)}\xrightarrow{P}Var(\hat{\beta}_1).$$

Then, by Slutsky's theorem, we conclude that

$$\frac{\hat{\beta}_1-\beta_1}{\sqrt{\widehat{Var}(\hat{\beta}_1)}}\xrightarrow{D}N(0,1).$$

Proof 2.

Consider equation (4.4) and we note that $E(c_i\epsilon_i)=0$ for every $i=1,\dots,n\ and$ $\sigma_i^2=E(c_i^2\epsilon_i^2)=\left(\frac{S_{x_1x_1}(x_{i2}-\bar{x}_2)-S_{x_1x_2}(x_{i1}-x_1)}{R}\right)^2\left(\sigma^2+\left(1+h_2(\tau)\right)\lambda^2\right)$. 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,

$$S_n^2 = \frac{\left(\sigma^2 + \left(1+h_2(\tau)\right)\lambda^2\right)W}{R^2},$$

where

$$W = S_{x_2x_2}\left(S_{x_1x_1}\right)^2 - 2S_{x_1x_1}\left(S_{x_1x_2}\right)^2 + S_{x_1x_1}\left(S_{x_1x_2}\right)^2.$$

Hence, we have

$$\left(\frac{\Gamma_n}{S_n}\right)^{2+\delta} = \frac{A_\delta \sum_{i=1}^{n} \left| S_{x_1x_1}(x_{i2}-\bar{x}_2) - S_{x_1x_2}(x_{i1}-\bar{x}_1) \right|^{2+\delta}}{\left(\sigma^2 + \left(1+h_2(\tau)\right)\lambda^2\right)^{\frac{2+\delta}{2}} W^{\frac{2+\delta}{2}}}.$$

After applying Berry-Esseen theorem, we get

$$\sup_{-\infty<x<\infty} \left| P\left(\sum_{i=1}^{n} c_i\epsilon_i \le y \sqrt{\frac{\left(\sigma^2 + \left(1+h_2(\tau)\right)\lambda^2\right)W}{R^2}} \right) - \Phi(y) \right|$$
$$\le \psi_{\delta,\lambda} \frac{\sum_{i=1}^{n} \left| S_{x_1x_1}(x_{i2}-\bar{x}_2) - S_{x_1x_2}(x_{i1}-\bar{x}_1) \right|^{2+\delta}}{W^{\frac{2+\delta}{2}}},$$

where

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

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

$$\left(\sum_{i=1}^{n} \left| S_{x_1x_1}(x_{i2}-\bar{x}_2) - S_{x_1x_2}(x_{i1}-\bar{x}_1) \right|^2 \right)^{1+\frac{\delta}{2}}$$
$$\ge n^{\frac{\delta}{v}} \sum_{i=1}^{n} \left| S_{x_1x_1}(x_{i2}-\bar{x}_2) - S_{x_1x_2}(x_{i1}-\bar{x}_1) \right|^{2+\delta},$$

and with more simplification, we obtain

$$n^{\frac{\delta}{v}}\sum_{i=1}^{n}\left|S_{x_1x_1}(x_{i2}-\bar{x}_2)-S_{x_1x_2}(x_{i1}-\bar{x}_1)\right|^{2+\delta}$$
$$=n^{\frac{\delta}{v}+1-\frac{1}{p}}\left(\sum_{i=1}^{n}\left|S_{x_1x_1}(x_{i2}-\bar{x}_2)-S_{x_1x_2}(x_{i1}-\bar{x}_1)\right|^{(2+\delta)p}\right)^{1/p}$$
$$\leq n^{\frac{\delta}{2v}(2-v)}\left(\sum_{i=1}^{n}\left|S_{x_1x_1}(x_{i2}-\bar{x}_2)-S_{x_1x_2}(x_{i1}-\bar{x}_1)\right|^{2}\right)^{1+\frac{\delta}{2}}.$$

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

$$\sup_{-\infty<x<\infty}\left\{\left|P\left(\sum_{i=1}^{n}c_i\epsilon_i\leq y\sqrt{\frac{\left(\sigma^2+\left(1+h_2(\tau)\right)\lambda^{2}\right)W}{R^2}}\right)-\Phi(y)\right|\right\}\leq \psi_{\delta,\lambda}\frac{1}{n^{\frac{\delta}{v}}}.$$

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

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

where

$$\widehat{Var}(\widehat{\beta_2})=\frac{SSE(\,S_{x_1x_1})}{(n-3)\left(S_{x_1x_1}S_{x_2x_2}-\left(S_{x_1x_2}\right)^2\right)}\xrightarrow{P}Var(\hat{\beta}_2).$$

Then, by Slutsky's theorem, we conclude that

$$\frac{\left(\hat{\beta}_2-\beta_2\right)}{\sqrt{\widehat{Var}(\hat{\beta}_2)}}\xrightarrow{D}N(0,1).$$

Proof 3.

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

$E(a_i\epsilon_i)=0$ for every $i=1,2,,\dots,n$.

Also, from Berry-Esseen theorem, we have

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

where

$$G = R^2Z^2 + n^2U^2S_{x_1x_1} + 2n^2UVS_{x_1x_2} + n^2V^2S_{x_2x_2} + 2nRU(x_{i1} - \bar{x}_1)Z + 2nRV(x_{i2} - \bar{x}_2)Z.$$

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

$$\Gamma_n^{2+\delta} = A_\delta \sum_{i=1}^{n} \left| \frac{RZ + nU(x_{i1} - \bar{x}_1) + nV(x_{i2} - \bar{x}_2)}{nRZ} \right|^{2+\delta},$$

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

By combining the two expressions of Γ_n and S_n, we get

$$\left(\frac{\Gamma_n}{S_n}\right)^{2+\delta} = \frac{A_\delta \sum_{i=1}^{n} |RZ + nU(x_{i1} - \bar{x}_1) + nV(x_{i2} - \bar{x}_2)|^{2+\delta}}{\left(\sigma^2 + \left(1 + h_2(\tau)\right)\lambda^2\right)^{\frac{2+\delta}{2}} G^{\frac{2+\delta}{2}}}.$$

Thus, we get

$$\sup_{-\infty<x<\infty} \left| P\left(\sum_{i=1}^{n} a_i\epsilon_i \le y \sqrt{\frac{\left(\sigma^2 + \left(1 + h_2(\tau)\right)\lambda^2\right)G}{(nRZ)^2}} \right) - \Phi(y) \right| \le \frac{C_\delta A_\delta \sum_{i=1}^{n} |RZ + nU(x_{i1} - \bar{x}_1) + nV(x_{i2} - \bar{x}_2)|^{2+\delta}}{\left(\sigma^2 + \left(1 + h_2(\tau)\right)\lambda^2\right)^{\frac{2+\delta}{2}} G^{\frac{2+\delta}{2}}},$$

which leads to the following conclusion:

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

where

$$\widehat{Var}(\widehat{\beta_0}) = \frac{SSE}{(n-3)} \left\{ \frac{1}{n} + \frac{\bar{x}_1^2 S_{x_2x_2} + \bar{x}_2^2 S_{x_1x_1}}{S_{x_1x_1}S_{x_2x_2} - \left(S_{x_1x_2}\right)^2} \right\} \xrightarrow{P} Var(\hat{\beta}_0).$$

Then, by Slutsky's theorem, we conclude that

$$\frac{\left(\hat{\beta}_0 - \beta_0\right)}{\sqrt{\widehat{Var}(\hat{\beta}_0)}} \xrightarrow{D} N(0,1).$$