Ingeniería y Ciencia
ISSN:1794-9165 | ISSN-e: 2256-4314
ing. cienc., vol. 17, no. 33, pp. 45–70, enero-junio. 2021.
http://www.eafit.edu.co/ingciencia

Inference in Multiple Linear Regression Model with Generalized Secant Hyperbolic Distribution Errors

Alvaro Alexander Burbano Moreno[1], Oscar Orlando Melo Matínez[2] and M. Qamarul Islam[3]

Received: 30-07-2020 | Accepted: 16-12-2020 | Online: 10-05-2021

MSC: 62J20, 60E05

doi:10.17230/ingciencia.17.33.3

Abstract

We study multiple linear regression model under non-normally distributed random error by considering the family of generalized secant hyperbolic distributions. We derive the estimators of model parameters by using modified maximum likelihood methodology and explore the properties of the modified maximum likelihood estimators so obtained. We show that the proposed estimators are more efficient and robust than the commonly used least square estimators. We also develop the relevant test of hypothesis procedures and compared the performance of such tests vis-a-vis the classical tests that are based upon the least square approach.

Keywords: Maximum likelihood; modified maximum likelihood; least square; generalized secant hyperbolic distribution; robustness; hypothesis testing.

[1] Universidade Federal de Minas Gerais, aaburbanom@ufmg.br, Belo Horizonte, Brasil.
[2] Universidad Nacional de Colombia, oomelom@unal.edu.co, Bogotá, Colombia.
[3] Middle East Technical University, mislam@metu.edu.tr, Ankara, Turkey.

Inferencia en modelo de regresión lineal múltiple con errores de distribución secante hiperbólica generalizada

Resumen
Estudiamos el modelo de regresión lineal múltiple bajo errores aleatorios no distribuidos normalmente considerando la familia de distribuciones hiperbólicas secantes generalizadas. Derivamos los estimadores de los parámetros del modelo utilizando la metodología modificada de máxima verosimilitud y exploramos las propiedades de los estimadores modificados de máxima verosimilitud así obtenidos. Mostramos que los estimadores propuestos son más eficientes y robustos que los estimadores de mínimos cuadrados comúnmente utilizados. También desarrollamos la prueba relevante de los procedimientos de hipótesis y comparamos el rendimiento de tales pruebas con las pruebas clásicas que se basan en el enfoque de mínimos cuadrados.

Palabras clave: Máxima verosimilitud; máxima verosimilitud modificada; mínimo cuadrados; distribución secante hiperbólica generalizada; robustez; prueba de hipótesis.

1 Introduction

In most applications of multiple linear regression (MLR) the random errors involved are assumed to have a normal distribution. However, there are situations in which a non-normal distribution for the errors may be an appropriate alternative to the normal one; see, e.g., Pearson [1], Huber [2]. Whereas, it is common to use least square (LS) method as a tool for estimating the parameter of a MLR model, it is known that the resulting LS estimators (LSE) are substantially less efficient when normality assumption is violated (see, Tukey [3]). Moreover, the LSE are prone to various data anomalies, e.g., presence of outliers in the sample. Alternatively, one can suggest to use the well-known maximum likelihood (ML) method that provides ML estimators (MLE) having many attractive properties. However, the method may not provide explicit solution in problems involving non-normal distributions as the likelihood equations to be solved are in intricate nonlinear form. One can get numerical solution but it may be riddled with difficulties. The iterative solution may not converge or may converge to wrong value, particularly when the data contains outliers (see, Barnett [4], Vaughan [5]). A modification in ML methods is suggested by Tiku and Suresh [6] resulting the estimators that are known as modified maximum likelihood estimators (MMLE). Furthermore, Vaughan and Tiku [7] showed that the MMLE are equivalent to MLE, asymptotically. For location-scale distributions the MMLE can be obtained in analytically closed form and they are found to be unbiased and substantially more efficient than the LSE. Robustness to outliers and to other data anomalies are also associated with these estimators (see, Tiku and Akkaya [8]).

The latest contribution, related with multiple regression analysis, is that of Islam and Tiku [9]. They considered three families of non-normal distributions: (a) Symmetric long-tailed distributions, (b) Symmetric short-tailed distributions, and (c) Generalized logistic distributions. A more general and flexible family of symmetric distributions named as generalized secant hyperbolic (GSH) family is introduced by Vaughan [10] and the MMLE are derived for its parameters. This family consists of symmetric distributions, with kurtosis ranging from 1.8 to infinity, i.e., log-tailed (kurtosis greater than 3) and short-tailed (Kurtosis less than 3), includes the logistic as a special case, the uniform as a limiting case, and closely approximates normal and Student t with corresponding kurtosis. Later, Yilmaz and Akkaya [11] presented one-way classification model in experimental design with errors having GSH distribution. These results motivated us to study the MLR models under the assumption that the distribution of the errors belongs to the GSH family. We derive the MMLE of the parameters in the model and compare them with the LSE. We show that the MMLE are considerably more efficient and robust than the LSE. We will use the classical frequency method of construction of the statistical tests, that is, the T-statistic and the F-statistic.

2 Multiple linear regression model

Consider the model

$$y_i = \theta_0 + \sum_{j=1}^{k} \theta_j x_{ij} + e_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where y_i denotes the ith observation on the dependent variable, x_{ij} the ith observation on the jth independent variable, e_i is the random error, and

θ_0 and θ_j $(j = 1, 2, \ldots, k)$ are the regression parameters. The model in (1) can be written in matrix form as follows.

$$\boldsymbol{Y} = \boldsymbol{1}\theta_0 + \boldsymbol{X\theta} + \boldsymbol{e}, \tag{2}$$

where

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \ \boldsymbol{1}_{(n\times 1)} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix}, \ \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_k \end{bmatrix}, \ \boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$$

Suppose that the errors e_i are independently and identically distributed as GSH with probability density function

$$GSH(0, \sigma^2; h) : f(e) = \frac{c_1}{\sigma} \frac{\exp\left(c_2(e/\sigma)\right)}{\exp(2c_2(e/\sigma)) + 2a\exp(c_2(e/\sigma)) + 1} \quad (-\infty < e < \infty), \tag{3}$$

where

$$a = \begin{cases} \cos(h), & \text{for } -\pi < h \leq 0 \\ \cosh(h), & \text{for } h > 0 \end{cases}, \quad c_2 = \begin{cases} \sqrt{\dfrac{\pi^2 - h^2}{3}}, & \text{for } -\pi < h \leq 0 \\ \sqrt{\dfrac{\pi^2 + h^2}{3}}, & \text{for } h > 0 \end{cases},$$

$$\text{and} \qquad c_1 = \begin{cases} \dfrac{\sin(h)}{h} c_2, & \text{for } -\pi < h \leq 0 \\ \dfrac{\sinh(h)}{h} c_2, & \text{for } h > 0 \end{cases}.$$

The shape parameter h controls the amount of kurtosis. In particular, the usual coefficient of kurtosis $\mathbb{K}(h)$ i.e. the fourth standardized moment. For $h > \pi$, $h < \pi$ and $h = \pi$, $(GSH0, \sigma^2; h)$ represents short-tailed, long-tailed and approximately normal distributions.

The LSE of θ_0 and $\boldsymbol{\theta}$ are obtained as

$$\widetilde{\theta}_0 = \overline{y} - \sum_{j=1}^{k} \widetilde{\theta}_j \overline{x}_j \quad , \quad \widetilde{\boldsymbol{\theta}} = (\boldsymbol{\mathcal{X}}'\boldsymbol{\mathcal{X}})^{-1}\boldsymbol{\mathcal{X}}'\boldsymbol{\mathcal{Y}}, \tag{4}$$

where $\boldsymbol{\mathcal{X}}$ is a matrix with elements $(x_{ij} - \overline{x}_j)$ and $\boldsymbol{\mathcal{Y}}$ is a vector with elements $(y_i - \overline{y})$, $(1 \leq i \leq n,\ 1 \leq j \leq k)$, $\overline{x}_j = \sum_{i=1}^{n} x_{ij}/n$ and

$\overline{y} = \sum_{i=1}^{n} y_i/n$. The LSE of the scale parameter σ is given by

$$\widetilde{\sigma} = \sqrt{\sum_{i=1}^{n}\left\{y_i - \overline{y} - \sum_{j=1}^{k}\widetilde{\theta}_j(x_{ij} - \overline{x}_j)\right\}^2 \Big/ (n-k-1)}. \tag{5}$$

3 Modified maximum likelihood

The likelihood function is given by

$$L = \left(\frac{c_1}{\sigma}\right)^n \prod_{i=1}^{n} \frac{\exp\left(c_2(e_i/\sigma)\right)}{\exp(2c_2(e_i/\sigma)) + 2a\exp(c_2(e_i/\sigma)) + 1}.$$

The MLE are the solution of the equations

$$\partial \ln L/\partial\theta_0 = 0, \;\; \partial \ln L/\partial\theta_j = 0, \;\; \text{and} \;\; \partial \ln L/\partial\sigma = 0. \tag{6}$$

The equations have no explicit solutions. The only way to solve them is by iteration, but that is problematic, see Barnett [12], Vaughan [5], Tiku and Akkaya [8]. If the data contain atypical values, the iterations with these equations are often non-convergent or converge to wrong values (for details, see Puthenpura and Sinha [13]). In order to overcome such difficulties, we modify these equations and obtained an asymptotically equivalent set of equations. First the equations (6) are expressed in terms of ordered variates $e_{(i)}$ obtained by ordering (ascending) $e_i = y_i - \theta_0 - \sum_{j=1}^{k}\theta_j x_{ij}$ $(1 \leq i \leq n)$ such that

$$e_{(i)} = y_{[i]} - \theta_0 - \sum_{j=1}^{k}\theta_j x_{[i]j}, \tag{7}$$

where the vector $\left(y_{[i]}, x_{[i]1}, \ldots, x_{[i]k}\right)$, said to be concomitant vector, is the vector $(y_i, x_{i1}, \ldots, x_{ik})$ of observations that corresponds to $e_{(i)}$. The likelihood equations (6) are now expressions in terms of the functions

$$g(z_{(i)}) = (\exp(2c_2 z_{(i)}) + a\exp(c_2 z_{(i)}))/\left(\exp(2c_2 z_{(i)}) + 2a\exp(c_2 z_{(i)}) + 1\right),$$

where $z_{(i)} = e_{(i)}/\sigma$.

The modified likelihood equations

$$\partial \ln L^*/\partial\theta_0 = 0, \;\; \partial \ln L^*/\partial\theta_j = 0, \;\; \text{and} \;\; \partial \ln L^*/\partial\sigma = 0,$$

are obtained replacing $g(z_{(i)})$ by linear functions $g(z_{(i)}) \cong \alpha_i + \beta_i z_{(i)}$ $(1 \leq i \leq n)$. The coefficients α_i and β_i are obtained from the first two terms of a Taylor series expansion of $g(z_{(i)})$ around $t_{(i)} = E(z_{(i)})$. They are

$$\alpha_i = \frac{\exp(2c_2 t_{(i)}) + a\exp(c_2 t_{(i)})}{\exp(2c_2 t_{(i)}) + 2a\exp(c_2 t_i) + 1} - \beta_i t_{(i)} \tag{8}$$

and
$$\beta_i = \frac{ac_2\exp(3c_2 t_{(i)}) + 2c_2\exp(2c_2 t_{(i)}) + ac_2\exp(c_2 t_{(i)})}{\left(\exp(2c_2 t_{(i)}) + 2a\exp(c_2 t_{(i)}) + 1\right)^2}. \tag{9}$$

Although the formulation to obtain the exact values of the expected values $t_{(i)}$ $(1 \leq i \leq n)$ is available at Vaughan [10], it is difficult to implement. Therefore, we use the following approximate values which is often done in practice, see Senoglu and Tiku [14], Tiku and Akkaya [8], Vaughan and Tiku [7],

$$t_{(i)} = \begin{cases} \dfrac{1}{c_2}\ln\left(\dfrac{\sin(hq_i)}{\sin(h(1-q_i))}\right), & \text{for } -\pi < h < 0; \\ \dfrac{\sqrt{3}}{\pi}\ln\left(\dfrac{q_i}{1-q_i}\right), & \text{for } h = 0; \\ \dfrac{1}{c_2}\ln\left(\dfrac{\sinh(hq_i)}{\sinh(h(1-q_i))}\right), & \text{for } h > 0, \end{cases} \tag{10}$$

where $q_i = i/(n+1)$.

The solutions of the modified likelihood equations give the MMLE of θ_0, $\boldsymbol{\theta}$, and σ as follows.

$$\widehat{\theta}_0 = \bar{y}_{[.]} - \sum_{j=1}^{k}\widehat{\theta}_j \bar{x}_{[.]j}, \tag{11}$$

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X}'_{[.]}\boldsymbol{\beta}\boldsymbol{X}_{[.]})^{-1}\left[(\boldsymbol{X}'_{[.]}\boldsymbol{\beta}\boldsymbol{Y}_{[.]}) - \widehat{\sigma}(\boldsymbol{X}'\boldsymbol{\Delta}\mathbf{1})\right], \tag{12}$$

and
$$\widehat{\sigma} = \left\{-c_2 B + \sqrt{(c_2 B)^2 + 2nc_2 C}\right\}\Big/ n, \tag{13}$$

where, $Y_{[i]} = y_{[i]} - \bar{y}_{[.]}$, $X_{[i]j} = x_{[i]j} - \bar{x}_{[.]j}$ $(1 \leq j \leq k)$,

$$\boldsymbol{Y}_{[.]} = \begin{bmatrix} Y_{[1]} \\ Y_{[2]} \\ \vdots \\ Y_{[n]} \end{bmatrix}, \quad \mathbf{1}_{(n\times 1)} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \boldsymbol{X}_{[.]} = \begin{bmatrix} X_{[1]1} & X_{[1]2} & \dots & X_{[1]k} \\ X_{[2]2} & X_{[2]2} & \dots & X_{[2]k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{[n]1} & X_{[n]2} & \dots & X_{[n]k} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \beta_n \end{bmatrix}, \quad \boldsymbol{\Delta} = \begin{bmatrix} \Delta_1 & 0 & \dots & 0 \\ 0 & \Delta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Delta_n \end{bmatrix},$$

$$\bar{x}_{[.]j} = \sum_{i=1}^{n} \beta_i x_{[i]j} \Big/ \sum_{i=1}^{n} \beta_i, \ \bar{y} = \sum_{i=1}^{n} \beta_i y_{[i]} \Big/ \sum_{i=1}^{n} \beta_i, \ \Delta_i = 1/2 - \alpha_i,$$

$$\boldsymbol{K} = (\boldsymbol{X}'_{[.]}\boldsymbol{\beta}\boldsymbol{X}_{[.]})^{-1}(\boldsymbol{X}'_{[.]}\boldsymbol{\beta}\boldsymbol{Y}_{[.]}) = (K_j) \ \ (1 \le j \le k),$$

$$B = \sum_{i=1}^{n} \Delta_i \left\{ y_{[i]} - \overline{y}_{[.]} - \sum_{j=1}^{k} K_j (x_{[i]j} - \overline{x}_{[.]j}) \right\},$$

$$C = \sum_{i=1}^{n} \beta_i \left\{ y_{[i]} - \overline{y}_{[.]} - \sum_{j=1}^{k} K_j (x_{[i]j} - \overline{x}_{[.]j}) \right\}^2 .$$

Computations. In order to initialize ordering of e_i, we first calculate the LSE $\widetilde{\theta}_0$ and $\widetilde{\boldsymbol{\theta}}$ using (4) and obtain the estimated residuals

$$\widetilde{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{1}\widetilde{\theta}_0 - \boldsymbol{X}\widetilde{\boldsymbol{\theta}}. \tag{14}$$

So, the ith concomitant vector $\left(y_{[i]}, x_{[i]1}, \dots, x_{[i]k}\right)$ corresponds to ith ordered value $\widetilde{e}_{(i)}$ in $\widetilde{\boldsymbol{e}}$. The MMLE are then calculated from (11-12) and are replaced in (14) in order to get a new concomitant vector to be used to compute the final MMLE. It is found that the estimates are stabilized in these two steps.

4 Asymptotic variances and relative efficiencies

It is known that MMLE are asymptotically equivalent to the MLE (fot a rigorous proof see Appendix A, Vaughan and Tiku [7]). Hence, the Asymptotic variances and covariances of the MMLE are given by I^{-1}, where I Is the Fisher information matrix given below.

$$I = P \begin{bmatrix} 1 & \overline{x}_{.1} & \overline{x}_{.2} & \cdots & \overline{x}_{.k} & 0 \\ \overline{x}_{.1} & \frac{1}{n}\sum_{i=1}^{n} x_{i1}^2 & \frac{1}{n}\sum_{i=1}^{n} x_{i2}x_{i1} & \cdots & \frac{1}{n}\sum_{i=1}^{n} x_{ik}x_{i1} & 0 \\ \overline{x}_{.2} & \frac{1}{n}\sum_{i=1}^{n} x_{i1}x_{i2} & \frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 & \cdots & \frac{1}{n}\sum_{i=1}^{n} x_{ik}x_{i2} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \overline{x}_{.k} & \frac{1}{n}\sum_{i=1}^{n} x_{i1}x_{ik} & \frac{1}{n}\sum_{i=1}^{n} x_{i2}x_{ik} & \cdots & \frac{1}{n}\sum_{i=1}^{n} x_{ik}^2 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{P}\left\{-E\left(\frac{\partial^2 \ln L}{\partial \sigma^2}\right)\right\} \end{bmatrix} \tag{15}$$

where

$$P = \begin{cases} -\dfrac{c_2^2 n\left(h - \sin(h)\cos(h)\right)}{2\sigma^2 h \sin^2(h)}, & \text{for } -\pi < h < 0; \\ -\dfrac{nc_2^2}{3\sigma^2}, & \text{for } h = 0; \\ -\dfrac{c_2^2 n\left(\sinh(h)\cosh(h) - h\right)}{2\sigma^2 h \sinh^2(h)}, & \text{for } h > 0, \end{cases}$$

$$E\left(\frac{\partial^2 \ln L}{\partial \sigma^2}\right) = \begin{cases} -\dfrac{n}{6\sigma^2}\left(\dfrac{\pi^2 - h^2}{\sin^2(h)} - \dfrac{(\pi^2 - 3h^2)\cos(h)}{h\sin(h)}\right), & \text{for } -\pi < h < 0; \\ -\dfrac{n(3+\pi^2)}{9\sigma^2}, & \text{for } h = 0; \\ -\dfrac{n}{6\sigma^2}\left(\dfrac{(\pi^2 + 3h^2)\cosh(h)}{h\sinh(h)} - \dfrac{\pi^2 + h^2}{\sinh^2(h)}\right), & \text{for } h > 0. \end{cases}$$

In order to show that the MMLE are remarkably efficient, we provide in Table 1 the asymptotic (Asy.) variances obtained from (15) and the simulated (Sim.) variances of the estimates for $k = 3$. The small differences observed between the two sets of values support our claim. All the simulated values in paper are based on $[100000/n]$ (integer value) Monte Carlo runs. A set of design points x_{ij} $(1 \le i \le n, 1 \le j \le k)$ was randomly generated from a $N(0,1)$ and are kept common to all random samples.
We now evaluate the relative efficiencies (RE) of the MMLE, relative to the LSE. We have simulated their means and variances for $h = -\pi/2, 0, \pi\sqrt{3}, \pi\sqrt{11}$, and $n = 30, 50, 100$ $(k = 3)$. Without loss of generality, we take $\theta_0 = 0$,

$\theta_j = 1$ $(1 \leq j \leq k)$ and $\sigma = 1$. The relative efficiencies (%) are calculated as follows,

$$RE = 100 * variance(MMLE)/variance(LSE).$$

Both estimators have negligible biases, and the MMLE are found to be substantially more efficient than the LSE. The results are presented in Table 2.

Table 1: Values of $n \times Variance$ of the MML Estimators $(k = 3)$.

	θ_0		θ_1		θ_2		θ_3		σ	
h	Asy.	Sim.	Asy.	Sim.	Asy.	Sim.	Asy.	Sim.	Asy.	Sim.
					$n = 50$					
$-\pi/4$	0.89	0.94	1.05	1.09	1.05	1.08	1.12	1.18	0.72	0.73
0	0.91	0.96	1.04	1.10	1.07	1.10	1.15	1.20	0.70	0.71
$\pi\sqrt{3}$	0.83	0.89	0.94	0.98	0.98	1.02	1.04	1.08	0.33	0.36
$\pi\sqrt{11}$	0.53	0.56	0.60	0.68	0.62	0.66	0.66	0.69	0.19	0.21
					$n = 100$					
$-\pi/4$	0.89	0.92	0.92	0.94	0.94	0.96	0.98	1.01	0.72	0.72
0	0.91	0.94	0.94	0.97	0.96	0.98	1.00	1.02	0.70	0.70
$\pi\sqrt{3}$	0.83	0.86	0.85	0.88	0.87	0.88	0.91	0.93	0.33	0.34
$\pi\sqrt{11}$	0.53	0.55	0.54	0.58	0.55	0.58	0.58	0.59	0.19	0.20

5 Hypothesis testing

5.1 Individual hypothesis

Let, we want to test a null hypothesis $H_{0j}: \theta_j = 0$ against an alternative $H_{1j}: \theta_j \neq 0$ for each j $(j = 1, 2, \ldots, k)$. It can be shown easily that the modified likelihood equation for $\widehat{\boldsymbol{\theta}}$ can rearranged to be written as

$$\frac{\partial lnL}{\partial \boldsymbol{\theta}} \approx \frac{2c_2}{\sigma^2}(\boldsymbol{X}'_{[.]}\boldsymbol{\beta}\boldsymbol{X}_{[.]})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}).$$

Hence, $\widehat{\boldsymbol{\theta}}$ (given σ) is best asymptotically normal estimator of $\boldsymbol{\theta}$ with the variance given by

$$V(\widehat{\boldsymbol{\theta}}) = \frac{\sigma^2}{2c_2}(\boldsymbol{X}'_{[.]}\boldsymbol{\beta}\boldsymbol{X}_{[.]})^{-1}.$$

Table 2: Mean, variance and relative efficiencies of the MMLE and the LSE, $\theta_0 = 0$, $\theta_j = 1$ $(1 \le j \le k)$, $\sigma = 1$, $k = 3$.

	Mean		Variance		RE	Mean		Variance		RE
	MML	LS	MML	LS	(%)	MML	LS	MML	LS	(%)
			$h=-\pi/2$					$h=0$		
					$n=30$					
θ_0	-0.001	-0.002	0.030	0.035	87	-0.003	-0.003	0.032	0.034	95
θ_1	1.000	1.002	0.038	0.042	88	1.002	1.002	0.028	0.029	95
θ_2	1.000	1.001	0.060	0.067	90	0.997	0.997	0.041	0.043	97
θ_3	0.998	0.997	0.044	0.049	90	1.002	1.002	0.039	0.041	94
σ	0.940	0.987	0.026	0.033	79	0.943	0.987	0.023	0.028	84
					$n=50$					
θ_0	-0.001	-0.002	0.018	0.021	83	-0.002	-0.003	0.019	0.020	95
θ_1	1.001	1.006	0.026	0.030	88	0.997	0.997	0.021	0.022	94
θ_2	0.996	0.996	0.019	0.022	87	1.000	1.002	0.023	0.024	94
θ_3	1.001	0.999	0.022	0.026	87	1.002	1.002	0.023	0.024	93
σ	0.967	0.994	0.016	0.019	81	0.967	0.994	0.014	0.017	86
					$n=100$					
θ_0	0.000	-0.002	0.008	0.010	81	-0.003	-0.004	0.010	0.010	96
θ_1	0.998	0.999	0.011	0.013	85	0.998	0.997	0.010	0.011	92
θ_2	1.005	1.004	0.010	0.012	83	1.001	1.001	0.010	0.011	93
θ_3	0.997	0.997	0.008	0.010	82	1.003	1.002	0.013	0.014	93
σ	0.986	0.999	0.008	0.009	84	0.985	0.998	0.007	0.008	87
			$h=\pi\sqrt{3}$					$h=\pi\sqrt{11}$		
					$n=30$					
θ_0	-0.003	-0.003	0.031	0.034	91	-0.002	-0.001	0.031	0.035	90
θ_1	1.002	1.002	0.027	0.029	93	1.001	1.001	0.038	0.042	90
θ_2	0.997	0.997	0.040	0.043	93	1.001	0.998	0.059	0.068	86
θ_3	1.001	1.002	0.039	0.041	95	0.998	0.999	0.043	0.048	89
σ	0.932	0.995	0.021	0.014	86	0.881	0.997	0.011	0.011	99
					$n=50$					
θ_0	-0.002	-0.003	0.018	0.020	87	-0.001	-0.001	0.015	0.021	70
θ_1	0.997	0.997	0.020	0.022	94	1.006	1.004	0.021	0.030	70
θ_2	1.003	1.001	0.022	0.024	89	0.998	0.996	0.015	0.022	69
θ_3	1.003	1.002	0.021	0.024	88	0.998	1.001	0.017	0.026	67
σ	0.963	0.998	0.007	0.008	91	0.953	1.000	0.004	0.006	78
					$n=100$					
θ_0	-0.004	-0.003	0.009	0.011	82	-0.003	-0.001	0.006	0.010	64
θ_1	0.996	0.997	0.009	0.011	86	1.002	0.999	0.008	0.013	60
θ_2	1.000	1.000	0.009	0.011	86	1.001	1.004	0.007	0.012	60
θ_3	1.001	1.002	0.012	0.014	87	0.999	0.996	0.006	0.010	58
σ	0.983	1.000	0.003	0.004	93	0.981	1.001	0.002	0.003	75

Therefore, we can propose the following test-statistics based upon the MMLE.

$$T_j = \sqrt{2c_2}\frac{\widehat{\boldsymbol{\theta}}_j}{\widehat{\sigma}\sqrt{(\boldsymbol{X}'_{[.]}\boldsymbol{\beta}\boldsymbol{X}_{[.]})^{-1}|_{jj}}} \tag{16}$$

Corresponding test statistic based upon the LSE can be taken as

$$T_j^* = \frac{\widetilde{\boldsymbol{\theta}}_j}{\widetilde{\sigma}\sqrt{(\boldsymbol{X}'\boldsymbol{X})^{-1}|_{jj}}}. \tag{17}$$

Here, for a symmetric matrix M, $M|_{jj}$ denotes the jth diagonal element of it. The null distribution of both T_j and T_j^* $(j = 1, 2, \ldots, k)$ statistics can be taken as Student's t with $n - k - 1$ degrees of freedom.

5.2 Joint hypothesis

For testing a null hypothesis $H_0 : \theta_1 = \theta_2 = \ldots = \theta_k = 0$ against an alternative hypothesis $H_1 : \theta_j \neq 0$ for at least one j $(j = 1, 2, \ldots, k)$. We consider decomposition of sum of squares in the same way as we do in normal samples.
Under H_0, the MMLE of σ is

$$\widehat{\sigma}_0 = \frac{-c_2 B_0 + \sqrt{(c_2 B_0)^2 + 2nc_2 C_0}}{n},$$

where $B_0 = \sum_{i=1}^{n} \Delta_i \left\{y_{[i]} - \overline{y}_{[.]}\right\} = \boldsymbol{\Delta Y}_{[.]}$ and $C_0 = \sum_{i=1}^{n} \beta_i \left\{y_{[i]} - \overline{y}_{[.]}\right\}^2 = \boldsymbol{Y}_{[.]}^{'}\boldsymbol{\beta}\boldsymbol{Y}_{[.]}$.

Now $\widehat{\sigma}_0$ is rewritten as

$$\widehat{\sigma}_0 = \frac{\sqrt{nC_0}}{n}\left[-c_2\left(\frac{B_0}{\sqrt{nC_0}}\right) + \sqrt{c_2^2\left(\frac{B_0}{\sqrt{nC_0}}\right)^2 + 2c_2}\right].$$

For large n, $\dfrac{B_0}{\sqrt{nC_0}} \cong 0$ (for details, see Vaughan and Tiku [7]) and we have

$$n\widehat{\sigma}_0^2 \cong 2c_2 C_0 = 2c_2\boldsymbol{Y}_{[.]}^{'}\boldsymbol{\beta}\boldsymbol{Y}_{[.]}. \tag{18}$$

Under H_1, the MMLE of σ is

$$\widehat{\sigma} = \frac{-c_2 B + \sqrt{(c_2 B)^2 + 2nc_2 C}}{n},$$

where

$$B = \sum_{i=1}^{n} \Delta_i \left\{ y_{[i]} - \overline{y}_{[.]} - \sum_{j=1}^{k} \theta_j (x_{[i]j} - \overline{x}_{[.]j}) \right\} = \boldsymbol{\Delta} \left(\boldsymbol{Y}_{[.]} - \boldsymbol{X}_{[.]} \boldsymbol{\theta} \right),$$

and

$$C = \sum_{i=1}^{n} \beta_i \left\{ y_{[i]} - \overline{y}_{[.]} - \sum_{j=1}^{k} \theta_j (x_{[i]j} - \overline{x}_{[.]j}) \right\}^2 = \boldsymbol{Y}'_{[.]} \boldsymbol{\beta} \boldsymbol{Y}_{[.]} - \boldsymbol{\theta}' \boldsymbol{X}'_{[.]} \boldsymbol{\beta} \boldsymbol{Y}_{[.]}.$$

For n large, $\dfrac{B}{\sqrt{nC}} \cong 0$ and we have

$$n\widehat{\sigma}^2 \cong 2c_2 C = 2c_2 \left(\boldsymbol{Y}'_{[.]} \boldsymbol{\beta} \boldsymbol{Y}_{[.]} - \boldsymbol{\theta}' \boldsymbol{X}'_{[.]} \boldsymbol{\beta} \boldsymbol{Y}_{[.]} \right). \tag{19}$$

Therefore, the decomposition of sum of square of the observations is express in two independent terms: the first term containing the variability explained by the regression and the second term reflects the variability due to the random errors. Thus,

$$\begin{aligned}
\text{Total variability} &= 2c_2 \boldsymbol{Y}'_{[.]} \boldsymbol{\beta} \boldsymbol{Y}_{[.]} = SST, \\
\text{Explained Variability} &= 2c_2 \widehat{\boldsymbol{\theta}}' \boldsymbol{X}'_{[.]} \boldsymbol{\beta} \boldsymbol{Y}_{[.]} = SSR, \\
\text{Unexplained variability} &= 2c_2 \boldsymbol{Y}'_{[.]} \boldsymbol{\beta} \boldsymbol{Y}_{[.]} - 2c_2 \widehat{\boldsymbol{\theta}}' \boldsymbol{X}'_{[.]} \boldsymbol{\beta} \boldsymbol{Y}_{[.]} = SSE.
\end{aligned}$$

Hence, we have the decomposition of the total sum of squares as follows:

$$SST = SSR + SSE. \tag{20}$$

Asymptotically, SST/σ^2, SSR/σ^2 and SSE/σ^2 are distributed as Chi-square with $n-1$, k and $n-k-1$ degrees of freedom, respectively (see, Islam et al.[15], Tiku et al. [16]). Hence, SSR/σ^2 and SSE/σ^2 are independently distributed chi-square random variables. Therefore, if the null hypothesis is true, we can use the following MML test-statistic in order to test our hypothesis.

$$F = \frac{2c_2 \left(\widehat{\boldsymbol{\theta}}' \boldsymbol{X}'_{[.]} \boldsymbol{\beta} \boldsymbol{Y}_{[.]} \right)}{k\widehat{\sigma}^2}, \tag{21}$$

where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\sigma}$ are the MMLE of the corresponding parameters.
The test based upon the LSE can be carried out by using the following test statistic.

$$F^{*} = \frac{\widetilde{\boldsymbol{\theta}}'\boldsymbol{X}'\boldsymbol{Y}}{k\widehat{\sigma}^{2}} \tag{22}$$

where $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\sigma}$ are the LSE of the respective parameters.

The null distribution of F and F^{*} are central F with $(k, n-k-1)$ degrees freedom. Large values indicate that at least one θ_j is not equal to zero. We can now compare the power of the T and T^{*} tests in $(16, 17)$ and F and F^{*} tests in $(21, 22)$. The simulation results for the power of these tests are provided in Table 3 for an assumed test-size 0.05 ($k = 3$). It is found that the tests based upon MMLE (T and F) are enormously more powerful than the respective tests based upon the LSE (T^{*} and F^{*}).

6 Robustness of estimators and tests

In this study, we use the following definition of robustness. An estimator is called robust if it is fully efficient under the assumed model and maintains high efficiency under the plausible alternatives to the assumed model , see Tiku and Akkaya [8]. For illustration purpose, first we take a true model as $GSH(0, \sigma; -\pi/4)$ and name it Model (1). It is a long-tailed model with kurtosis equal to 4.36. We use the following sample models to represent a large number of plausible alternatives to the true model.

Table 3: Power of the T^* and T tests ($\theta_0 = 0$, $\theta_2 = \theta_3 = 1$, $\sigma = 1$) and F^* and F tests ($\theta_0 = 0$, $\theta_2 = \theta_3 = 0$, $\sigma = 1$), $k = 3$.

	$h = -\pi/2$		$h = 0$		$h = \pi\sqrt{3}$		$h = \pi\sqrt{11}$	
θ_1	T^*	T	T^*	T	T^*	T	T^*	T
				$n = 50$				
0.00	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
0.10	0.193	0.195	0.193	0.195	0.194	0.198	0.196	0.202
0.20	0.414	0.435	0.407	0.419	0.406	0.449	0.406	0.513
0.30	0.662	0.690	0.659	0.670	0.650	0.704	0.654	0.780
0.40	0.868	0.889	0.868	0.872	0.864	0.897	0.869	0.939
0.50	0.956	0.969	0.956	0.962	0.960	0.971	0.959	0.989
0.60	0.988	0.992	0.990	0.991	0.994	0.995	0.993	0.999
				$n = 100$				
0.00	0.051	0.051	0.051	0.051	0.050	0.049	0.050	0.050
0.05	0.108	0.127	0.113	0.115	0.108	0.116	0.113	0.169
0.10	0.199	0.256	0.210	0.230	0.218	0.224	0.216	0.310
0.15	0.415	0.475	0.423	0.450	0.423	0.439	0.424	0.601
0.20	0.581	0.658	0.592	0.621	0.596	0.614	0.584	0.791
0.25	0.766	0.828	0.773	0.795	0.767	0.799	0.766	0.924
0.30	0.875	0.925	0.883	0.900	0.877	0.976	0.875	0.973
θ_1	F^*	F	F^*	F	F^*	F	F^*	F
				$n = 50$				
0.00	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
0.10	0.193	0.195	0.193	0.195	0.194	0.198	0.196	0.202
0.20	0.414	0.435	0.407	0.419	0.406	0.449	0.406	0.513
0.30	0.662	0.690	0.659	0.670	0.650	0.704	0.654	0.780
0.40	0.868	0.889	0.868	0.872	0.864	0.897	0.869	0.939
0.50	0.956	0.969	0.956	0.962	0.960	0.971	0.959	0.989
0.60	0.988	0.992	0.990	0.991	0.994	0.995	0.993	0.999
				$n = 100$				
0.00	0.051	0.051	0.051	0.051	0.050	0.049	0.050	0.050
0.05	0.108	0.127	0.113	0.115	0.108	0.116	0.113	0.169
0.10	0.199	0.256	0.210	0.230	0.218	0.224	0.216	0.310
0.15	0.415	0.475	0.423	0.450	0.423	0.439	0.424	0.601
0.20	0.581	0.658	0.592	0.621	0.596	0.614	0.584	0.791
0.25	0.766	0.828	0.773	0.795	0.767	0.799	0.766	0.924
0.30	0.875	0.925	0.883	0.900	0.877	0.976	0.875	0.973

(a) **Misspecification of the distribution**: Samples are generated from a $GSH(0, \sigma; -\pi/2)$ with kurtosis equal to 5.0.

(b) **Outliers model**: Outliers are generated by selecting 10% observations randomly and multiplying these by a factor 4.

(c) **Mixture model**: 90% observations are from the true model and 10% observations are drawn randomly from a $GSH(0, \sigma; -\pi\sqrt{2/3})$.

(d) **Contamination model**: 90% observations are taken from the true model and 10% are randomly taken from a Uniform distribution $U(-1/2, 1/2)$.

(e) **Contamination model**: 90% observations are taken from the true model and 10% observations are randomly taken from a standard normal distribution.

Next, we take $GSH(0, \sigma; \pi\sqrt{3})$ as a true model, call it Model (2), which represents a short-tailed distribution with kurtosis equal to 2.4. For alternatives, the following sample models are taken.

(f) **Misspecification of the distribution**: Samples are generated from $GSH(0, \sigma; \pi\sqrt{11})$ with the kurtosis equal to 2.0.

(g) **Lamda Family**: Samples are generated from Tukey's Lambda family of distributions with shape parameter taking a value 0.585 that corresponds to a kurtosis 2.0 (see, Tiku et al. [17]).

(h) **Inliers model**: Inliers are generated by reciprocating the smallest 5% and the largest 5% observations in the sample (for details, see Joiner and Rosenblatt [18]).

Table 4: Simulated values of means, variances and relative efficiency of the LS estimators; $\theta_0 = 0$, $\theta_j = 1$ $(1 \le j \le k)$, $\sigma = 1$, $k = 3$, $n = 50$.

	Mean		Variance		RE	Mean		Variance		RE
	MML	LS	MML	LS	(%)	MML	LS	MML	LS	(%)
	Model (1)					Model (a)				
θ_0	-0.002	-0.003	0.019	0.020	93	-0.002	-0.003	0.014	0.016	87
θ_1	0.997	0.997	0.021	0.022	93	0.998	0.997	0.015	0.018	87
θ_2	1.000	1.002	0.022	0.024	93	1.001	1.002	0.017	0.019	87
θ_3	1.002	1.002	0.022	0.024	92	1.002	1.002	0.017	0.019	86
σ	0.967	0.994	0.015	0.017	85	0.851	0.888	0.013	0.016	79
	Model(b)					Model(c)				
θ_0	-0.000	-0.003	0.025	0.049	52	0.002	0.001	0.019	0.021	89
θ_1	1.001	0.999	0.030	0.058	51	1.003	1.002	0.138	0.022	89
θ_2	1.003	1.001	0.032	0.062	52	1.006	1.007	0.021	0.023	88
θ_3	0.998	0.995	0.029	0.058	50	1.003	1.002	0.019	0.021	89
σ	1.312	1.538	0.052	0.146	35	0.929	0.962	0.016	0.019	85
	Model (d)					Model (e)				
θ_0	-0.002	-0.003	0.016	0.018	87	-0.002	-0.003	0.019	0.020	93
θ_1	0.997	0.995	0.018	0.021	84	0.996	0.996	0.021	0.023	93
θ_2	1.002	1.001	0.019	0.022	87	0.997	0.996	0.021	0.023	92
θ_3	0.997	0.998	0.018	0.020	87	0.998	0.999	0.022	0.024	93
σ	0.909	0.945	0.014	0.017	83	0.972	0.997	0.014	0.016	86
	Model (2)					Model (f)				
θ_0	-0.002	-0.003	0.018	0.021	87	-0.003	-0.004	0.041	0.062	67
θ_1	0.997	0.997	0.020	0.022	90	0.995	0.995	0.047	0.067	71
θ_2	1.003	1.001	0.022	0.024	89	1.004	1.001	0.051	0.073	71
θ_3	1.003	1.002	0.021	0.024	88	1.004	1.003	0.051	0.072	70
σ	0.963	0.998	0.007	0.008	91	1.643	1.730	0.013	0.018	76
	Model (g)					Model (h)				
θ_0	-0.001	-0.002	0.010	0.015	70	-0.001	-0.002	0.011	0.014	81
θ_1	0.997	0.997	0.012	0.016	74	1.002	1.002	0.008	0.012	63
θ_2	1.002	1.000	0.013	0.017	73	0.998	0.998	0.008	0.012	63
θ_3	1.002	1.002	0.013	0.017	72	1.003	1.002	0.007	0.011	63
σ	0.807	0.846	0.003	0.004	74	0.677	0.723	0.003	0.004	88

The simulated values of the means, variances for the MMLE and the LSE of the model parameters θ_j $(j = 0, \ldots, k)$ and σ under the alternative

models and the RE of the LSE are presented in Table 4. Clearly, the MMLE are not only less biased but also substantially more efficient and robust as compared to the LSE. The robustness of the MMLE to outliers is a consequence of the fact that the values of β_i $(i = 1, \ldots, n)$ are in symmetric umbrella ordering. Hence, the extreme residuals receive small weights and the influence is automatically depleted. Similarly, the robustness to inliers is the consequence of the inverted umbrella ordering of β_i $(i = 1, \ldots, n)$. As a result small residuals receive smaller weights. Note a disturbing feature of the LSE that their relative efficiencies decrease as sample size increases.
In order to show the robustness property of the T and F, tests we use the following definition of robustness formulated by Box [19] and Tiku et al. [20].
Criterion robustness, if the Type I error of a test is not substantially higher under plausible alternatives than that attained under an assumed model, the test is said to have criterion robustness.
Efficiency robustness, if the power of a test is the highest possible (or nearly so) under an assumed model but stays high for all plausible models, the test is said to have efficiency robustness.
The simulated values of the power of test for models (a) to (h) are presented in Table 5. It can be seen that the T and F tests are more powerful than the T^* and F^* tests, respectively, in all the models. Therefore, it can be said that tests based upon MMLE have criterion robustness as well as the efficiency robustness.

7 Applications

7.1 The demand for imports into the UK over a period 1973-2005

Barrow [21] presented a study on estimating the demand equation for imports into the UK over a period 1973-2005. The variables are defined as follows:

- Imports (variable M): imports of goods and services into the UK, at current prices, in £bn.

Table 5: Power of the T^* and T tests ($\theta_0 = 0$, $\theta_2 = \theta_3 = 1$, $\sigma = 1$) and F^* and F tests ($\theta_0 = 0$, $\theta_2 = \theta_3 = 0$, $\sigma = 1$), $k = 3$, $n = 50$.

	Model (a)		Model (b)		Model (c)		Model (d)	
θ_1	T^*	T	T^*	T	T^*	T	T^*	T
0.00	0.050	0.047	0.050	0.042	0.051	0.050	0.050	0.049
0.10	0.207	0.215	0.121	0.121	0.169	0.174	0.171	0.183
0.20	0.463	0.488	0.225	0.272	0.388	0.414	0.415	0.455
0.30	0.729	0.774	0.393	0.493	0.663	0.702	0.662	0.713
0.40	0.920	0.944	0.532	0.674	0.854	0.877	0.865	0.909
0.50	0.978	0.987	0.678	0.836	0.948	0.969	0.957	0.973
0.60	0.996	0.998	0.795	0.933	0.989	0.994	0.988	0.992
	Model (e)		Model (f)		Model (g)		Model (h)	
θ_1	T^*	T	T^*	T	T^*	T	T^*	T
0.00	0.053	0.051	0.047	0.043	0.052	0.045	0.051	0.041
0.10	0.178	0.188	0.118	0.123	0.211	0.258	0.238	0.292
0.20	0.400	0.428	0.203	0.238	0.489	0.592	0.591	0.741
0.30	0.663	0.694	0.329	0.421	0.768	0.868	0.875	0.959
0.40	0.850	0.876	0.474	0.590	0.933	0.977	0.983	0.999
0.50	0.950	0.964	0.627	0.736	0.989	1.000	0.998	1.000
0.60	0.987	0.990	0.756	0.866	0.999	1.000	0.999	1.000
	Model (a)		Model (b)		Model (c)		Model (d)	
θ_1	F^*	F	F^*	F	F^*	F	F^*	F
0.00	0.052	0.050	0.051	0.049	0.053	0.051	0.051	0.049
0.10	0.212	0.224	0.119	0.125	0.171	0.189	0.182	0.195
0.20	0.451	0.512	0.241	0.283	0.299	0.401	0.401	0.462
0.30	0.696	0.771	0.386	0.423	0.518	0.692	0.652	0.711
0.40	0.882	0.945	0.501	0.683	0.799	0.892	0.825	0.901
0.50	0.969	0.981	0.669	0.834	0.941	0.986	0.962	0.981
0.60	0.989	0.992	0.801	0.946	0.979	0.992	0.989	0.995
	Model(e)		Model (f)		Model (g)		Model (h)	
θ_1	F^*	F	F^*	F	F^*	F	F^*	F
0.00	0.051	0.051	0.049	0.046	0.051	0.046	0.052	0.043
0.10	0.179	0.185	0.112	0.125	0.202	0.261	0.217	0.293
0.20	0.386	0.412	0.201	0.242	0.426	0.589	0.588	0.745
0.30	0.596	0.708	0.312	0.412	0.694	0.860	0.796	0.884
0.40	0.761	0.892	0.473	0.598	0.882	0.986	0.898	0.982
0.50	0.899	0.965	0.612	0.742	0.981	0.999	0.992	0.999
0.60	0.988	0.991	0.761	0.902	0.999	1.000	0.999	1.000

- Income (GDP): UK gross domestic product (GDP) at factor cost, at current prices, in £bn.

- The GDP deflator (P_{GDP}): an index of the ratio of nominal to real GDP, 1985 = 100. This is an index of general price increases and may be used to transform nominal GDP to real GDP.

- The price of imports (P_M): the unit value index of imports, 1990 = 100.

- The price of competing products (P): the retail price index (RPI), 1985 = 100.

Before calculating the regression equation, the data must be transformed. This is because the expenditures on imports and GDP have not been adjusted for price changes (inflation). Also, we need to adjust the import prices that influence the demand for imports. People make their spending decisions by looking at the price of an imported good relative to prices of competing products. Hence, we divide the price of imports by the RPI to give the relative, or real, price of imports. In summary, the transformed variables are derived as follows:

- Real imports (M/P_M): this series is obtained by dividing the nominal series for imports by the unit value index (i.e. the import price index). The series gives imports at 1990 prices (in £bn). (Note that the nominal and real series are identical in 1990.)

- Real income (GDP/P_{GDP}): this is the nominal GDP series divided by the GDP deflator to give GDP at 1990 prices (in £bn).

- Real import prices (P_M/P): the unit value index is divided by the RPI to give this series. It is an index number series, with its value set to 100 in 1990.

The model to be estimated is therefore

$$\left(\frac{M}{P_M}\right)_i = \theta_0 + \theta_1 \left(\frac{GDP}{P_{GDP}}\right)_i + \theta_2 \left(\frac{P_M}{P}\right)_i + e_i$$

expressed in terms of the original variables. To simplify the notation, we rewrite this in terms of the transformed variables, as

$$m_i = \theta_0 + \theta_1\, gdp_i + \theta_2\, pm_i + e_i.$$

where $m_i = (M/P_M)_i$, $gdp_i = (GDP/P_{GDP})_i$ and $pm_i = (P_M/P)_i$.
The Q-Q plot of the estimated least square residuals

$$\widetilde{e}_i = m_i - \widetilde{\theta}_0 - \widetilde{\theta}_1\, gdp_i - \widetilde{\theta}_2\, pm_i,$$

reveals that the distribution of the residuals follow a symmetric long-tailed pattern, see Figure 1. Thus, a GSH model is proposed with the shape parameter $h = \pi\sqrt{11}$, the value obtained by performing the procedure given in Appendix 8. The value of coefficient of determination R^2 is found to be 0.9686. In Table 6, we present the MMLE and LSE along with their standard errors (SE). It is clear that the MMLE are superior to the corresponding LSE.

Table 6: Parameter estimates and their and standard errors.

Parameter	MML		LS	
	Estimate	SE	Estimate	SE
θ_0	-193.12	1.68	-217.43	76.097
θ_1	0.6170	0.0135	0.638	0.065
θ_2	0.1024	0.08	0.208	0.384
σ	13.26	0.996	14.19	1.24[a]

[a] Is the value of $(\widehat{\sigma}/\sqrt{2n})\sqrt{1+\lambda_4/2}$, $\lambda_4 = 2$.

7.2 The delivery time data

Montgomery et al. [22] considers the following application. A soft drink bottler is analyzing the vending machine service routes in his distribution system. He is interested in predicting the amount of time required by the route driver to service the vending machines in an outlet. This service activity includes stocking the machine with beverage products and minor maintenance or housekeeping. The industrial engineer responsible for the

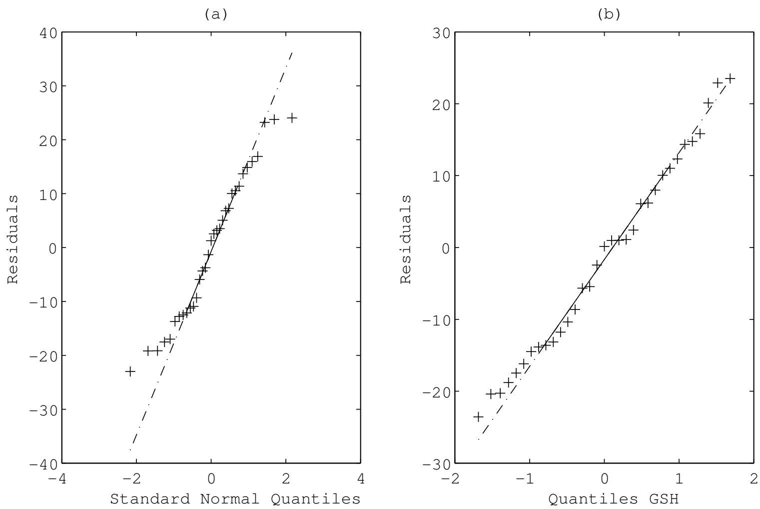

Figure 1: Q-Q plot of residuals: Application 7.1

study has suggested that the two most important variables affecting the delivery time (Y) are the number of cases of product stocked (X_1) and the distance walked by the route driver (X_2).

The following multiple linear regression model is appropriate to be used, see Mongomery et al. [22].

$$y_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + e_i, \quad (i = 1, \ldots, 25)$$

We determine the most plausible value of h as explained in Appendix 8. It comes out to be $-\pi\sqrt{3/4}$. See Figure 2 for corresponding Q-Q plots. The determination coefficient value R^2 comes out to be 0.9596. The results are presented in Table 7.

Table 7: Parameter estimates and their and standard errors.

parameter	MML		LS	
	Estimate	SE	Estimate	SE
θ_0	2.4060	0.5581	2.3412	1.0967
θ_1	1.7333	0.0828	1.6159	0.1707
θ_2	0.0124	0.0018	0.0144	0.0036
σ	4.6701	1.0609	3.259	1.0509[b]

[b] Is the value of $(\widehat{\sigma}/\sqrt{2n})\sqrt{1+\lambda_4/2}$, $\lambda_4 = 8.4$.

The MMLEs are clearly more efficient than the LSEs.

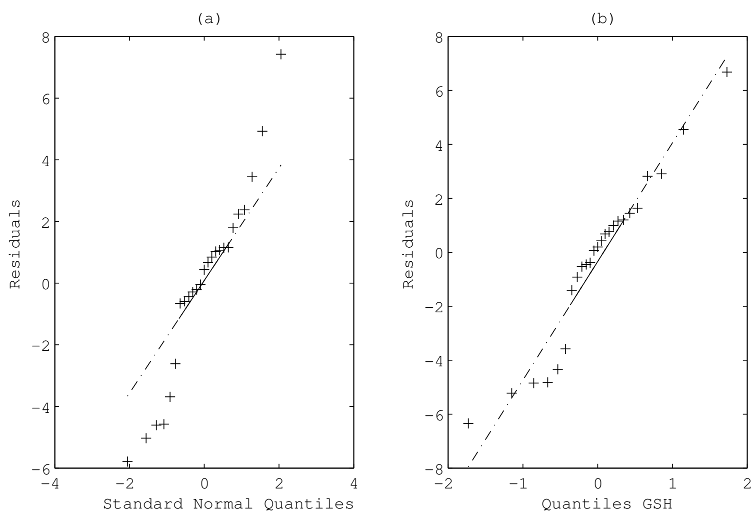

Figure 2: Q-Q plots of residuals: Application 7.2.

8 Conclusions

In the context of MLR model, with the assumption that random errors are having a distribution in the *GSH* family, we used the MML method for the estimation of the model parameters. This is theoretically and computationally simple to implement and it also provides explicit solutions. Simulation studies show that the MMLE are more efficient and robust than the commonly used LSE. Furthermore, in hypothesis testing the tests based upon MMLE are substantially more powerful as compared to the equivalent tests based upon LSE. In the applications presented here, it is observed that the linear least square regression residuals do not follow a normal distribution which is commonly assumed in practice. Rather, it suggests using distributions that can be selected from the flexible family of GSH distributions. Hence, the MML method can be easily adopted for the purpose of making statistical inferences in such cases.

References

[1] E. Pearson, “The analysis of variance in cases of nonnormal variation,” *Biometrika.*, vol. 22, no. 1/2, pp. 231–235, 1986. https://doi.org/10.2307/2333631 46

[2] P. Huber, *Robust Statistics*, 2nd ed. Jonh Wiley: New York, 1981. 46

[3] J. Tukey, *A survey of sampling from contaminated distributions.* Stanford University Press, Stanford: Contributions to Probability and Statistics, 1960. 46

[4] V. D. Barnett, “Order statistic estimators of the location of the cauchy distribution,” *Journal of American Statistical Association.*, vol. 61, no. 316, pp. 1205–1218, 1966. https://doi.org/10.2307/2283210 46

[5] D. C. Vaughan, “On the tiku-suresh method of estimation,” *Communications in Statistics - Theory and Methods*, vol. 21, no. 2, pp. 451–469, 1992. https://doi.org/10.1080/03610929208830788 47, 49

[6] M. L. Tiku and R. P. Suresh, “A new method of estimation for location and scale parameters,” *Journal of Statistical Planning and Inference.*, vol. 30, no. 2, pp. 281–292, 1992. https://doi.org/10.1016/0378-3758(92)90088-A 47

[7] D. C. Vaughan and M. L. Tiku, “Estimation and hypothesis testing for a nonnormal bivariate distribution with applications,” *Mathematical and Computer Modelling*, vol. 32, no. 1/2, pp. 53–67, 2000. https://doi.org/10.1016/S0895-7177(00)00119-9 47, 50, 51, 55

[8] M. L. Tiku and A. D. Akkaya, *Robust Estimation and Hypothesis Testing*, 2nd ed. New York: New Age, 2004. 47, 49, 50, 57

[9] M. Q. Islam and M. L. Tiku, "Multiple linear regression model under non-normality," *Communications in Statistics - Theory and Methods.*, vol. 33, no. 10, pp. 2443–2467, 2004. https://doi.org/10.1081/STA-200031519 47

[10] D. C. Vaughan, "The generalized secant hyperbolic distribution and its properties," *Communications in Statistics - Theory and Methods.*, vol. 31, no. 2, pp. 219–238, 2002. https://doi.org/10.1081/STA-120002647 47, 50

[11] Y. E. Yilmaz and A. D. Akkaya, "Analysis of variance and linear contrasts in experimental design with generalized secant hyperbolic distribution," *Journal of Computational and Applied Mathematics.*, vol. 216, no. 2, pp. 545–553, 2008. https://doi.org/10.1016/j.cam.2007.06.001 47

[12] V. D. Barnett, "Evaluation of the maximum likelihood estimator when the likelihood equation has multiple roots," *Biometrika.*, vol. 53, no. 1/2, pp. 151–165, 1996. https://doi.org/10.2307/2334061 49

[13] S. Puthenpura and N. K. Sinha, "Modified maximum likelihood method for the robust estimation of system parametrs from very noisy data," *Automatica.*, vol. 22, pp. 231–235, 1986. https://doi.org/10.1016/0005-1098(86)90085-3 49

[14] B. Senoglu and M. L. Tiku, "Analysis of variance in experimental design with non-normal error distributions," *Communications in Statistics - Theory and Methods.*, vol. 30, pp. 1335–1352, 2001. https://doi.org/10.1081/STA-100104748 50

[15] M. Q. Islam, M. L. Tiku, and F. Yildirim, "Nonnormal regression. I. skew distributions," *Communications in Statistics - Theory and Methods.*, vol. 30, no. 6, pp. 993–1020., 2001. https://doi.org/10.1081/STA-100104347 56

[16] M. L. Tiku, W. K. Wong, and G. Bian, "Estimating parameters in autoregressive models in non-normal situations: symmetric innovations," *Communications in Statistics - Theory and Methods.*, vol. 2, no. 28, pp. 315–341, 1999. https://doi.org/10.1080/03610929908832300 56

[17] M. L. Tiku, M. Q. Islam, and A. S. Selcuk, "Nonnormal regression. II. symmetric distributions," *Communications in Statistics - Theory and Methods.*, vol. 30, no. 6, pp. 1021–1045, 2001. https://doi.org/10.1081/STA-100104348 59

[18] B. L. Joiner and J. R. Rosenblatt, "Some properties of the range in samples from tukey's symmetric lambda distributions," *Journal of American Statistical Association*, vol. 66, pp. 394–399, 1971. https://doi.org/10.2307/2283943 59

[19] G. Box, "Non-normality and test of variances," *Biometrika.*, vol. 40, pp. 336–346, 1953. https://doi.org/10.2307/2333350 61

[20] M. L. Tiku, W. Y. Tan, and N. Balakrishnan, *Robust Inference.* New York: Marcel Dekker, 1986. 61

[21] M. Barrow, *Statistics for Economics, Accounting and Business Studies*, 5th ed. Pearson Education, 2009. 61

[22] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*, 5th ed. John wiley & Sons, 2015. 64, 65

Appendices

Determination of the shape parameter h

For a chosen value of h calculate the values of $\widehat{\theta}_0$, $\widehat{\boldsymbol{\theta}}$ and $\widehat{\sigma}$ from the equations(11), (12) and (13). Then get the value of

$$\frac{1}{n}\ln L = \ln\left(\frac{c_1}{\widehat{\sigma}}\right) + \frac{1}{n}c_2\sum_{i=1}^{n}\widehat{z}_i - \frac{1}{n}\sum_{i=1}^{n}\ln\left\{\exp\left(2c_2\widehat{z}_i\right) + 2a\exp\left(c_2\widehat{z}_i\right) + 1\right\}, \quad (23)$$

where

$$\widehat{z}_i = \left(y_i - \widehat{\theta}_0 - \sum_{j=1}^{k}\widehat{\theta}_j x_{ij}\right) \Big/ \widehat{\sigma} \qquad (1 \le i \le n).$$

Repeat this procedure for a series of values of h. The value that maximizes $\frac{1}{n}\ln L$ is taken as the most plausible value for the shape parameter.