ISSN:1306-3111
e-Journal of New World Sciences Academy
2011, Volume: 6, Number: 3, Article Number: 3A0041

PHYSICAL SCIENCES
Received: June 2011
Accepted: July 2011
Series : 3A
ISSN : 1308-7304
© 2010 www.newwsa.com

Pelin Kasap
Birdal Şenoğlu
Olcay Arslan
Şükrü Acıtaş
Ankara University
pkasap@science.ankara.edu.tr
Ankara-Turkey

ESTIMATING THE LOCATION AND SCALE PARAMETERS OF THE GT DISTRIBUTION

ABSTRACT

Arslan and Genç [1] obtained the M estimators of the location and scale parameters of the Generalized t (GT) distribution. In this study, different than their studies, we obtain the modified maximum likelihood (MML) estimators of the model parameters. We compare the efficiencies of the M and MML estimators via Monte Carlo Simulation study for some pre-determined p and q values.

Keywords: GT Distribution, Maximum Likelihood, Modified Maximum Likelihood, M Estimators, Efficiency

GT DAĞILIMININ KONUM VE ÖLÇEK PARAMETRELERININ TAHMINI

ÖZET

Arslan and Genç [1] Genelleştirilmiş t (GT) dağılımının konum ve ölçek parametrelerinin M tahmin edicilerini elde etmişlerdir. Bu çalışmada, onların çalışmalarından farklı olarak, model parametrelerinin Uyarlanmış En Çok Olabilirlik (MML) tahmin edicileri elde edilmiştir. M ve MML tahmin edicilerinin etkinlikleri, p ve q şekil parametrelerinin bazı değerleri için Monte Carlo simülasyonu yardımı ile karşılaştırılmıştır.

Anahtar Kelimeler: GT Dağılımı, En Çok Olabilirlik, Uyarlanmış En Çok Olabilirlik, M Tahmin Edicileri, Etkinlik

1. INTRODUCTION (GİRİŞ)

GT distribution is widely used in the areas of economy and finance. It was first introduced by McDonald and Newey [6]. They used it for estimating the regression parameters. Then, it has been followed up by many other authors. Theodossiou [10] developed a skewed extension of GT distribution and derived the some mathematical properties of the distribution. Arslan and Genç [1] obtained maximum likelihood estimators for the location and the scale parameters of a GT distribution. Bali and Peng [2] used GT distribution for estimating the GARCH-in-mean models with daily data. Butler et al. [3] applied a partially adaptive technique to estimate the parameters of regression models, see also Kantar et al. [5] in which GT is used within partially adaptive estimation.

In this study, we obtain the estimators of the location and scale parameters by using the MML methodology introduced by Tiku [11 and 12]. Then we compare the efficiencies of these estimators with the M estimators obtained from the likelihood equations studied by Arslan and Genç [1] via Monte Carlo simulation study for some selected *p*, *q* and *n* values.

In spite of the fact that M and MML estimators are both asymptotically equivalent to the ML estimators, we do not known their behaviour for small and moderate samples. This study will also provide an answer for this problem.

Traditional estimators of the model parameters are the least squares (LS) estimators. Therefore, we include them into the simulation study for making this study precise.

- **GT Distribution:**

The probability density function (PDF) of GT distribution is given by

$$f\left(x;\mu,\sigma,p,q\right)=\frac{p}{2\sigma q^{1/p}B\left(\frac{1}{p},q\right)}\left\{1+\frac{\left|x-\mu\right|^{p}}{q\sigma^{p}}\right\}^{-\left(q+\frac{1}{p}\right)} \tag{1.1}$$

where $x,\ \mu\in R$, $\sigma,\ p,\ q\in R^{+}$. Here, μ is a location parameter, σ is a scale parameter, p and q are the shape parameters and $B(.)$ is the beta function. The corresponding cumulative distribution function (CDF) is:

$$F\left(x\right)=\begin{cases}\frac{1}{2}\left[1+I_{1-\left\{1+\left(x/\sigma q^{1/p}\right)^{p}\right\}^{-1}}\left(\frac{1}{p},q\right)\right] & , \quad x\geq 0\\ \frac{1}{2}\left[1-I_{1-\left\{1+\left(-x/\sigma q^{1/p}\right)^{p}\right\}^{-1}}\left(\frac{1}{p},q\right)\right] & , \quad x<0\end{cases}. \tag{1.2}$$

See Nadarajah and Kotz [7], Nadarajah [8].
In (1.2), I is incomplete beta function and is defined by

$$I_{y}\left(a,b\right)=\frac{1}{B\left(a,b\right)}\int_{0}^{y}w^{a-1}\left(1-w\right)^{b-1}dw. \tag{1.3}$$

The mean and the variance of GT distribution are

$$E(X) = 0, \quad Var(X) = \frac{q^{2/p}\Gamma(3/p)\Gamma(q-2/p)}{\Gamma(1/p)\Gamma(q)} \quad \text{if } pq > 2, \tag{1.4}$$

respectively, see McDonald and Newey [6].

The skewness $(\sqrt{\beta_1})$ and the kurtosis (β_2) values of the GT distribution are obtained by using the following formulas

$$\sqrt{\beta_1} = \mu_3/(\mu_2)^{3/2} \quad , \quad \beta_2 = \mu_4/\mu_2^2 \tag{1.5}$$

where

$$\mu_n = \left[(-1)^n + 1\right] p^n B\left(\frac{n+1}{p}, q - \frac{n}{p}\right)\left[2B\left(\frac{1}{p}, q\right)\right]^{-n-1}. \tag{1.6}$$

See Hueng and Brooks [4].

For better understanding the shape of the GT distribution see the following table, which is taken from Hueng and Brooks [4], showing the kurtosis values. See also Figure 1 given below.

Table 1. β_2 values of the GT distribution for some p and q values
(Tablo 1. GT dağılımının bazı p ve q değerleri için β_2 değerleri)

(p, q)	$(0.5, 0.5)$	$(0.5, 1)$	$(0.5, 10)$	$(0.5, 50)$	$(0.5, 100)$
β_2	–	–	635.040	35.835	29.798
(p, q)	$(1, 0.5)$	$(1, 1)$	$(1, 10)$	$(1, 50)$	$(1, 100)$
β_2	–	–	10.286	6.527	6.251
(p, q)	$(10, 0.5)$	$(10, 1)$	$(10, 10)$	$(10, 50)$	$(10, 100)$
β_2	3.550	2.070	1.892	1.886	1.885
(p, q)	$(50, 0.5)$	$(50, 1)$	$(50, 10)$	$(50, 50)$	$(50, 100)$
β_2	1.821	1.809	1.805	1.804	1.804
(p, q)	$(100, 0.5)$	$(100, 1)$	$(100, 10)$	$(100, 50)$	$(100, 100)$
β_2	1.805	1.802	1.801	1.801	1.801

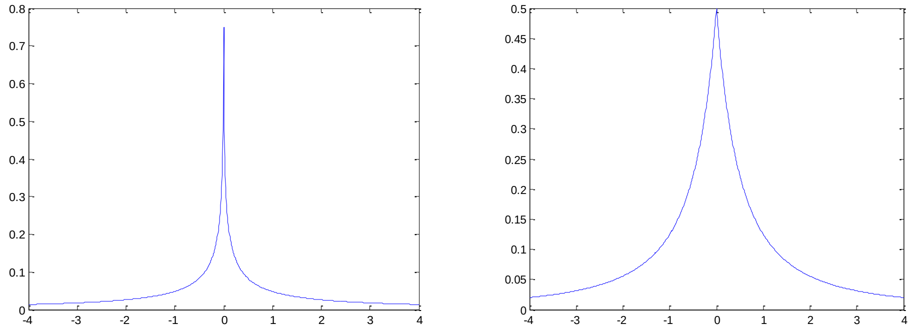

Figure 1(a). $p = q = 0.5$ Figure 1(b). $p = q = 1$

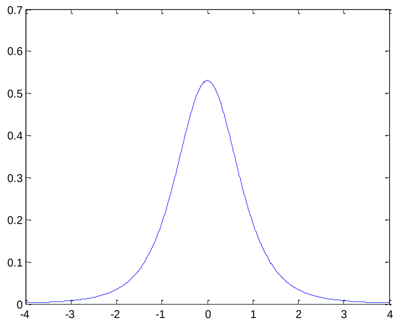

Figure 1(c). $p = q = 2$

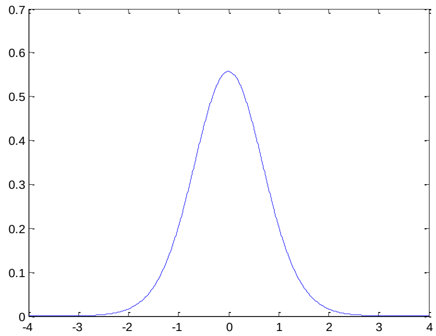

Figure 1(d). $p = 2, q = 10$

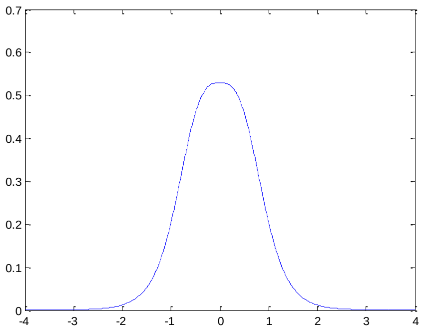

Figure 1(e). $p = 3, q = 2$

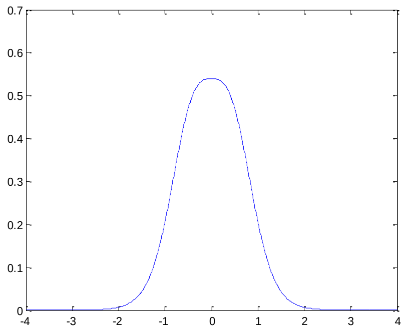

Figure 1(f). $p = q = 3$

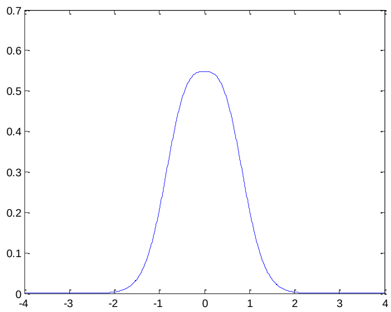

Figure 1(g). $p = 3, q = 5$

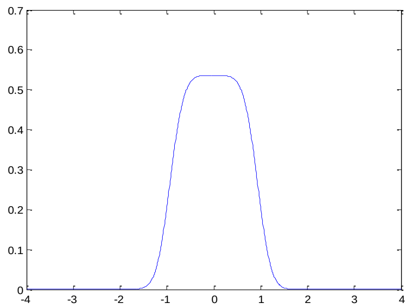

Figure 1(h). $p = q = 5$

It should be noted that skewness values are not given in Table 1 since GT distribution is symmetric. Therefore its skewness values are all 0.

It is clear from Table 1 and Figure 1 that GT distribution has sometimes thick tails and sometimes thin tails according to the values of p and q. If the shape parameters p and q are small then GT has thick tails, otherwise it has thin tails. GT distribution covers various other distributions, such as Student's t and Normal. When $p=2$ GT distribution reduces to the well known Student's t distribution with $\upsilon = 2q$ degrees of freedom. It is clear that when $p = 2$ and $q = \infty$ it becomes the normal distribution. See Arslan and Genç [1] for more detailed information.

2. RESEARCH SIGNIFICANCE (ÇALIŞMANIN ÖNEMİ)

In this study, we have considered parameter estimation for the GT distribution. There are many methods used to estimate the model parameters in a given distribution. The traditional method for parameter estimation is the LS method. Estimators obtained by the LS method are the most efficient under normal distribution. But, efficiencies of the LS estimators decrease when there are departures from normality [9]. Other estimators are the M estimators obtained from the likelihood equations for the location and the scale parameters. However, these estimators cannot be explicitly expressed so that some numerical algorithms have to be used to compute them. Another widely used method is the MML method introduced by Tiku [11,12]. Unlike the M estimators, the MML estimators are explicitly expressed in terms of the observations.

In this study, LS, M and MML estimators are compared in terms of bias and efficiency for the finite sample size.

3. METHOD (METOT)

Given a random sample $x_1, x_2, \ldots, x_n$ from (1.1), the log-likelihood function is

$$\ln L(\mu, \sigma, p, q; \underline{x}) = n \ln p + n \ln \Gamma\left(q + \frac{1}{p}\right) - n \ln 2 - n \ln \sigma - \frac{n}{p} \ln q - n \ln \Gamma\left(\frac{1}{p}\right)$$
$$-n\Gamma(q) - \left(q + \frac{1}{p}\right) \sum_{i=1}^{n} \ln\left\{1 + \frac{|x_i - \mu|^p}{q\sigma^p}\right\}. \qquad (3.1)$$

We assume that the shape parameters p and q are known. Therefore, taking the partial derivatives of (3.1) with respect to μ and σ, the likelihood equations are

$$\frac{\partial \ln L(\mu, \sigma, p, q; \underline{x})}{\partial \mu} = \frac{pq+1}{q} \sum_{i=1}^{n} \operatorname{sgn}(z_i)\, g(z_i) = 0 \qquad \text{if } p > 1 \qquad (3.2)$$

$$\frac{\partial \ln L(\mu, \sigma, p, q; \underline{x})}{\partial \sigma} = -\frac{n}{\sigma} + \frac{pq+1}{q\sigma} \sum_{i=1}^{n} z_i g(z_i) = 0 \qquad (3.3)$$

where

$$z_i = \frac{x_i - \mu}{\sigma}, \quad g(z) = \frac{|z|^{p-1}}{\left\{1 + \frac{1}{q}|z|^p\right\}}. \qquad (3.4)$$

It is obvious that these equations cannot be solved analytically because of the awkward function $g(z)$. One can use any numerical method to solve these equations to find exact or approximate solutions. However, in this study we used the MML methodology which provides explicit estimators of the model parameters to find the approximate solutions of these equations.

First, we express equations (3.2)-(3.3) in terms of order statistics. Since complete sums are invariant to ordering that is

$$\sum_{i=1}^{n} g(z_i) = \sum_{i=1}^{n} g(z_{(i)}), \qquad (3.5)$$

the likelihood equations equivalently written as

$$\frac{\partial \ln L(\mu, \sigma, p, q; \underline{x})}{\partial \mu} = \frac{pq+1}{q} \sum_{i=1}^{n} \operatorname{sgn}(z_{(i)})\, g(z_{(i)}) = 0 \qquad \text{if } p > 1 \qquad (3.6)$$

$$\frac{\partial \ln L(\mu, \sigma, p, q; \underline{x})}{\partial \sigma} = -\frac{n}{\sigma} + \frac{pq+1}{q\sigma} \sum_{i=1}^{n} z_{(i)} g(z_{(i)}) = 0. \qquad (3.7)$$

Further, since $g(z)$ is almost linear in small intervals around $t_{(i)} = E\left(z_{(i)}\right)\ (1 \le i \le n)$, we linearize the function $g(z)$ by using the first two terms of Taylor series expansion, i.e.,

$$g\left(z_{(i)}\right) \cong \alpha_i + \beta_i z_{(i)} \tag{3.8}$$

where

$$\alpha_i = \frac{|t|^{p-1}}{\left\{1 + \frac{1}{q}|t|^p\right\}} - t \cdot \frac{(p-1)|t|^{p-2}\left\{1 + \frac{1}{q}|t|^p\right\} - \frac{p}{q}|t|^{2p-2}}{\left\{1 + \frac{1}{q}|t|^p\right\}^2} \tag{3.9}$$

and

$$\beta_i = \frac{(p-1)|t|^{p-2}\left\{1 + \frac{1}{q}|t|^p\right\} - \frac{p}{q}|t|^{2p-2}}{\left\{1 + \frac{1}{q}|t|^p\right\}^2}. \tag{3.10}$$

The approximate values of $t_{(i)}$s are the solutions of the following equation

$$F\left(t_{(i)}\right) = \frac{i}{n+1}\ (1 \le i \le n). \tag{3.11}$$

Then $t_{(i)}$ values are obtained as

$$t_{(i)} = \begin{cases} \left(\frac{y_1}{1-y_1}\right)^{1/p} \sigma q^{1/p},\ y_1 = 1 - \left\{1 + \left(\left(i/n+1\right)/\sigma q^{1/p}\right)^p\right\}^{-1} & \text{, for } x \ge 0 \\ -\left(\frac{y_2}{1-y_2}\right)^{1/p} \sigma q^{1/p},\ y_2 = 1 - \left\{1 + \left(-\left(i/n+1\right)/\sigma q^{1/p}\right)^p\right\}^{-1} & \text{, for } x < 0 \end{cases}$$
(3.12)

By inserting (3.8) into (3.6)-(3.7), we obtain the following MML estimators

$$\hat{\mu}_{MML} = \bar{x}_{(.)} \quad \text{and} \quad \hat{\sigma}_{MML} = \frac{B + \sqrt{B^2 + 4nC}}{2\sqrt{n(n-1)}} \tag{3.13}$$

where

$$\bar{x}_{(.)} = \sum_{i=1}^{n} \beta_i x_{(i)} \Big/ m,\quad m = \sum_{i=1}^{n} \beta_i$$

$$B = \frac{(pq+1)}{q}\sum_{i=1}^{n} \alpha_i \left(x_{(i)} - \bar{x}_{(.)}\right),\quad C = \frac{(pq+1)}{q}\sum_{i=1}^{n} \beta_i \left(x_{(i)} - \bar{x}_{(.)}\right)^2. \tag{3.14}$$

It should be noted that original denominator of $\hat{\sigma}_{MML}$ is replaced by $2\sqrt{n(n-1)}$ for bias correction.

4. FINDINGS (BULGULAR)

In this simulation study, we compare the LS, M and MML estimators as mentioned earlier. The LS estimators for the parameters μ and σ are

$$\hat{\mu}_{LS} = \sum_{i=1}^{n} x_i \Big/ n \quad \text{and} \quad \hat{\sigma}_{LS} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}. \tag{4.1}$$

Note that in simulation study we make a bias correction for $\hat{\sigma}_{LS}$ using the variance of GT distribution.

Corresponding M estimators for μ and σ obtained from the GT distribution are

$$\hat{\mu}_M = \sum_{i=1}^{n} w_i x_i \Big/ \sum_{i=1}^{n} w_i \quad \text{and} \quad \hat{\sigma}_M = \sqrt{\frac{1}{n}\sum_{i=1}^{n} w_i \left(x_i - \hat{\mu}_M\right)^2} \tag{4.2}$$

where

$$w_i = \frac{(pq+1)\left(|x_i - \hat{\mu}_M|/\hat{\sigma}_M\right)^{p-2}}{q + \left(|x_i - \hat{\mu}_M|/\hat{\sigma}_M\right)^{p}}. \tag{4.3}$$

These estimators can be computed using iteratively reweighting algorithm. For the details of these estimators and the algorithm see Arslan and Genç [1].

In the simulation study, we generate data from GT distribution with $\mu = 0$ and $\sigma = 1$ without loss of generality. The sample sizes are taken as $n = 20, 50 \text{ and } 100$. The results for different values of the shape parameters p and q are given in Table 2. In these Table, the mean, the variance and the RE values

$$RE_M = \frac{Var(\Box_M)}{Var(\Box_{LS})} \times 100 \quad \text{and} \quad RE_{MML} = \frac{Var(\Box_{MML})}{Var(\Box_{LS})} \times 100 \tag{4.4}$$

are given.

Table 2. Simulated means, variances and RE values for some *p* and *q* values.
(Tablo 2. Bazı *p* ve *q* değerleri için simülasyonla elde edilen ortalama, varyans ve RE değerleri)

	(p, q)	$(3, 2)$			$(3, 3)$		
	n	20	50	100	20	50	100
mean	$\hat{\mu}_{LS}$	-0,0010	0,0033	0,0007	0,0000	0,0031	0,0016
	$\hat{\mu}_{MML}$	-0,0010	0,0036	0,0006	0,0005	0,0027	0,0014
	$\hat{\mu}_{M}$	-0,0020	0,0036	0,0006	0,0016	0,0026	0,0015
$n \times$ var	$\hat{\mu}_{LS}$	0,5460	0,5163	0,5111	0,4593	0,4358	0,4530
	$\hat{\mu}_{MML}$	0,5389	0,4971	0,4849	0,4577	0,4265	0,4352
	$\hat{\mu}_{M}$	0,5261	0,4945	0,4852	0,4545	0,4232	0,4348
RE_{MML}		99	96	95	100	98	96
RE_{M}		96	96	95	99	97	96
mean	$\hat{\sigma}_{LS}$	0,9817	0,9968	0,9977	0,9751	0,9954	0,9953
	$\hat{\sigma}_{MML}$	1,0298	1,0181	1,0076	0,9985	1,0093	1,0027
	$\hat{\sigma}_{M}$	0,9695	0,9911	0,9949	0,9597	0,9888	0,9916
$n \times$ var	$\hat{\sigma}_{LS}$	0,6860	0,6591	0,7540	0,5262	0,4884	0,5582
	$\hat{\sigma}_{MML}$	0,7193	0,5886	0,5850	0,5521	0,4920	0,5378
	$\hat{\sigma}_{M}$	0,5936	0,5440	0,5651	0,5057	0,4663	0,5159
RE_{MML}		105	89	78	105	101	96
RE_{M}		87	83	75	96	95	92

Table 2. (Continue)
(Tablo 2. (Devam))

	(p, q)	(2, 2)			(2, 3)		
	n	20	50	100	20	50	100
mean	$\hat{\mu}_{LS}$	0,0118	0,0075	0,0005	0,0007	0,0022	0,0020
	$\hat{\mu}_{MML}$	0,0053	0,0055	-0,0006	0,0037	0,0016	0,0013
	$\hat{\mu}_{M}$	0,0044	0,0053	-0,0006	0,0045	0,0015	0,0013
$n \times$ var	$\hat{\mu}_{LS}$	0,9793	1,0263	0,9693	0,7855	0,7676	0,7396
	$\hat{\mu}_{MML}$	0,6903	0,7159	0,6525	0,6855	0,6664	0,6444
	$\hat{\mu}_{M}$	0,6828	0,7109	0,6511	0,6801	0,6655	0,6450
RE_{MML}		70	70	67	87	87	87
RE_{M}		70	69	67	87	87	87
mean	$\hat{\sigma}_{LS}$	1,3619	1,3863	1,3952	1,2077	1,2156	1,2189
	$\hat{\sigma}_{MML}$	1,0596	1,0111	0,9956	1,0476	1,0195	1,0081
	$\hat{\sigma}_{M}$	0,9840	0,9962	0,9968	0,9812	0,9939	0,9966
$n \times$ var	$\hat{\sigma}_{LS}$	3,0810	4,1863	4,3135	1,3872	1,5526	1,6381
	$\hat{\sigma}_{MML}$	1,1011	1,2164	1,1361	0,8890	0,8222	0,7767
	$\hat{\sigma}_{M}$	0,8843	0,9159	0,8990	0,7613	0,7725	0,7576
RE_{MML}		36	29	26	64	53	47
RE_{M}		29	22	21	55	50	46

Table 2. (Continue)
(Tablo 2. (Devam))

	(p, q)	(3, 5)			(2, 10)		
	n	20	50	100	20	50	100
mean	$\hat{\mu}_{LS}$	0,0022	-0,0022	0,0009	0,0061	-0,0026	0,0002
	$\hat{\mu}_{MML}$	0,0014	-0,0011	0,0011	0,0055	-0,0026	-0,0001
	$\hat{\mu}_{M}$	0,0021	-0,0011	0,001	0,0052	-0,0026	-0,0001
$n \times$ var	$\hat{\mu}_{LS}$	0,4402	0,4164	0,4092	0,5459	0,5621	0,5306
	$\hat{\mu}_{MML}$	0,4309	0,3977	0,3801	0,5421	0,5518	0,5173
	$\hat{\mu}_{M}$	0,4317	0,3997	0,3798	0,5434	0,5501	0,5161
RE_{MML}		98	96	93	99	98	98
RE_{M}		98	96	93	100	98	97
mean	$\hat{\sigma}_{LS}$	0,9951	0,9947	0,996	1,0415	1,0481	1,0516
	$\hat{\sigma}_{MML}$	0,9999	0,9999	0,9993	1,0064	1,0033	1,0027
	$\hat{\sigma}_{M}$	0,9709	0,9859	0,9916	0,9707	0,9872	0,9941
$n \times$ var	$\hat{\sigma}_{LS}$	0,4475	0,3907	0,4085	0,6543	0,6579	0,6697
	$\hat{\sigma}_{MML}$	0,4503	0,3964	0,4064	0,6062	0,5887	0,5962
	$\hat{\sigma}_{M}$	0,4259	0,3859	0,4028	0,5636	0,5684	0,5855
RE_{MML}		101	101	99	93	89	89
RE_{M}		95	99	99	86	86	87

5. RESULTS AND SUGGESTIONS (SONUÇLAR VE ÖNERİLER)

As can be seen from Table 2, the bias for LS, M and MML estimators are negligible and when the sample size increases the estimators becomes very close to the exact parameter values 0 and 1, respectively. For all sample sizes, M and MML estimators are more efficient than the LS estimators. Further, the efficiency of the M and MML increases with *n*. Overall the simulation results show that the M and MML estimators are remarkably efficient compared to the conventional LS estimators. The M and MML estimators can be both used to estimate the location and the scale of the GT distribution, but since the MML estimates are easy to compute one can prefer to use MML estimates over the M estimates.

REFERENCES (KAYNAKLAR)

1. Arslan, O. and Genc, A.I., (2003). Robust Location and Scale Estimation Based on the Univariate Generalized t Distribution, Communications in Statistics: Theory and Methods, Vol. 32 No.8 pp.1505-1525.
2. Bali, T.G. and Peng, L., (2006). Is there a risk-return trade-off? Evidence from high-frequency data, Journal of Applied Econometrics, 21:1169-1198.
3. Butler, R.J., McDonald, J.B., Nelson, R.D., and White, S.B. (1990). Robust and partially adaptive estimation of regression models. The *Review of Economics and Statistics.72:* 321-327.
4. Hueng, C.J. and Brooks, R., (2000). Forecasting asymmetries in stock returns: Evidence from higher moments and conditional densities. Department of Economics, Finance and Legal Studies Working Paper Series Index: www.cba.ua.edu. Culverhouse College of Commerce and Business Administration The University of Alabama.
5. Kantar, Y.M., Usta, I., and Acıtaş, Ş. (2010). A Monte Carlo simulation study on partially adaptive estimators of linear regression models. Journal of Applied Statistics, DOI: 10.1080/02664763.2010.516389.
6. McDonald, J.B. and Newey, W.K., (1988). Partially adaptive estimation of regression models via the generalized t distribution. *Econometric Theory* 4: 428–457.
7. Nadarajah, S. and Kotz, S., (2007). Skew models II. Acta Appl. Math., 98, 29-46.
8. Nadarajah, S., (2008). On the generalized t(GT) distribution. Statistics, 42, 467-473.
9. Şenoğlu, B., (2005). Robust 2^k factorial design with Weibull error distributions, Journal of Applied Statistics, Vol. 32, No. 10,pp. 1051-1066.
10. Theodossiou, P., (1998). Financial data and the skewed generalized *t* distribution. Management Science Part I, 12:1650-1661.
11. Tiku, M.L., (1967). Estimating the mean and standard deviation from censored normal samples, *Biometrika*, 54, pp. 155–165.
12. Tiku, M.L., (1968). Estimating the parameters of log-normal distribution from censored samples. J. Amer. Stat Assoc., 63, 134–140.