On the determination of the number of outliers in a Geometric sample

Mathachan Pathiyil and E.S Jeevanand
Department of Mathematics and Statistics,
Union Christian College,Aluva 683 102,India.
E-Mail**:** *radhajeevanand@yahoo.co.in*,
mathachanpathiyil@yahoo.co.in,
Fax: 91 – 0484 - 2607534.

Abstract: *In this paper we suggest procedures for the determination of the number of outliers present in a sample taken from geometric distribution when the data is in the form of a frequency distribution. We also compare our procedure with identification of outlier procedure based on the posterior density proposed by Kale and Kale and that based on the method of least square by Wu.*

Keywords: Geometric distribution, Outliers, Survival function, Bayes method.

AMS Classification: 62C, 62E, 62N.

1. Introduction

In an experimental situation, many a times an experimenter comes across some of the observations which are far removed from the main body of the data and hence are outliers. We shall define an outlier as a set of data to be an observation (or subset of observations) that appears to be inconsistent with the remainder of the data. The problem of outliers is of considerable importance in almost all experimental fields and has received continued attention in statistical literature. The prime consideration of the scientists was to devise rules for detection of such outlying observations.

A familiar topic in the vast amount of literature available on outliers is the problem of identification or accommodation pertaining to specific continuous probability models like the Normal, Gamma, Weibull, Exponential, Pareto etc.

Anscombe [1], Grubbs [6], Stigler [19,20], Barnett [2], Kale [16], Hawkins [7], Barnett and Lewis [3], Joshi [14], Pettit [18], Gather and Kale [5], Jeevanand and Nair [8, 9, 10, 11, 12,13], Wu [21] have considered the problem of accommodation or identification of outliers for the continuous probability distributions.

Reliability analysis of devices or systems through failure time data when time is treated as discrete is an emerging area of research and the geometric distribution due to its lack of memory property is widely used to model such systems. It is a good model for studies on the distribution of the demand (supply) of commodities in the field of Economics, on the analysis of offer distribution of wage workers by labour economists dealing with job search models, on the entropy distribution in decoding of information in information theory and cryptography, on the modeling of the frequency distribution of natal dispersal distances counted as the number of home range diameters moved by vertebrate animals, in the construction of statistical control limits for attributes in process control of near zero-defects processes based on the cumulative count of conforming items between two non conforming ones etc.. Further, the model belongs to the class of long tailed distributions and as such, the occurrence of extreme observations is quite common and their identification as outliers or not becomes important. It seems that the problem of identification of the outliers in the geometric set up is not much discussed in the available literature and the present work is an attempt to fill this gap.

The remaining part of the paper is organized as follows. In Section 2,we derive a procedure to determine the number of outliers present in geometric data when it is in the frequency form. The procedure can also be used even if the data is not in a grouped form. Kale and Kale [19] proposed a procedure to identify the number of outliers present in an exponential sample based on posterior density and we extend this result to the discrete case in Section 3. Section 4 is devoted to the extension of the identification of the outliers using the least square method based on the empirical distribution function proposed by Wu [24]. Finally in Section 5, the use of the procedures discussed so far is demonstrated with the help of simulated and real data sets.

2. Detection of the number of outliers by the least square method based on the survival function

In this section, we derive a method to identify the upper outliers based on the survival function when the data is in the form a frequency distribution of n observations. We assume that $(n-k)$ of the observations are independently and identically distributed as geometric with probability density function

$$f\left(x \mid \theta\right)=\theta\left(1-\theta\right)^{x},\ 0<\theta<1,\ x=0,1,\ldots \tag{2.1}$$

while the remaining k observations are independently and identically distributed as geometric with probability density function $f\left(x \mid \alpha\theta\right)$, $0<\alpha<1$. Under the assumption of this k - outlier model, we have the survival function of X as

$$S\left(x\right)=P\left(X\geq x\right)=\left(1-\theta\right)^{x}. \tag{2.2}$$

Then

$$x_{(i)}=\frac{1}{\ln\left(1-\theta\right)}\ln S\left(x_{(i)}\right).$$

Define

$$Y_i = x_{(i)}, \quad A = \frac{1}{\ln(1-\theta)} \quad \text{and} \quad X_i = \ln S\left(x_{(i)}\right).$$

Then (2.2) takes the form $Y_i = AX_i$ and consequently, by the least square procedure

$$\hat{A} = \frac{\sum_{i=1}^{n} x_{(i)} \ln S\left(x_{(i)}\right)}{\sum_{i=1}^{n} \left(\ln S\left(x_{(i)}\right)\right)^2}$$

and the mean square error(MSE) of $\hat{A}$ is,

$$\omega_n = \frac{\sum_{i=1}^{n} \left[x_{(i)} - \left(\hat{A} \right) \ln S\left(x_{(i)}\right) \right]^2}{(n-2) \sum_{i=1}^{n} \left[\ln S\left(x_{(i)}\right) \right]^2}. \tag{2.3}$$

We may use the change in the values of (2.3) when outliers are present in the data profitably to detect the number of outliers present. For this we compute ω_n, the value of (2.3) based on all the n observations and then $\omega_n^{(d)}$, the value after excluding $d = x_{(n)}$, the largest observation in the set and obtain $\mu = \omega_n - \omega_n^{(d)}$. Then compute $\omega_n^{(d,e)}$, the value of (2.3) after excluding the last two observations $d = x_{(n)}$ and $e = x_{(n-1)}$. Then find $\mu = \omega_n^{(d)} - \omega_n^{(d,e)}$ and so on. The procedure ends when a μ takes a negative value for the first time. Now we conclude that the observations hitherto discarded belong to $f\left(x \mid \alpha\theta\right)$.

3. Detection of the number of outliers based on posterior density

Kale and Kale [19] propose a procedure based on posterior probability to identify the number of outliers present in exponential samples. For the sake of comparison, in this section, we have extended those results to geometric distribution. Let $\underline{x} = (x_1, x_2, ..., x_n)$ be a random sample containing k outliers, where k is unknown. We assume that $(n-k)$ of these observations are independently and identically distributed as geometric with probability density function (2.1) and the remaining k observations $x_{v_1}, x_{v_2}, ..., x_{v_k}$ are independently and identically distributed as geometric with probability density function $f\left(x \mid \alpha\theta\right)$, $0 < \alpha < 1$, where the indexing set of observations $v = \left(v_1, v_2, ..., v_k\right)$ is treated as a parameter over $N_{(k)}$, the subset of k integers out of n. Here we derive a procedure to determine the number of outliers present in the sample based on the posterior density $\Psi(v \mid \underline{x})$ of v. Since $0 < \alpha < 1$, Kale [17] ensures that the outliers are among the largest order statistics.

Let $g(\theta, \alpha)$ be any prior for (θ, α). Then the joint probability density function of $(\underline{x}, \theta, \alpha)$ is given by,

$$L_k\left(\underline{x}, \theta, \alpha, v\right) = C_1 P\left(v\right) g\left(\theta, \alpha\right) L_k\left(\underline{x} \mid \theta, \alpha, v\right)$$

where C, with various suffixes, represent the normalizing constants. Taking, $P(v) = \frac{1}{\binom{n}{k}}$, the joint probability distribution of $(\underline{x}, v)$ is

$$P_k(\underline{x},\nu)=\int_0^1\int_0^1 g(\theta,\alpha)\prod_{i=1}^{n-k}\left[\theta(1-\theta)^{x_{(i)}}\right]\prod_{i=n-k+1}^{n}\left[(\alpha\theta)(1-\alpha\theta)^{x_{(i)}}\right]d\theta\ \ d\ \alpha$$

$$=\frac{1}{\binom{n}{k}}\int_0^1\int_0^1 g(\theta,\alpha)\theta^n(1-\theta)^{\sum_{i=1}^{n-k}x_{(i)}}\alpha^k(1-\alpha\theta)^{T_\nu}\ d\theta\ \ d\ \alpha \tag{3.1}$$

where $T_\nu=\sum_{\nu_j\in\nu}x_{\nu_j}$.

Now the marginal probability density function of $\underline{x}$ is

$$q_k(\underline{x})=\sum_{\nu\in N_{(k)}}P_k(\underline{x},\nu).$$

Hence the posterior distribution of ν given $\underline{x}$ will be

$$\Psi_k(\nu\,|\,\underline{x})=\frac{P_k(\underline{x},\nu)}{q_k(\underline{x})}.$$

If we assume beta prior for (α,θ) and they are independently distributed, the joint prior density for (α,θ) is

$$g(\theta,\alpha)=C_1\theta^{\,p-1}(1-\theta)^{q-1}\alpha^{\,s-1}(1-\alpha)^{t-1},\quad 0<\alpha<1\ ,0<\theta<1,\ p,q,s,t>0\,.$$

Then

$$P_k(\underline{x}|\ \nu)=C_2\sum_{i=0}^{T_\nu}(-1)^i\binom{T_\nu}{i}B(P+i,N)B(S+i,t)$$

and

$$q_k(\underline{x})=C_2\sum_{\nu\in N_{(k)}}\sum_{i=0}^{T_\nu}(-1)^i\binom{T_\nu}{i}B(P+i,N)B(S+i,t)$$

where $P=n+p$, $N=\sum_{i=1}^{n-k}x_{(i)}+q$ and $S=s+k$. Hence

$$\Psi_k\left(v \mid \underline{x}\right)=\frac{\sum_{i=0}^{T_v}(-1)^i\binom{T_v}{i}B\left(P+i,N\right)B\left(S+i,t\right)}{\sum_{v\in N_{(k)}}\sum_{i=0}^{T_v}(-1)^i\binom{T_v}{i}B\left(P+i,N\right)B\left(S+i,t\right)}. \tag{3.2}$$

Now for each $k \le \left[\frac{n}{2}\right]$, $\Psi_k\left(\hat{v} \mid \underline{x}\right)$ with $\left(x_{\hat{v}_1}, x_{\hat{v}_2}, ..., x_{\hat{v}_k}\right) \equiv \left(x_{(n-k+1)}, x_{(n-k+2)}, ..., x_{(n)}\right)$ is determined. Then the number of outliers, k, present in the sample is that value of k, say k_0, for which

$$\Psi_{k_0}\left(\hat{v} \mid \underline{x}\right) = \underset{k}{Max}\ \Psi_k\left(v \mid \underline{x}\right).$$

4. Detection of the number of outliers by the least square method based on the empirical distribution function

Wu [24] proposed the least square procedure to determine the number of upper outliers in the continuous univariate distributions (Exponential, Pareto, Gumbel, Weibull etc.). In this section we have extended this procedure to find the number of outliers in the geometric sample by minimizing the sample mean square error (SMSE). For developing the theory, we consider a random sample $\underline{x} = (x_1, x_2, ..., x_n)$ in which the first $(n-k)$ ordered observations are independent and identically distributed as geometric with density (2.1), while the remaining k ordered observations $x_{(n-k+1)}, x_{(n-k+2)}, ..., x_{(n)}$ are independently and identically distributed as geometric with density $f\left(x \mid \alpha\theta\right)$.

The survival function of the geometric distribution with density (2.1) is

$$P\left(X \geq x_{(i)}\right) = (1-\theta)^{x_{(i)}}, \quad i = 1, 2, \ldots, (n-k)$$

$$= (1-\alpha\theta)^{x_{(i)}}, \quad i = n-k+1, \ldots, n.$$

Then

$$\ln P\left(X \geq x_{(i)}\right) = x_{(i)} \ln(1-\theta), \quad i = 1, 2, \ldots, (n-k)$$

$$= x_{(i)} \ln(1-\alpha\ \theta), \quad i = n-k+1, \ldots, n.$$

The empirical value of the survival function is equal to $\left(1-\frac{i}{n}\right)$. In order to avoid $\ln(0)$, D' Agostino and Stephans [5] suggested that it might be approximated by $\left(1-\frac{i-c}{n-2c+1}\right)$, $i = 1, 2, \ldots$ and $0 \leq c \leq 1$, generally. So we may write

$$x_{(i)} \ln(1-\theta) = \ln\left(\frac{n-i-c+1}{n-2c+1}\right), \quad i = 1, 2, \ldots, (n-k)$$

and

$$x_{(i)} \ln(1-\alpha\ \theta) = \ln\left(\frac{n-i-c+1}{n-2c+1}\right), \quad i = n-k+1, \ldots, n.$$

For each value of $k \in \left(0, 1, \ldots, \left[\frac{n}{2}\right]\right)$, we may find a least square solution of θ and α using the system of equations given above. They are

$$\hat{\theta}_k = 1 - Exp\left[\frac{\sum_{i=1}^{n-k} \ln\left(\frac{n-i-c+1}{n-2c+1}\right)}{\sum_{i=1}^{n-k} x_{(i)}}\right] \tag{4.1}$$

and

$$\hat{\alpha}_k = \frac{1}{\hat{\theta}_k}\left[1 - Exp\left(\frac{\sum_{i=n-k+1}^{n} \ln\left(\frac{n-i-c+1}{n-2c+1}\right)}{\sum_{i=n-k+1}^{n} x_{(i)}}\right)\right], \quad \text{for } 0 \le c \le 1. \tag{4.2}$$

Now the SMSE are

$$SMSE = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{(i)}\ln\left(1-\hat{\theta}_k\right) - \ln\left(\frac{n-i-c+1}{n-2c+1}\right)\right)^2, \quad \text{for} \quad k=0 \tag{4.3}$$

and

$$SMSE = \frac{1}{n-2}\left[\sum_{i=1}^{n-k}\left(x_{(i)}\ln\left(1-\hat{\theta}_k\right) - \ln\left(\frac{n-i-c+1}{n-2c+1}\right)\right)^2\right.$$

$$\left. + \sum_{i=n-k+1}^{n}\left(x_{(i)}\ln\left(1-\hat{\alpha}_k\hat{\theta}_k\right) - \ln\left(\frac{n-i-c+1}{n-2c+1}\right)\right)^2\right], \tag{4.4}$$

for $k \in \left(0,1,\ldots,\left[\frac{n}{2}\right]\right)$ and $0 \le c \le 1$.The value of k which result in the minimum SMSE is taken as the optimal solution.

5. Discussion and Conclusions

The procedures derived in the previous sections are assessed by a numerical study with simulated samples with different values for the parameters of the model, the hyper parameters of the prior distributions and for the real data set.

a) Simulation results

The results obtained for a few selected sets of parameter values when the samples contain one, two and three outliers are presented in Tables 1,2,3, in Tables

4,5,6 and in Tables 7,8,9 respectively and are given in the Appendix. The * denotes the number of outliers recognized by the procedure in the particular situation.

To assess the performance of the various procedures in specific situations, we obtained the rate of correct identification using Monte Carlo experiments. For this purpose, we generated random samples of sizes 10, 20, 50 and 100 and obtained the rate of identification of the actual number of outliers in the samples empirically, using 1000 Monte Carlo runs for different choices of the parameters when the sample contains one, two and three outliers. The result of the same for some selected parameter values are presented in Tables 10,11 and 12 in the Appendix.

b) Real Data Application

The pattern of natal dispersal in vertebrate animals is an important factor affecting the genetic and demographic processes within and between populations. The geometric probability distribution is a common way to model the frequency distribution of vertebrate dispersal distances (Porter and Dooley [21], Greenwood et al., [7]). Let X define the number of units (home ranges, habitat, nest sites, territories etc. with a fixed diameter) moved before stopping (settling and / or dying) and θ , the probability of stopping while crossing any one unit of habitat before moving an additional home- range diameter.

For an illustration of the present study, we use the following data about the dispersal distance (in units of 200 meters diameter) from natal site to first year breeding site for different categories of 177 one-year-old male great tits given in page 141,Appendix I, Greenwood et al. [7].

X	Frequency
0	17
1	24
2	23
3	17
4	12
5	10
6	4
7	2
8	3
9	1
10	1
11	1
12	1
16	1

It is very interesting to note that the SMSE criterion and the Ψ_k criterion based on the posterior probability procedure result in the wrong conclusion that the last observation in the set (the natal distance 16 units) is an outlier. But the μ criterion based on the MSE procedure gives the correct conclusion that it is not an outlier (see Table 13 in the Appendix). The normal probability plot of the data indicates that the observation is not an outlier. When Grubb's test for outlier detection (Burke S., [4]) is applied to test whether the observation 16 is actually an outlier or not, we get the conclusion that it is not an outlier at 95% and 99% levels of significance.

The conclusions from the empirical study are

1. The μ criterion based on the MSE procedure is the only procedure among the above three methods which can be applied when a frequency data is given instead of the set of individual observations.

2. In the two and three outlier cases, decisions based on the SMSE criterion result in better identification rates than the posterior probability method even when θ and α are large.
3. When one outlier is present, for all values of θ and α, the Ψ_k criterion based on the posterior probability procedure identifies it faster.
4. The μ criterion is good only in cases where θ and α are small. In the two and three outlier cases, this procedure ends with better rates of correct identification.
5. In the above real life application, the conclusion based on the μ criterion is correct and the other two methods result in wrong conclusions that the largest observation on X is an outlier.

Acknowledgements

The authors are thankful to the University Grants Commission, New Delhi, India, for the financial support.

References

1. F.J. Anscombe, Rejection of outliers, *Technometrics*, 2 (1960), 123-147.
2. V. Barnett, The Study of Outliers: Purpose and Model, *Applied Statistics* 27 (3) (1978), 242-250.
3. V. Barnett, T. Lewis, *Outliers in Statistical Data*, 2nd edition, John Wiley and Sons, New York (1984).
4. S. Burke, Missing values, Outliers, Robust statistics and Non Parametric methods, *VAM Bulletin*, 19 (1998), Autumn, 22-27.

5. R.B.D'Agostino, M.A. Stephens, *Goodness-of-fit Techniques*, Marcel Dekker, Inc., New York (1986).

6. U. Gather, B.K. Kale, Outlier generating models-a review, in Venugopal, N. (Eds.), *Contributions to Stochastics*, Wiley Eastern, New Delhi (1992), pp.57-86.

7. P.J. Greenwood, P.H. Harvey, and C.M. Perrins, The role of dispersal in the Great Tit (Parus Major): The causes, consequences and heritability of natal dispersal, *Journal of Animal Ecology*, 48 (1979), 123-142.

8. F.E. Grubbs, Procedures for detecting outlying observations in samples, *Technometrics* 11 (1969), 1-21.

9. D.M. Hawkins, *Identification of Outliers*, Chapman and Hall, London (1980).

10. E.S. Jeevanand, N.U. Nair, Bayes estimate of Pareto parameters in the presence of outliers, *Proceedings of the Symposium on Statistical Inference, Trivandrum, 16-18 June 1992,* (1992) 53 – 70,Educational Publishers and Distributors, Kerala.

11. E.S. Jeevanand, N.U. Nair, Estimation of Exponential parameters in the presence of outliers, *Biometrical Journal* 35(4) (1993a), 471- 478.

12. E.S. Jeevanand, N.U. Nair, Prediction of future observations in the presence of outliers, *Statistica* ,53 (2) (1993b), 171-176.

13. E.S. Jeevanand, N.U. Nair, Estimating P [X>Y] from exponential samples containing spurious observations, *Comm. in Statistics - Theory and Methods* 23(9) (1994), 2629-2642.

14. E.S. Jeevanand, N.U. Nair, Bayes prediction interval for the exponential and Pareto populations in the presence of an outlier, *Far East Journal of Mathematical Sciences,* Professor Bhattacharya Memorial Volume (1996), 81-95.

15. E.S. Jeevanand, N.U. Nair, On determining the number of outliers in exponential and Pareto sample, *Statistical Papers* 39 (3) (1998), 277 - 290.

16. P.C. Joshi, Estimation and testing under exchangeable exponential model with a single outlier, *Communications in Statistics - Theory and Methods* 17 (1988), 2315 - 2326.

17. B.K. Kale, Detection of outliers, *Sankhya* Series B 38 (1976), 356-363.

18. B.K. Kale, Outliers-a Review, *Journal of the Indian Statistical Association* 17 (1979), 51-67.

19. V.B. Kale, B.K. Kale, Outliers in exponential sample- a Bayesian approach, *Gujrat Statistical Review*, *Professor Khatri memorial Volume* (1992).

20. L.I. Pettit, Bayes method for outliers in exponential sample, *Journal of the Royal Statistical Society* Series B 50 (1988), 371-380.

21. Porter, J.H and Dooley J.L (1993): Animal dispersal patterns: A reassessment of simple mathematical models, *Ecology*, 74(8), 2436-2443.

22. S.M. Stigler, Simmon Newcomb, Percy Daniell, and history of robust estimation 1885-1950, *Journal of the American Statistical Association* 68 (1973), 872-879.

23. S.M. Stigler, Studies in the history of probability and statistics XXXVIII. - R.H. Smith, a Victorian interested in robustness, *Biometrika* 67 (1980), 217-221.

24. J.W. Wu, A note on determining the number of outliers in exponential sample by least squares procedure, *Statistical Papers* 42 (4) (2001), 489-503.

Appendix

Table 1: Detection of Outliers using posterior density, when one outlier is present in the sample

Parameters						k	Ψ_k
α = 0. 5							
n	θ	p	q	s	t		
						*1	0.9957
						2	0.4732
10	0.1	1	1	1	1	3	0.4417
						*1	0.9998
						2	0.2377
20	0.3	1	1	1	2	3	0.1169
						*1	0.4629
						2	0.0600
50	0.5	1	1	2	2	3	0.0150
						*1	0.4630
						2	0.0229
100	0.7	2	2	2	2	3	0.0012

*** :** *The number of outliers identified in the sample*

Table 2: Detection of Outliers using empirical Survival function by least square procedure, when one outlier is present in the sample

Parameters		k	SMSE		
α = 0. 5					
n	θ		c = 0. 0	c = 0. 3	c = 0. 5
10	0.1	0	0.6365	0.5924	0.5458
		*1	0.0472	0.0406	0.0371
		2	0.1835	0.1866	0.1852
		3	0.353	0.3541	0.3469
20	0.3	0	1.1721	1.1256	1.0726
		*1	0.1741	0.1591	0.1481
		2	0.2029	0.1979	0.1916
		3	0.2595	0.257	0.2496
50	0.5	0	1.0732	1.0268	0.9783
		*1	0.2479	0.2386	0.2319
		2	0.2719	0.2681	0.2639
		3	0.3271	0.3238	0.3183
100	0.7	0	7.2641	7.2493	7.2105
		*1	1.1814	1.1734	1.1674
		2	1.2643	1.2686	1.2712
		3	1.4716	1.4832	1.4901

*** :** *The number of outliers identified in the sample*

Table 3: Detection of Outliers using Survival function by least square procedure, when one outlier is present in the sample

Parameters			
α = 0. 5			
n	**θ**	**k**	**μ**
		*1	-0.5868
		2	0.2212
		3	0.0747
10	0.1	4	0.3798
		*1	-0.2156
		2	0.0034
		3	0.0226
20	0.3	4	0.0172
		*1	-0.0018
		2	0.0012
		3	0.0032
50	0.5	4	0.0023
		*1	-0.0066
		2	0.0012
		3	0.0008
100	0.7	4	0.0007

*** :** *The number of outliers identified in the sample*

Table 4: Detection of Outliers using posterior density, when two outliers are present in the sample

Parameters						k	Ψ_k
α = 0. 5							
n	**θ**	**p**	**q**	**s**	**t**		
						1	0.7629
						*2	0.9999
						3	0.6665
10	0.1	1	1	1	1	4	0.6396
						1	0.8451
						*2	0.9883
						3	0.1918
20	0.3	1	1	1	2	4	0.0966
						1	0.6552
						*2	0.6745
						3	0.0463
50	0.5	1	1	2	2	4	0.0061
						1	0.6306
						*2	0.8745
						3	0.1946
100	0.7	2	2	2	2	4	0.0522

*** :** *The number of outliers identified in the sample*

Table 5: Detection of Outliers using empirical Survival function by least square procedure, when two outliers are present in the sample

Parameters		k	SMSE		
α = 0. 5					
n	θ		c = 0. 0	c = 0. 3	c = 0. 5
		0	0.5186	0.4991	0.4846
		1	0.4348	0.4062	0.3843
		*2	0.0085	0.0167	0.0299
		3	0.1311	0.1399	0.1507
10	0.1	4	0.2342	0.2396	0.2456
		0	0.768	0.725	0.6788
		1	0.2268	0.2203	0.2152
		*2	0.0662	0.069	0.0694
		3	0.2154	0.2203	0.2193
20	0.3	4	0.3421	0.346	0.3416
		0	1.9669	1.9182	1.8634
		1	0.6103	0.5959	0.5845
		*2	0.2175	0.2113	0.2057
		3	0.3027	0.2997	0.2948
50	0.5	4	0.4164	0.4146	0.409
		0	3.0147	2.9757	2.934
		1	2.2663	2.2582	2.2503
		*2	0.8722	0.8703	0.8697
		3	0.9743	0.9743	0.9743
100	0.7	4	1.074	1.0748	1.0747

*** :** *The number of outliers identified in the sample*

Table 6: Detection of Outliers using Survival function by least square procedure, when two outliers are present in the sample

Parameters		k	μ
α = 0. 5			
n	θ		
		1	1.8469
		*2	- 0.7958
		3	0.4062
10	0.1	4	0.4037
		1	0.2559
		*2	- 0.2081
		3	0.0176
20	0.3	4	0.0412
		1	0.0232
		*2	- 0.0137
		3	0.0022
50	0.5	4	0.0034
		1	0.0001
		*2	- 0.0087
		3	0.0004
100	0.7	4	0.0006

*** :** *The number of outliers identified in the sample*

Table 7: Detection of Outliers using posterior density, when two outliers are present in the sample

Parameters						k	Ψ_k
α = 0. 5							
n	θ	p	q	s	t		
						1	0.3421
						2	0.4353
						*3	0.9999
10	0.1	1	1	1	1	4	0.2942
						1	0.4687
						2	0.4034
						*3	0.7082
20	0.3	1	1	1	2	4	0.2014
						1	0.7408
						2	0.4718
						*3	0.9925
50	0.5	1	1	2	2	4	0.0632
						1	0.3405
						2	0.5357
						*3	0.9907
100	0.7	2	2	2	2	4	0.0831

*** :** *The number of outliers identified in the sample*

Table 8: Detection of Outliers using empirical Survival function by least square procedure, when three outliers are present in the sample

Parameters α = 0. 5			SMSE		
n	θ	k	c = 0. 0	c = 0. 3	c = 0. 5
		0	0.1978	0.1597	0.1269
		1	0.1009	0.0876	0.0783
		2	0.0458	0.0383	0.0328
		*3	0.0396	0.0326	0.0255
10	0.1	4	0.1051	0.0917	0.0776
		0	0.9699	0.9231	0.8744
		1	0.5438	0.5275	0.5147
		2	0.2385	0.2286	0.2202
		*3	0.0802	0.0739	0.0663
20	0.3	4	0.2113	0.2055	0.196
		0	1.9933	1.9528	1.9097
		1	1.5293	1.5189	1.5097
		2	1.0645	1.0601	1.0565
		*3	0.2021	0.1979	0.195
50	0.5	4	0.2924	0.2895	0.2865
		0	4.0912	4.0662	4.0372
		1	3.9018	3.9015	3.8985
		2	2.868	2.874	2.8785
		*3	0.6829	0.6831	0.6848
100	0.7	4	0.7608	0.7631	0.7658

*** :** *The number of outliers identified in the sample*

Table 9: Detection of Outliers using Survival function by least square procedure, when three outliers are present in the sample

Parameters $\alpha = 0.5$			
n	**θ**	**k**	**μ**
		1	2.2368
		2	0.0528
		*3	- 0.1814
10	0.1	4	0.2159
		1	0.2414
		2	0.0115
		*3	- 0.0322
20	0.3	4	0.0041
		1	0.0334
		2	0.0053
		*3	- 0.0321
50	0.5	4	0.0048
		1	0.0059
		2	0.0011
		*3	- 0.0159
100	0.7	4	0.0007

*** :** *The number of outliers identified in the sample*

Table 10: Rate of identification of the number of outliers correctly by the procedures proposed – when one outlier is present in the data

Parameter	θ = 0.1			θ = 0. 3			θ = 0. 5			θ = 0. 8		
Criteria α	Ψ_k	SMSE	μ	Ψ_k	SMSE	μ	Ψ_k	SMSE	μ	Ψ_k	SMSE	μ
0. 2	0.935	0.967	0.958	0.925	0.954	0.935	0.934	0.915	0.928	0.921	0.882	0.919
0. 4	0.928	0.973	0.796	0.926	0.927	0.452	0.918	0.908	0.412	0.915	0.876	0.421
0. 6	0.944	0.981	0.352	0.918	0.912	0.311	0.921	0.911	0.309	0.911	0.881	0.287
0. 8	0.929	0.959	0.214	0.903	0.918	0.186	0.907	0.892	0.215	0.909	0.854	0.254

Table 11: Rate of identification of the number of outliers correctly by the procedures proposed – when two outliers are present in the data.

Parameter	θ = 0. 1			θ = 0. 3			θ = 0. 5			θ = 0. 8		
Criteria α	Ψ_k	SMSE	μ	Ψ_k	SMSE	μ	Ψ_k	SMSE	μ	Ψ_k	SMSE	μ
0. 2	0.954	0.941	0.928	0.921	0.901	0.919	0.854	0.906	0.924	0.812	0.802	0.931
0. 4	0.923	0.881	0.901	0.865	0.873	0.904	0.724	0.876	0.916	0.506	0.864	0.865
0. 6	0.936	0.853	0.658	0.831	0.844	0.825	0.639	0.812	0.803	0.313	0.801	0.811
0. 8	0.944	0.846	0.547	0.820	0.831	0.812	0.515	0.809	0.813	0.196	0.698	0.802

Table 12: Rate of identification of the number of outliers correctly by the procedures proposed – when three outliers are present in the data.

Parameter	θ = 0. 1			θ = 0. 3			θ = 0. 5			θ = 0. 8		
Criteria α	Ψ_k	SMSE	μ	Ψ_k	SMSE	μ	Ψ_k	SMSE	μ	Ψ_k	SMSE	μ
0. 2	0.825	0.921	0.165	0.728	0.912	0.182	0.686	0.841	0.201	0.552	0.689	0.191
0. 4	0.805	0.913	0.211	0.713	0.856	0.241	0.515	0.816	0.211	0.386	0.608	0.241
0. 6	0.768	0.906	0.203	0.585	0.834	0.319	0.504	0.808	0.254	0.213	0.521	0.208
0. 8	0.729	0.921	0.315	0.525	0.811	0.346	0.468	0.796	0.378	0.185	0.412	0.341

Table 13: Detection of Outliers in the data given by Greenwood et al. [7].

k	Ψ_k	**SMSE**	**μ**
0		0.0179	
1	0.0928*	0.0168* *	0.0125
2	0.0113	0.0176	0.0009
3	0.0019	0.0177	0.0004

* : Highest value of Ψ_k, * * : Least value of SMSE