Jordan Journal of Mathematics and Statistics (JJMS) 4(1), 2011, pp.47 - 60

LONGITUDINAL DATA ANALYSIS USING GENERALIZED MAXIMUM ENTROPY APPROACH

MOHAMMAD Y. AL-RAWWASH[(1)] AND AMJAD D. AL-NASSER[(2)]

Abstract. Marginal generalized linear models are frequently used for the analysis of repeated measurements and longitudinal data. During the last three decades, researchers used parametric, nonparametric as well as Bayesian methods as useful approaches to model such kind of data. The correlation among the repeated measurements is considered a vital factor to increase the estimation efficiency of the model's parameters for different correlation structures. This article suggests using the generalized maximum entropy (GME) as an efficient method for the joint modelling of mean and correlation parameters that permits the estimation with minimum distributional assumptions. Moreover, we present a simulation study to compare the performance of the GME method with a set of well known estimation methods in the longitudinal data literatures.

1. Introduction

In the literature of longitudinal data there are two broad approaches based on the idea of generalized estimating equations and mixed models for estimation of regression and correlation parameters ([9] and [17]). The generalized estimating equations (GEE) as well as its various modifications are widely used in health, environmental and social sciences to estimate the regression and correlation parameters (see [7]; [3];

2000 *Mathematics Subject Classification.* 62J12.

Key words and phrases. Gaussian estimation; Generalized estimating equations; Generalized maximum entropy.

Received: July 15 , 2010 Accepted : Dec. 23 , 2010 .

[4] and [10]). On the other hand, longitudinal data analysis based on mixed models usually exploits the Gaussian likelihood function as its objective function along with the often tentative normality assumption to estimate the parameters. In fact, longitudinal data analysis is a byproduct of generalized linear models and time series analysis where we address the relationships between a response variable and the corresponding explanatory variables that are measured repeatedly over time.To elaborate on these ideas, we consider longitudinal data that can be described as follows. Let $y_i = (y_{i1}, y_{i2}, \cdots, y_{in})'$ be a vector of repeated measurements taken on a generic subject at times $t = (t_1, t_2, \cdots, t_n)$ such that $t_1 < t_2 < \cdots < t_n$ and the associated covariates $x_j = (x_{j1}, \cdots, x_{jp})'$. In general, no distributional assumption is made about y_i other than those about its first two moments. The analysis is carried out based on a fixed effect model for longitudinal data, thus the model for subject i is given as

$$y_i = X_i\beta + \epsilon_i, \quad i = 1, 2, \cdots, m \tag{1.1}$$

where X_i is an $(n \times p)$ design matrix, $\beta = (\beta_1, \beta_2, \cdots, \beta_p)'$ is a set of unknown parameters and $\epsilon_i = (\epsilon_1, \cdots, \epsilon_n)'$ is a set of random errors with mean 0 and an $n \times n$ positive definite covariance matrix Σ. However, in many cases we prefer to concentrate on a parsimonious working covariance structure (autoregressive, equicorrelated, tridiagonal, $\cdots$, etc.) to comprise and model the within-individual error. Some of these structures depend on a single parameter ρ that represents the correlation between two adjacent observations collected on a given subject. In this article, we assume that the vector of within-individual errors follows a first order autoregressive model such that

$$\epsilon_{ij} = \rho\epsilon_{i(j-1)} + a_{ij}, \tag{1.2}$$

where a_{ij} is error term with mean 0 and variance σ^2 and ρ is the correlation parameter. ([4]) developed the so-called generalized Gaussian likelihood function as:

$$GL(\beta, \mathbf{\Sigma}^*, \mathbf{\Sigma}; \mathbf{Y}) = \log |\mathbf{\Sigma}^*| + \mathbf{Z}(\beta)'\mathbf{\Sigma^{-1}}\mathbf{Z}(\beta), \tag{1.3}$$

where $\mathbf{Z}$ is the "standardized" counterpart of $\mathbf{Y}$ such that $\mathbf{Z} = V^{-\frac{1}{2}}(\mathbf{y} - \mu)$ and the matrix $V =$ diag $(v(\mu_1), \cdots, v(\mu_n))$ and $\mu = \mu(\beta) = (\mu_1, \cdots, \mu_n)'$. The variance function $v(.)$ depends sometimes on the mean (Poisson or count data). Let $\mathbf{\Sigma}$ be a block-diagonal covariance matrix with non-zero blocks $\Sigma_i = \text{cov}(\mathbf{y_i})$ and $\mathbf{\Sigma}^*$ is an $N \times N$ covariance matrix (possibly depending on $\mathbf{\Sigma}$ and X_i's), for $N = nm$ and $y = (y_1, y_2, \cdots, y_m)$

The generalized Gaussian likelihood function subsumes many of the estimating functions developed in the literature ([7]; [3] and [4]). The estimation strategy in such situations focuses on minimizing such objective function with respect to the parameters of interest (see [3]). It is clear that when $\mu_i = X_i\beta$, then the parameter estimate that minimizes (1.3) is $\hat{\beta} = (\sum_{i=1}^{m} X_i'\Sigma_i^{-1}X_i)^{-1}(\sum_{i=1}^{m} X_i'\Sigma_i^{-1}Y_i)$. In other cases such as binary and count data, a link function may be introduced to present the relation between μ and β. One major pitfall that many authors had discussed is the feasibility of the correlation parameter estimates (see [3]; [8] and [4]). In many situations, the parametric model may get complicated to the extent that an alternative estimation strategy become imminent. Not only nonparametric and semiparametric models have been proposed in the literature but also Bayesian and other significant approaches have been discussed extensively and substantial contributions were introduced in this area. These ideas focused on modeling the mean and association in the longitudinal data setup to avoid model misspecification [5].

In this paper, we intend to introduce the GME [12], as an alternative methodology for the classical estimation methods of the regression and correlation parameters in the longitudinal data setup. The list of merits that motivates us to use the GME approach includes but not limited to the following. First, the GME tends to dominate traditional estimation methods when we have small samples. Also, the data may be non-experimental noisy data and finally the data might be observed from a weak designed experiment. Moreover, the maximum likelihood estimator is unattractive approach when we adopt nonlinear model since it may not be robust to the underlying (unknown) distribution. Also, the GME estimator is expected to be highly efficient under various circumstances including skewed distribution, the existence of outliers and the existence of many highly correlated covariates, ([12]; [15]; [11] and [6]). Finally, the GME is considered a promising approach when there are no restrictive sampling assumptions or in the case of negative degrees of freedom (ill-posed problems) or undetermined ones.

The organization of the paper is as follows. Section 2 discusses the idea of generalized maximum entropy and presents the notations and model. In section 3, we derive the parameter estimators of the regression and correlation parameters using the GME approach. The methodology is illustrated in section 4 using a simulation study and the results are compared using different estimation methodologies. Finally, section 5 provides concluding remarks on the needs and merits of the generalized maximum entropy.

2. **Model Formulation and Estimation**

In this section, we introduce the setup and model formulation using the generalized maximum entropy method for the longitudinal data framework when the repeated measurements are collected for different subjects. The GME methodology has been repeatedly proposed in various fields of applications covering not only statistical aspects but also economical, agricultural as well as engineering. It is anticipated that

GME will compete with the classical estimation methodologies especially when we expect to have skewed or heavy tailed distributions([12] and [6]).

2.1. **Generalized maximum entropy formulation.** The history of entropy goes back to [18] where he defined the information entropy of the distribution (discrete events) with corresponding probabilities $P = \{p_1, p_2, \cdots, p_n\}$ as $H(p) = -\sum_{i=1}^{n} p_i \ln(p_i)$, where $0\ln 0 = 0$. Note that the entropy function exposes the uncertainty regarding the appearance of a set of events. Many extensions and enhancements have appeared in the literature of entropy that presented various and more complicated information entropy measures such as Rényi and Tsallis (see [13]).

Maximizing entropy subject to various side conditions is well known in the literature as a method of deriving the most uniform distribution compatible with the prior information ([14]; [19]; [1] and [2]). The basic idea underlying the GME approach is to view each one of the unknown parameters and the error terms as an expected value of some proper probability distribution defined over a given bounded support. Accordingly, the GME idea is to rewrite the model in terms of the new convex reparametrizations. Then, the estimation methodology is to recover simultaneously the probability distribution of the parameters using the data as well as some normalization constraints. It is noteworthy that if the support values of the regression parameters are not available, the researcher should provide support values that are symmetric around zero while the support values for the error terms are chosen with respect to the 3 sigma rule [16]. Hence, maximizing the joint entropies subject to the data represented by the model and the add up normalization constraints may provide competitive estimates with minimal distributional assumptions.

2.2. **Model Formulation and Estimation.** The linear model given in (1.1) and (1.2) can be rewritten as

$$y_{ij} = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ijk} + \epsilon_{ij},$$

where $\epsilon_{ij} = \rho\epsilon_{i(j-1)} + a_{ij}$.

Following the GME principles ([12]), the unknown parameters β_0, β_k, ρ as well as the error term a_{ij} for $k = 1, 2, \cdots, p$, $j = 1, 2, \cdots, n$ and $i = 1, 2, \cdots, m$ should be reparametrize to be in the form of an expected value of discrete random variable. Therefore, we may represent β_0 as a convex combination using a set of discrete random variable $b_r, r = 1, 2, \cdots, R$ with $R \geq 2$ possible realizations and their corresponding probabilities $q_r, r = 1, 2, \cdots, R$ which can be formulated as:

$\beta_0 = \sum_{r=1}^{R} b_r q_r$; $\quad \sum_{r=1}^{R} q_r = 1$; $\quad$ where $q_r \in (0, 1)$

Usually the values of b_r may be used as supports on β_0 and we may choose these values based on prior information of β_0. Similarly, we may reparametrize the rest of the unknown parameters. The error term is treated as a finite and discrete random variable that might be pictured as bounds of the error term. Therefore, we may have the following

(1) $\beta_k = \sum_{r=1}^{T} z_{kr} f_{kr}$; $\quad \sum_{r=1}^{T} f_{kr} = 1$; $\quad k = 1, 2, \cdots, p$

(2) $a_{ij} = \sum_{r=1}^{H} v_{ijr} w_{ijr}$; $\quad \sum_{r=1}^{H} w_{ijr} = 1$; $\quad j = 1, 2, \cdots, n$ and $i = 1, 2, \cdots, m$

(3) $\rho = \sum_{r=1}^{L} \psi_r \xi_r$; $\quad \sum_{r=1}^{L} \xi_r = 1,$

where z_{kr} is a set of T realizations of a discrete random variable with their corresponding probabilities f_{kr}. Also, v_{ijr} and ψ_r are realizations of random variables with the corresponding probabilities w_{ijr} and ξ_r, respectively.

The restrictions imposed on the parameter space through the values of b and z while reparametrizing β_0 and β_k's are specified uniformly and symmetrically around zero with equally spaced distance discrete points. In fact, [12] conducted some monte carlo experiments and found that the greatest improvement in precision comes from using 5 support points. On the other hand, the actual bounds for v_{ijr} depend on the observed values or any empirical information about the error. If such information is not available, then we may use the three-sigma rule (see [16]) to specify these bounds over the interval [-3S_y, 3S_y] where S_y is the standard deviation of the observed values y. Moreover, the empirical GME literature indicates that the value of H is 3. Finally, the parameter space of ρ will be chosen in the interval [-1,1].

3. GME System and Solution for Longitudinal Data

The GME system is a nonlinear programming problem that is built to have shannon's entropy as our main objective function in addition to some important constraints obtained from the data formulation and reparametrization. Consequently, our system is viewed as the maximization of the following function

$$\begin{aligned} H(\mathbf{q},\mathbf{f},\mathbf{w},\xi) &= -\mathbf{q}'\mathbf{ln}(\mathbf{q}) - \mathbf{f}'\mathbf{ln}(\mathbf{f}) - \mathbf{w}'\mathbf{ln}(\mathbf{w}) - \xi'\mathbf{ln}(\xi) \\ &= -\sum_{r=1}^{R} q_r ln(q_r) - \sum_{k=1}^{p}\sum_{r=1}^{T} f_{kr} ln(f_{kr}) - \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{r=1}^{H} w_{ijr} ln(w_{ijr}) \\ &\quad -\sum_{r=1}^{L} \xi_r ln(\xi_r) \end{aligned} \tag{3.1}$$

subject to the following model of interest and the add up normalization constraints

$$y_{ij} = \sum_{r=1}^{R} b_r q_r + \sum_{k=1}^{p} x_{ijk}\Big(\sum_{r=1}^{T} z_{kr} f_{kr}\Big) + \rho\epsilon_{i(j-1)} \sum_{r=1}^{H} v_{ijr} w_{ijr},$$

$\sum_{r=1}^{R} q_r = 1, \quad \sum_{r=1}^{L} \xi_r = 1, \quad |\sum_{r=1}^{L} \psi_r \xi_r| \le 1, \quad \sum_{r=1}^{T} f_{kr} = 1, \quad k = 1, 2, \cdots, p$

and

$\sum_{r=1}^{H} w_{ijr} = 1, \qquad i = 1, 2, \cdots, m, \quad j = 1, 2, \cdots, n$

It can be noted that the number of unknowns is equal to $(R + L + PT + nmH)$ while the number of constraints is equal to $(2mn + P + 3)$. The number of unknowns and constraints get larger as both n and m increase which indicates that GME should perform better for small samples. In order to solve such a nonlinear programming system analytically, we need to simplify the inequality constraints such that it satisfies Kuhn-Tucker conditions. Hence, there will be a need to introduce some slack variables in order to form the following lagrangian function:

$$\begin{aligned} L &= H(q, f, w, \xi) - \sum_{i=1}^{m}\sum_{j=1}^{n} \lambda_{ij}[y_{ij} - \sum_{r=1}^{R} b_r q_r - \sum_{k=1}^{p} x_{ijk}(\sum_{r=1}^{T} z_{kr} f_{kr}) - \sum_{r=1}^{H} v_{ijr} w_{ijr}] \\ &- \mu_1[\sum_{r=1}^{R} q_r - 1] - \sum_{k=1}^{p} \eta_k[\sum_{r=1}^{T} f_{kr} - 1] - \sum_{i=1}^{m}\sum_{j=1}^{n} \theta_{ij}[\sum_{r=1}^{H} w_{ijr} - 1] \\ &- \mu_2[\sum_{r=1}^{L} \xi_r - 1] - \mu_3[\sum_{r=1}^{L} \psi_r \xi_r + S_1^2] - \mu_4[-\sum_{r=1}^{L} \psi_r \xi_r + S_2^2], \end{aligned}$$

where $\lambda_{ij}, \theta_{ij}, \eta_k$ and μ_d are the lagrangian multipliers for $i = 1, 2, \cdots, m; j = 1, 2, \cdots, n; k = 1, 2, \cdots, p$ and $d = 1, 2, 3, 4$. Also, we introduce S_1^2 and S_2^2 as slack variables. Accordingly, we use lagrange's method of undetermined multipliers to derive the following solutions:

$$\hat{q}_r = \frac{exp\{-b_r[1 - \hat{\rho}] \sum_{i=1}^{m} \sum_{j=1}^{n} \hat{\lambda}_{ij}\}}{\sum_{r=1}^{R} exp\{-b_r[1 - \hat{\rho}] \sum_{i=1}^{m} \sum_{j=1}^{n} \hat{\lambda}_{ij}\}}$$

$$\hat{f}_{kr} = \frac{exp\{-z_{kr} \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ijk}^{*} \hat{\lambda}_{ij}\}}{\sum_{r=1}^{T} exp\{-z_{kr} \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ijk}^{*} \hat{\lambda}_{ij}\}}$$

$$\hat{w}_{ijr} = \frac{exp\{-\hat{\lambda}_{ij} v_{ijr}\}}{\sum_{r=1}^{H} exp\{-\hat{\lambda}_{ij} v_{ijr}\}}$$

and

$$\hat{\xi}_r = \frac{exp\{-\hat{\beta}_0\psi_r \sum_{i=1}^{m}\sum_{j=1}^{n}\hat{\lambda}_{ij}\}}{\sum_{r=1}^{L} exp\{-\hat{\beta}_0\psi_r \sum_{i=1}^{m}\sum_{j=1}^{n}\hat{\lambda}_{ij}\}}.$$

Then, the estimated parameters can be obtained via:

$\hat{\beta}_0 = \sum_{r=1}^{R} b_r\hat{q}_r$, $\hat{\beta}_k = \sum_{r=1}^{T} z_{kr}\hat{f}_{kr}$, $\hat{a}_{ij} = \sum_{r=1}^{H} v_{ijr}\hat{w}_{ijr}$, and $\hat{\rho} = \sum_{r=1}^{L} \psi_r\hat{\xi}_r$.

4. **Simulation Study**

In this section, we intend to use simulated longitudinal data to compare the performance of GME with some existing techniques used in the literature including GEE and GE obtained from (1.3). We assume that the true and working correlation structure in this simulation is set to be an autoregressive of order one (AR(1)). The simulation study is conducted to illustrate the performance of the GME method in estimating the correlation as well as the regression parameters in the longitudinal data analysis. [9] provided a simulation study to investigate the efficiency of the ordinary least squares using the AR(1) correlation structure assumption. We use the same simulation setup and consider m simulated subjects $y_1, y_2, \cdots, y_m$ each has five repeated measurements at five different occasions $t = (-2, -1, 0, 1, 2)$. The simulation is performed by generating 1000 data sets from a contaminated multivariate t distribution under the following conditions

$$y_{ij} = \beta_0 + \beta_1 t_j + \epsilon_{ij},$$

for $j = 1, 2, \cdots, n$ and $i = 1, 2, \cdots, m$, the number of repeated measurements n is set to be 5, and m = 10, 15, 25, 30. The model parameters β_0 and β_1 are initialized by 2 while we allow ρ to take the values -0.8(0.2)0.8. For GME estimates, we set

3 support points of the parameters space β_0 and β_1 to be symmetric around zero in the interval [-100, 0, 100], however we set 5 data points for the ρ in the interval [-1,1]. Finally the support points for the residual obtained by setting 3 data points symmetrically around zero based on the 3 σ-rule, [-3S, 0, 3S] where S is the standard deviation of the dependent variable. Then the simulated mean square error(MSE) is calculated using the following formula:

$$MSE = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\beta}_i - \beta)^2$$

The detailed results of our simulation comparing the methods GE, GEE and GME considered in this study are given in Table 1 and Table 2 according to the initial values of ρ. The results illustrated in the tables show the estimate values of ρ as well as the relative efficiency of GME compared with GEE and GE for different values of ρ. The relative efficiency is obtained as follows

$$eff(\beta_{GME}) = \frac{MSE(\beta^*)}{MSE(\beta_{GME})}$$

where β^* is the estimate of β using the GEE approach or the GE method. All results show that the GME method is better than its counterpart estimates in view of the relative efficiency values when we have small number of subjects. Moreover, the GME starts to be less efficient compared to the other methods when we increase the sample size. The estimate values of the ρ are obtained using the GME approach and it seems to be acceptable.

Table 1: The estimate value of ρ as well as the relative efficiency of GME compared with GEE for different values of ρ

		$m = 10$		$m = 15$		$m = 25$		$m = 30$	
ρ	$\hat{\rho}$	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
-0.8	-0.86	1.24	2.80	0.74	1.99	0.40	1.36	0.39	0.69
-0.6	-0.54	1.58	2.25	0.87	1.85	0.42	1.27	0.57	0.81
-0.4	-0.32	1.41	2.54	0.81	1.56	0.51	1.22	0.48	0.62
-0.2	-0.23	2.71	2.35	1.09	2.40	1.73	0.96	0.72	0.49
0.0	-0.09	1.11	1.21	1.05	1.12	1.04	0.98	0.97	0.73
0.2	0.15	2.15	2.08	1.33	2.03	1.12	0.91	0.92	0.83
0.4	0.34	2.74	2.35	1.47	2.15	1.34	1.05	1.02	0.78
0.6	0.65	2.98	2.11	1.52	2.12	1.41	1.00	0.95	0.76
0.8	0.71	2.65	2.91	1.89	1.74	1.26	1.11	0.91	0.98

Table 2: The estimate value of ρ as well as the relative efficiency of GME compared with GE for different values of ρ

		$m = 10$		$m = 15$		$m = 25$		$m = 30$	
ρ	$\hat{\rho}$	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
-0.8	-0.86	1.79	3.77	1.09	2.67	0.64	1.66	0.52	0.84
-0.6	-0.54	1.88	2.85	1.17	2.85	1.21	1.25	0.61	0.76
-0.4	-0.32	2.41	2.54	1.12	2.56	1.42	1.28	0.49	0.64
-0.2	-0.23	2.97	2.45	1.23	2.70	3.44	1.17	0.70	0.60
0.0	-0.09	1.12	1.24	1.11	1.22	1.01	1.05	0.95	0.76
0.2	0.15	1.85	2.41	1.53	2.01	1.14	0.97	0.98	0.93
0.4	0.34	2.74	2.68	1.74	2.14	1.53	1.06	1.03	0.87
0.6	0.65	2.99	2.81	1.82	2.71	1.74	1.01	0.97	0.95
0.8	0.71	2.69	2.93	1.91	1.69	1.60	0.97	0.92	0.94

5. Conclusions and Remarks

This paper discusses the use of the GME estimation method in studying the longitudinal data models assuming an autoregressive type correlation among the repeated measurements. The main advantage of GME approach is the waiver of any possible distributional assumptions which allows the researcher to maneuver and extract the most available information to obtain the parameter estimators. However, one disadvantage of the GME is the difficulty of dealing with large sample size and the methodology will slow down when we increase the number of parameters. Under simulation assumptions, the comparisons results shown in Tables 1 and 2 indicate that the GME estimates are more accurate and more efficient for small samples compared to the traditional estimation methods. Based on such results, one may consider the GME as a competitive estimation method that provide feasible and efficient estimates compared to the existence estimation techniques.

Acknowledgement

We would like to thank the editor and the referees for providing thoughtful comments and suggestions which led to substantial improvement in the presentation of this article.

References

[1] A. Al-Nasser, Customer satisfaction measurement models: Generalized maximum entropy approach, *Pakistan Journal of Statistics*, 19(2) (2003), 213-226.

[2] A. Al-Nasser, Entropy type estimator to simple linear measurement error models, *Austrian Journal of Statistics*, 34(3) (2005), 283-294.

[3] M. Al-Rawwash, Gaussian estimation and modeling covariance in longitudinal data analysis, Ph.D. thesis, Northern Illinois University, 2001.

[4] M. Al-Rawwash and M. Pourahmadi, Gaussian estimation and joint modeling of dispersions and correlations in longitudinal data, *Computer methods and Programs in Biomedicine*, 82 (2006), 106-113.

[5] P. Buzkova and T. Lamley, Semiparametric log-linear regression for longitudinal measurements subject to outcome-dependent follow-up, *Journal of Statistical Planning and Inference*, 138 (2008), 2450-2461

[6] E. Ciavolino and A. D. Al-Nasser, Comparing Generalized Maximum Entropy and Partial Least Squares Methods for Structural Equation Models, Journal of Nonparametric Statistics (JNPS), 21(8) (2009), 1017-1036.

[7] N. R. Chaganty, An alternative approach to the analysis of longitudinal data via generalized estimating equations, *Journal of Statistical Planning and Inference*, 63 (1997), 39-54.

[8] M. Crowder, On repeated measure analysis with misspecified covariance structure, *Journal of Royal Statistical Society* series B, 63 (2001), 55-62.

[9] P. J. Diggle, K.Y. Liang and S.L. Zeger, *Analysis of Longitudinal Data*, Oxford: Oxford University Press, 1994.

[10] A. Gelman, Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis*, 1 (2006), 515-533.

[11] A. Golan, Information and Entropy Econometrics - A Review and Synthesis, *Foundations and Trends in Econometrics*, 2 (2008), 1-145.

[12] A. Golan, G. Judge and D. Miller, *A Maximum Entropy Econometrics: Robust Estimation with Limited Data*, New York: Wiley, 1996.

[13] A. Golan and J. M. Perloff, Comparison of maximum entropy and higher-order entropy estimators, *Journal of Econometrics*, 107(1) (2002), 195-211.

[14] E. T. Jaynes, Information and statistical mechanics II, *Physics Review*, 108 (1957), 171-190.

[15] L. Peeters, Estimating a random-coefficients sample-selection model using generalized maximum entropy, *Economics Letters*, 84 (2004), 87-92.

[16] F. Pukelshein, The three sigma rule, *The American Statistician*, 48(2) (1994), 88-91.

[17] G. Qin and Z. Zhu, Robust Estimation in Partial Linear Mixed Model for Longitudinal Data, *Acta Mathematica Scientia*, 28 (2008), 333-347.

[18] C. E. Shannon, A mathematical theory of communication, *Bell System Technical Journal*, 27 (1948), 379-423.

[19] J. Skilling, The axioms of maximum entropy, in: Skilling, J (Ed.), *maximum entropy and bayesian methods in science and engineering*, Kluwer Academic, Dordecht, (1989), 173-187.

(1)Department of Mathematics, University of Sharjah, Sharjah, UAE

E-mail address: malrawwash@sharjah.ac.ae

(2)Department of statistics, Yarmouk University , Irbid Jordan

E-mail address: amjadn@yu.edu.jo