STATISTICS IN TRANSITION-new series, July 2010
Vol. 11, No. 1, pp. 9—24

AN INFORMATION – THEORETIC APPROACH TO THE MEASUREMENT ERROR MODEL

Amjad D. Al-Nasser

ABSTRACT

In this paper, the idea of generalized maximum entropy estimation approach (Golan et al. 1996) is used to fit the general linear measurement error model. A Monte Carlo comparison is made with the classical maximum likelihood estimation (MLE) method. The results showed that, the GME is outperformed the MLE estimators in terms of mean squared error. A real data analysis is also presented.

Key words: Measurement Error Model, Generalized Maximum Entropy, Maximum Likelihood

1. Introduction

The traditional maximum entropy formulation is based on the entropy-information measure of Shannon (1948). It is developed and described in Jaynes (1957a, 1957b). Shannon's entropy measure reflects the uncertainty about the occurrence of a collection of events. Suppose we have a set of events $\{x_1,x_2,\ldots, x_k\}$ whose probabilities of occurrence are $p_1,p_2,\ldots,p_k$. then using an axiomatic method to define a unique function to measure the uncertainty of a collection of events, Shannon (1948) defines the entropy of the distribution (discrete events) as the average of self-information $H(P) = -\sum_{i=1}^{k} p_i \ln(p_i)$, where $0\ln(0) = 0$.

Since the 1990's many attempts have been made to apply the method of maximum entropy in the area of linear models. In 1996, Golan et al. proposed an estimator based on the maximum entropy formalism of Jaynes that they called the generalized maximum entropy (GME) estimator. The logic of using GME estimation method is that GME tends to dominate traditional estimation methods with small samples, and it does not rely on any distributional assumption, also it is robust in fitting nonlinear models (Golan, 2008; Peeter, 2004). The idea underling the GME approach is to view each unknown parameter and error terms as an expected value of some proper probability distribution; then by maximizing

the joint entropies subject to the data, represented by each unobserved value, and the requirement for proper probability distributions; better estimates can achieved with less assumptions (Csiszar (1991), Donho et al. (1992), Golan, et al. (1997), Al-Nasser (2003), Al-Nasser (2004) and Al-Nasser et al (2006)).

The remainder of this paper is divided into six sections. Section 2 presents the Ultrastructural model. Section 3 presents the generalized maximum entropy estimation approach idea in general linear model. Section 4 introduces the GME to ultrastructural model, Section 5 gives a real data analysis to study the relationships between number of crimes and number of unemployed citizens in Jordan , Section 6 presents Monte Carlo evidence on the numerical performance of GME and maximum likelihood estimation (MLE), and Section 6 presents concluding comments.

amjadn@yu.edu.jo
Department of Statistics, Science faculty
Yarmouk University, 21163 Irbid
Jordan

2. The Model

Consider the simple linear relationship between two mathematical variables ξ and η

$$\eta = \alpha + \beta\xi$$

where α is the intercept and β is the slope. The classical theory of regression analysis assumes that these variables are measured without error; particularly in the social sciences and natural science this assumption is often violated. Hence, this linear relationship is reformulated such that both variables are contaminated with measurement errors. Then, the observed values can be defined by:

$$\begin{aligned} x_i &= \xi_i + \delta_i , \\ y_i &= \eta_i + \varepsilon_i , \quad i = 1,2,...,n \end{aligned} \tag{1}$$

where δ and ε are the measurement errors associated with ξ and η, respectively. It is assumed that $\delta_1, \delta_2, ..., \delta_n$ are identically and independently distributed (*i.i.d*) with mean 0 and variance σ_δ^2. Similarly, $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ are *i.i.d* with mean 0 and variance σ_ε^2. Further, suppose that the true values of the variable (ξ) have possibly different means; say $\mu_1, \mu_2, ..., \mu_n$, so that we can write

$$\xi_i = \mu_i + \omega_i , \quad i = 1,2,...,n \tag{2}$$

where $\omega_1, \omega_2, ..., \omega_n$ are *i.i.d* with mean 0 and variance σ_ω^2. Finally, assume that the distribution of $(\delta_1, \delta_2, ..., \delta_n)$, $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ and $(\omega_1, \omega_2, ..., \omega_n)$ are mutually independent of each other. This provides the specification of ultrastructural model.

This model can be reduced to the functional measurement error model if we assumed that ξ_i are fixed; i.e. $\omega_i = 0$, $i = 1,2,...n$. Also it can be reduced to the structural measurement model if we assumed $\mu_i = \mu_j, \forall i, j = 1,2,..., n$. On the other hand, if $\delta_i = 0$, $i = 1,2,...n$ then the ultrastructural model reduces to the classical regression model with no measurement error; for more details see, Dolby (1976) and Cheng and Van Ness (1999).

Assuming Normal measurement errors, then the MLE for the parameters of the ultrastructural model are unidentifiable, Shalabh (1997). Hence, to solve the ultrastructural model (1-2), the MLE method required additional assumptions; Dolby (1976) derived the MLE estimates when the ratios $\lambda = \dfrac{\sigma_\varepsilon^2}{\sigma_\delta^2}$, and $\nu = \dfrac{\sigma_\omega^2}{\sigma_\delta^2}$ are known. Then, under the customary assumptions, the MLE for the ultrastructural model are the results in a quintic equation for $\hat{\beta}$;

$$h(\hat{\beta}) = \nu S_{xx}\hat{\beta}^5 + \left(3\nu\lambda S_{xx} - \nu S_{yy}\right)\hat{\beta}^3 - 2\lambda(\nu - 1)S_{xy}\hat{\beta}^2 + \left\{2\lambda^2(\nu + 1)S_{xx} - \lambda(\nu + 2)S_{xx}\right\}\hat{\beta} - 2\lambda^2(\nu + 1)S_{xy} = 0$$

This generally should be solved by iterative numerical methods. However, Gleser (1985) showed that the likelihood is maximized on the $\sigma_\omega^2 = 0$; boundary of the parameter set. Thus, the ML estimates for the ultrastructural model are just the ML estimates of the functional model; which leads to

$$\hat{\beta} = \frac{(S_{yy} - \lambda S_{xx}) + ((S_{yy} - \lambda S_{xx})^2 + 4\lambda S_{xy}^2)^{1/2}}{2S_{xy}}$$

The ML solution of the other parameters will be:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \quad \hat{\mu}_i = \frac{\lambda x_i + \hat{\beta}(y_i - \hat{\alpha})}{\lambda + \hat{\beta}^2}$$

where

$S_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2 \Big/ n$, $S_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2 \Big/ n$, $S_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \Big/ n$, $\bar{x}=\sum_{i=1}^{n} x_i \Big/ n$ and $\bar{y}=\sum_{i=1}^{n} y_i \Big/ n$

For more details see Cheng and Van Ness (1999), Carroll et al. (1995).

3. Generalized Maximum Entropy

Consider the general linear model; $y_i = f(x_i, \beta) + \varepsilon_i$; i = 1,2,…n. where $\beta = (\beta_1, \beta_2, ..., \beta_K)'$ is the vector of parameters to be estimated, the regressor variable x_i, $i = 1,2,...n$ are *K*-dimensional vectors whose values are assumed known and ε_i, $i = 1,2,...,n$ is the random error.

In GME, the model is fitted after some reformulation of the unknown parameters β, and the unknown error term ε_i, i = 1,2,3,…, n; if they are not in probability format. This can be done by reparameterize their possible outcome values probabilistically as a convex combination of random variables. This combination is presented as expected value of some proper probability distribution. For each unknown, assume that there exists a discrete probability distribution that defined over the parameter space [0, 1]; by a set of equally distanced discrete points; then the formulation of model parameter will be of the form $\beta = ZP$; where Z is a ($K \times KR$) matrix and P is a KR-vector of weights such that $p_k > 0$ and $P_k^{'} 1_R = 1$ for each k. Simply, each β_k, $k = 1,2,...,K$ can be defined by a set of equally distanced discrete points $Z_k' = [z_{k1}, z_{k2}, z_{k3}, ..., z_{kR}]$, where $R \geq 2$ with corresponding probabilities $P_k' = [p_{k1}, p_{k2}, p_{k3}, ..., p_{kR}] \in [0,1]$. That is,

$$\beta_k = \sum_{r=1}^{R} z_{kr} p_{kr} \ , \sum_{r=1}^{R} p_{kr} = 1, \ 0 \leq p_{kr} \leq 1, \ \text{k} = 1,2,\ldots,\text{K}$$

The support points Z for β are known. Golan et al. (1996) suggested that these values can be specified uniformly symmetric around *0* with large value of the lower and the upper bounds. For example, one can select the support point $z = (-c, 0, c)$, given that c is a large value (i.e. c = 100 or 1000 etc.). Moreover, assuming one specify Z to span the true values of β then the GME is a consistent estimator, which is an advantage of this method.

The disturbance ε_i can be treated in similar fashion. For a set $T \geq 2$ support points ε_i is assumed to be bound between two finite values v_t, v_T, which are symmetric around zero with corresponding unknown probability weights w_{1t}, w_{nT}.

That is, each error term may be modeled as; $\boldsymbol{\varepsilon} = VW$, where V is a $(n \times nT)$ matrix and W is a nT-dimensional vector of weights; that is to say:

$$\varepsilon_i = \sum_{j=1}^{T} v_{ij} w_{ij} \quad , \quad \sum_{j=1}^{T} w_{ij} = 1 \, , 0 \le w_{ij} \le 1 \, , \text{ i= 1,2,…,n}$$

Placing bounds for v_j is difficult in practice. Alternatively, Chebychev's inequality Pukelsheim (1994) may be used as a conservative means of specifying sets of error bounds. The empirical GME literature indicates that, in general, the number of support values R for unknown parameters is 5 and the number of support values of T of the error term is 3; see for example Paris (2001), Al-Nasser (2003), Al-Nasser (2005) and Golan et al. (1996).

Now, using the reparameterized unknowns' $\beta = ZP$ and $\varepsilon = VW$, we rewrite the general linear model as follows:

$$y = f(x_i, ZP) + VW$$

Then, maximum entropy principle may be stated in scalar summations with two nonnegative probability components, and the GME estimators can be achieved by solving the following non-linear programming problem:

$$\left.\begin{array}{l} Max \;\; H(P,W) = -P' \ln(P) - W' \ln(W) \\ \text{subject to the following constraints;} \\ \text{(i)} \;\; y = f(x, ZP) + VW \\ \text{(ii)} \quad 1_K = (I_K \otimes 1_R')P \\ \text{(iii)} \quad 1_n = (I_n \otimes 1_J')W \end{array}\right\} \quad (3)$$

Note that $\otimes$ is the Kronecker product, 1_K is a K-dimensional vector of ones, and $\ln(P) = (\ln(p_{11}), \ln(p_{12}), ..., \ln(p_{KR}))$. The GME system in (3) is a non-linear programming system and can be solved by applying the Lagrangian method, in which after finding the Lagrangian function, the first order conditions can be solved . For more details see Golan et al. (1996).

3.1.GME Procedure to ME Model

Without losing of generality, assume that ξ_i are fixed; i.e. $\omega_i = 0, \;\; i = 1,2,...n$. The compound model of ultrastructural model given in (1-2) can be rewritten as follows:

$$y_i = \alpha + \beta(x_i - \delta_i) + \varepsilon_i \quad , i = 1,2,..., n \qquad (4)$$

Then, by using the GME we can solve the problem after some reformulation of the unknown parameters α and β, and the unknown error terms δ_i, and ε_i, i = 1,2,3,…, n. In the compound model (4), there are two unknown parameters reparametrized as

$$\alpha = AQ \text{ ; where } 1_R^{'} Q = 1$$

where A is a row vector of size R and Q is a R-dimensional column vector of weights. The slope will be in the form

$$\beta = ZP \text{ ; where } 1_K^{'} P = 1$$

where Z is a row vector of size K and P is a K-dimensional column vector of weights. Also, there are two error terms that are reparametrized in following terms

$$\delta = V^* W^* \text{ where } \left(I_n \otimes 1_T^{'}\right) W^* = 1_n$$

$$\varepsilon = VW \text{ ; where } \left(I_n \otimes 1_J^{'}\right) W = 1_n$$

noting that V^* and V are ($n \times nT$) and ($n \times nJ$) matrices respectively, and W^* and W are nT-dimensional and nJ-dimensional vectors of weights respectively. Based on these reparametrization, the new ultrastructural model can be rewritten as

$$Y = AQ + X\left(ZP + V^* W^*\right) + VW$$

3.2. GME Solution

Generalized Maximum Entropy (GME) for the model given in (4) may be formulated as nonlinear programming (NP) system. The NP selects q, p, w^* and w ≥ 0 to maximize the joint Shannon entropy:

$$H(Q,P,W,W^*) = -Q^{'} \ln(Q) - P^{'} \ln(P) - W^{'} \ln(W) - W^{*'} \ln(W^*)$$

Subject to

$$Y = AQ + X\left(ZP + V^* W^*\right) + VW$$

$$1_R^{'} Q = 1$$

$$1_K^{'} P = 1$$

$$\left(I_n \otimes 1_T^{'}\right) W^* = 1_n$$

$$\left(I_n \otimes 1_J^{'}\right) W = 1_n$$

Here, we have $3n+2$ constraints and $R+K+n(T+J)$ unknowns. The solution of this system can be found by deriving the first order conditions of the Lagrangian function:

$$L = H(Q,P,W,W^*) + \gamma'[Y - AQ - X(ZP + V^*W^*) \cdot$$

$$- VW] + \theta_1'[1 - 1_R'Q] + \theta_2'[1 - 1_K'P]$$

$$+ \psi'[(I_n \otimes 1_T')W^* + \zeta'[(I_n \otimes 1_J')W$$

where, $\gamma' \in \Re^n, \theta_1 \in \Re^R, \theta_2 \in \Re^K, \psi \in \Re^n, and\ \zeta \in \Re^n$ are the associated vectors of Lagrangian multipliers. Taking the gradient of L to derive the first order condition and solving these conditions, the GME solution selects the most uniform distribution consistent with the information provided in the data and the add up constraints:

$$\hat{Q} = \exp(-A1_n'\hat{\gamma}) \odot \left\{1_R' \exp(-A1_n'\hat{\gamma})\right\}^{-1}$$

$$\hat{P} = \exp(-Z1_n'\hat{\gamma}(X - V^*\hat{W}^*)) \odot 1_K'\left\{\exp(-Z1_n'\hat{\gamma}(X - V^*\hat{W}^*))\right\}^{-1}$$

$$\hat{W} = \exp(-V'\hat{\gamma}) \odot \left\{(I_n \otimes 1_J 1_J') \exp(-V'\hat{\gamma})\right\}^{-1}$$

$$\hat{W}^* = \exp(-V^{*'}\hat{\gamma}1_K'Z\hat{P}) \odot \left\{(I_n \otimes 1_T 1_T') \exp(-V^{*'}\hat{\gamma}1_K'Z\hat{P})\right\}^{-1}$$

where $\odot$ is the Hadamard (element wise) product. To simplify matter the individual probabilities take the forms:

$$\hat{q}_r = \frac{\exp\left(-a_r \sum_{i=1}^{n} \hat{\gamma}_i\right)}{\sum_{r=1}^{R} \exp\left(-a_r \sum_{i=1}^{n} \hat{\gamma}_i\right)}, r = 1,2,...,R;$$

$$\hat{p}_k = \frac{\exp\left(-z_k \sum_{i=1}^{n} \hat{\gamma}_i \left(x_i - \sum_{t=1}^{T} v_t^* \hat{w}_{it}^*\right)\right)}{\sum_{k=1}^{K} \exp\left(-z_k \sum_{i=1}^{n} \hat{\gamma}_i \left(x_i - \sum_{t=1}^{T} v_t^* \hat{w}_{it}^*\right)\right)}, k = 1,2,3,...,K$$

$$\hat{w}_{ij} = \frac{\exp(-\hat{\gamma}_i v_j)}{\sum_{j=1}^{J} \exp(-\hat{\gamma}_i v_j)} \quad ,i = 1,2,3,\ldots,n \quad ,j = 1,2,3,\ldots,J;$$

$$\hat{w}^*_{it} = \frac{\exp\left(-\hat{\gamma}_i v^*_t \sum_{k=1}^{K} z_k \hat{p}_k\right)}{\sum_{t=1}^{T} \exp\left(-\hat{\gamma}_i v^*_t \sum_{k=1}^{K} z_k \hat{p}_k\right)} \quad ,i = 1,2,3,\ldots,n \quad ,t = 1,2,3,\ldots,T$$

Then the intercept and the slope of model (4) can be estimated as

$$\left.\begin{aligned} \hat{\alpha} &= A\hat{Q} \\ \hat{\beta} &= Z\hat{P} \end{aligned}\right\} \quad (5)$$

Lemma.1 The GME estimators given in (5) are unbiased estimators.

Proof: Simply, by taking the expected value of $\hat{\alpha}$, we have;

$$E(\hat{\alpha}) = E(\sum_r a_r \hat{p}_r) = E(a)\sum_r \hat{p}_r$$

since $\sum_r \hat{p}_r = 1$, then

$$E(a) = \sum_r a_r p_r = \alpha$$

Therefore, $\hat{\alpha}$ is an unbiased estimator of the intercept. Consequently, the estimated variance of $\hat{\alpha}$ is:

$$Var(\hat{\alpha}) = Var\left(\sum_r a_r \hat{p}_r\right) = \sum_{r=1}^{R} \hat{p}_r^2 \,\mathrm{var}(a_r)$$

$$= \sum \hat{p}_r^2 \left[\left(\sum a_r^2 \hat{p}_r\right) - \left(\sum a_r \hat{p}_r\right)^2\right]$$

$$= \sum \hat{p}_r^2 \left[\left(\sum a_r^2 \frac{\exp\left(-\hat{\gamma}_r a_r\right)}{\sum_{r=1}^{R} \exp\left(-\hat{\gamma}_r a_r\right)} \right) - \left(\sum a_r \frac{\exp\left(-\hat{\gamma}_r a_r\right)}{\sum_{r=1}^{R} \exp\left(-\hat{\gamma}_r a_r\right)} \right)^2 \right]$$

In similar ways we can prove that $\hat{\beta}$ is an unbiased estimator of the slope, and consequently we can derive the associated variance.

4. Monte Carlo Experiments

Monte Carlo experiments were performed to comment on the choice of the support points and number of support points of the unknown parameters and error terms in GME formulations. For fixed sample size, $n = 20$, a simulation study was carried out by generating 1000 samples according to the ultrastructural relationship $y_i = 1 + x_i + \varepsilon_i$ and $x_i = \frac{2+i}{2} + \delta_i$, $i = 1,2,...,n$. The errors terms generated independently from standard normal, i.e., $\varepsilon \sim N(0,1)$ and $\delta \sim N(0,1)$. Then, a comparison between the GME and MLE estimation methods was made in terms of bias and means squared error (*MSE*)

$$MSE(\hat{\beta}) = \frac{1}{1000} \sum_{i=1}^{1000} \left(\hat{\beta}_i - \beta\right)^2 \text{ and } Bias(\hat{\beta}) = \frac{1}{1000} \sum_{i=1}^{1000} \left(\hat{\beta}_i - \beta\right)$$

Three experiments were conducted:

Experiment 1.

The support parameter space of the error terms, V^* and V were fixed to be three in the intervals [-3S_x, 0, 3S_x] and [-3S_y, 0, 3S_y] for δ and ε; respectively. Then, this experiment was conducted for selecting the support values (a, z) and number of support values (R, K) for the unknown parameters $\alpha = AQ$ and $\beta = ZP$; i.e.,

$$\{a_i, i = 1,2,...,R\};\ \{z_j, j = 1,2,...,K\}$$

In the first part of this experiment, we fixed the number of these support values to be 3 in the intervals [-1, 0, 1] to [-500, 0, 500]. The results of this simulation study in Table.1 indicate that the best support values for both parameters should be in the interval [-100, 0, 100]. Hereafter, in the second part of this experiment, we start to increase number of support values within this interval to have 4, 5… or 7 support points and allocate them in an equidistant fashion. The results in Table.1 indicate that the greatest improvement in precision comes when

R and *K* were equal to 7. Moreover, it could be noted that for all choices of the parameter spaces, the GME results were more accurate and more efficient than the MLE results.

Table 1. Selecting the parameters support

Method		$\hat{\alpha}$		$\hat{\beta}$	
		Bias	*MSE*	*Bias*	*MSE*
MLE		-0.1465	0.6190	-0.0903	0.0549
GME	[-1,0,1]	-0.0906	0.0824	-0.0546	0.0323
	[-10,0,10]	-0.0765	0.0599	-0.0445	0.0296
	[-100,0,100]	-0.0579	0.0386	-0.0471	0.0311
	[-500,0,500]	-0.0647	0.0465	-0.0450	0.0294
	Increasing number of the support points				
	[-100,-50,50,100]	-0.0608	0.0405	-0.0464	0.0308
	[-100,-50,0,50,100]	-0.0744	0.0582	-0.0443	0.0291
	[-100,-50,-25,25,50,100]	-0.0588	0.0393	-0.0466	0.0307
	[-100,-50,-25,0,25,50,100]	-0.0544	0.0354	-0.0436	0.0278

Experiment 2.

To check the impact of error terms support space, and based on the experimental design outlines above, the Monte Carlo trial were repeated under the results of Experiment.1. The number of support values for each parameters *A* and *Z* were fixed to be 7 support values within the interval [-100, 0, 100]. Also, this experiment started by fixing the number of support values (*J*, *T*) to be 3, then the experiment repeated by shifting the support values V^* and V in the interval [-*hS*, 0, *hS*], where *h* = 1, 2, …, 7. Under the simulation assumptions, the results in Table.2 indicate that the greatest improvement in the precision comes from using *h* = 3. In similar ways of Experiment.1, the simulation is repeated by start increasing the number of support points in the interval [-*3S*, 0, *3S*] for each error term. Taking in our consideration both parameters, the greatest improvement comes when number of support points equals to 3.

Table 2. Selecting the error terms support

Method		$\hat{\alpha}$		$\hat{\beta}$	
		Bias	*MSE*	*Bias*	*MSE*
MLE		-0.1465	0.6190	-0.0903	0.0549
GME	[-S, 0, S]	-0.0540	0.0361	-0.0445	0.0282
	[-2S, 0,2 S]	-0.0534	0.0337	-0.0437	0.0279
	[-3S, 0, 3S]	-0.0544	0.0354	-0.0436	0.0278
	[-4S, 0, 4S]	-0.0557	0.0384	-0.0470	0.0311
	[-5S, 0, 5S]	-0.0714	0.0541	-0.0449	0.0296
	Increasing number of the support points				
	[-3S, -1.5 S,1.5S, 3S]	-0.0819	0.0672	-0.0397	0.0254
	[-3S, -1.5 S,0,1.5S, 3S]	-0.0771	0.0633	-0.0405	0.0260
	[-3S, -1.5 S, -0.75S, 0.75S, 1.5S, 3S]	-0.0603	0.0432	-0.0426	0.0266
	[-3S, -1.5 S, -0.75S,0, 0.75S, 1.5S, 3S]	-0.0644	0.0467	-0.0366	0.0219

Experiment 3.

Based on the previous experiments results, which related to the choice of parameter support, this experiment starts to increase the sample size; n = 20, 30, …, 100. The simulation results in Table.3 showed that the GME estimators have a lower *MSE* and lower bias for all sample sizes.

Table 3. Comparisons between GME and MLE

n	Method	$\hat{\alpha}$		$\hat{\beta}$	
		Bias	*MSE*	*Bias*	*MSE*
20	GME	-0.0544	0.0354	-0.0436	0.0278
	MLE	-0.1465	0.6190	-0.0903	0.0549
30	GME	-0.0488	0.0313	-0.0449	0.0284
	MLE	-0.1303	0.6721	-0.0896	0.0501
40	GME	-0.0396	0.0218	-0.0398	0.0240
	MLE	-0.0769	0.4213	-0.0798	0.0481
50	GME	-0.0411	0.0211	-0.0382	0.0239
	MLE	-0.0818	0.4186	-0.0764	0.0397
100	GME	-0.0375	0.0191	-0.0394	0.0234
	MLE	-0.0734	0.3381	-0.0789	0.0268

5. Empirical Example

To illustrate the computation of estimates by using the GME and MLE methods, a small data set adapted from department of statistics (2007) is used. The data given in Table.4 are number of crimes and number of unemployed collected at twelve main directorates in Jordan. The estimates of both variables contain measurement error arising from the observational and sampling error associated with the use of the register cases only (crimes or unemployed) to represent the population.

Table 4. Number of Unemployed and Number of General Crimes in Jordan by Location of Occurrence, 2007

Police Directorate	Population Size	Number of Crimes	Number of unemployed
Amman	2216000	20166	243760
Balqa	383400	2261	52526
Zarqa	852700	5649	113409
Madaba	143100	790	26617
Irbid	1018700	6296	197628
Mafraq	269000	949	43578
Jarash	171700	755	30563
Ajlun	131600	713	27504
Karak	223200	1199	29016
Tafiela	80100	405	13857
Ma'an	108800	805	21216
Aqaba	124700	1745	18082

Source: Statistical Yearbook, Amman, Jordan, 2007.

We believe that the number of crime is a linear function of the number of unemployed, and based on Figure.1 the suggested model for this data is

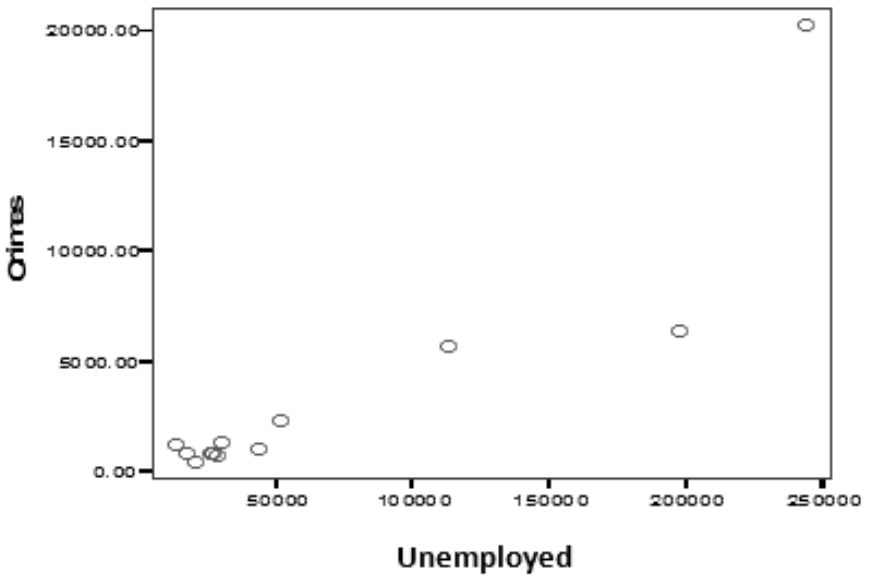
$$\eta_i = \alpha + \beta\xi_i \quad , i = 1,2,...,12$$

such that

$$x_i = \xi_i + \delta_i \,,$$
$$y_i = \eta_i + \varepsilon_i \,, \quad i = 1,2,...,12$$

where η is the number of crimes and ξ is the number of unemployed.

Figure 1. Scatter plot of number of Crimes vs Number of Unemployed



The statistics associated with the data of Table.4 are $(\bar{x}, \bar{y})$ = (68146.26, 3437.5833); (S_{xx}, S_{yy}, S_{xy}) = (5.373197E+09, 2.884094E+07, 3.557909E+08).

This model is solved by using both approaches (GME and MLE). While using the GME method we carried out all numerical calculations parallel to the theoretical development. The number of support values for each parameter were fixed to be 7 support values within the interval [-100, 0, 100]; however, the support parameter space of the error terms were fixed to be three in the intervals [$-3S_x$, 0, $3S_x$] and [$-3S_y$, 0, $3S_y$] for δ and ε, respectively. The results are given in Table.10

Table 5. MLE and GME estimates of the number of crimes data

Method	$\hat{\alpha}$	$\hat{\beta}$	Total Residual sum of squares (TRSS)
MLE	-1039.039	0.0663	-15.75
GME	-2656.31	0.0900	10.68

Figure 2. Comparison between the observed residuals of both methods

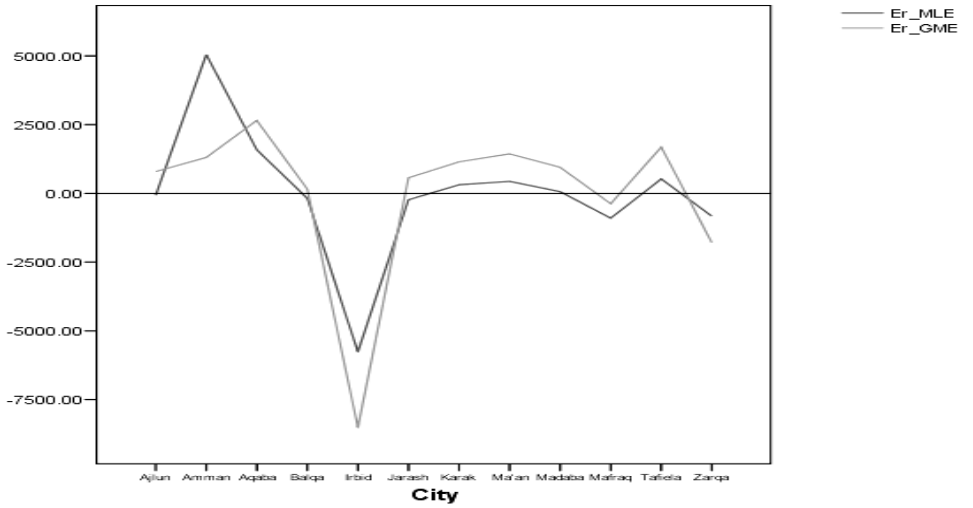


Figure.2 depicts a distinction between the observed errors by using both methods. The TRSS is -15.75 and 10.68 for the MLE and GME estimation methods, respectively; this appears to support that the GME is a good alternative estimation method to the MLE in fitting the measurement error models.

6. Concluding Remarks

This study gives the researcher a more precise method for estimating the parameters of the ultrastructural model by applying the GME estimation approach. The main idea of using GME is to improve the parameter estimation in the generalized measurement error models and to abstract the additional assumptions that are needed in the traditional MLE. In fact, all that the GME needs to be applicable can be obtained from the sample or can be specified by the researcher experiences. The Monte Carlo simulations provide a good evidence for the superiority of the GME on the MLE in terms of *MSE*. Also, the real data analysis gives similar results that confirm the robustness of GME over the MLE estimation method. Hence, the GME considered a good alternative in estimating the ultrastructural models.

REFERENCES

AL-NASSER, A. 2005. Entropy Type Estimator to Simple Linear Measurement Error Models. *Austrian Journal of Statistics*. 34(3). 283–294.

AL-NASSER, A. 2004. Estimation of Multiple Linear Functional Relationships. *Journal of Modern Applied Statistical Methods*.3 (1), 181–186.

AL-NASSER, A. 2003. Customer Satisfaction Measurement Models: Generalized Maximum Entropy Approach. *Pakistan Journal of Statistics*. 19(2), 213–226.

CARROLL, R. J., RUPPERT, D. and STEFANSKI, L. A. 1995. *Measurement Error in Nonlinear Models*. Chapman and Hall, London.

CHI-LUN CHENG and JOHN W. VAN NESS. 1999. Statistical Regression with Measurement Error. Arlond: N.Y: USA.

CSISZAR, I. 1991. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19, 2032–2066.

Department of Statistics. 2007. Statistical Yearbook, 58th Issue, Amman, Jordan.

DOLBY, G. R., 1976, The ultra-structural model: A synthesis of the functional and structural relations, *Biometrika* 63, 39–50.

DONHO, D. L, JOHNSTONE, I.M, HOCH, J.C, and STERN A.S 1992. Maximum entropy and nearly black object. *J. Royal, Statistical Society, Ser B*, 54, 41–81.

GOLAN, A. (2008). Information and Entropy Econometrics – A Review and Synthesis. Foundations and Trends in Econometrics. 2, 1–2, 1–145.

GOLAN, A. JUDGE, G. MILLER, D.1996. *A maximum Entropy Econometrics: Robust Estimation with limited data*, Wiley, New York.

GOLAN, A. JUDGE, G. PERLOFF, J. 1997. Estimation and Inference with Censored and Ordered Multinomial Response Data. *J. Econometrics*. 79, 23–51.

GLESER, L. J .1985. A note on G. R. Dolby's unreplicated ultrastructural model. *Biometrika*, 72, 117–124.

JAYNES, E. T. 1957(a,b). Information and Statistical Mechanics (I, II). *Physics Review* (106,108), (620–630, 171–190).

QUIRINO PARIS. 2001. Multicollinearity and Maximum Entropy Estimators. *Economics Bulletin*. Vol.3, No.11, 1–9.

PEETERS, L. (2004). Estimating a random-coefficients sample-selection model using generalized maximum entropy. Economics Letters. 84: 87–92

PUKELSHEIM, F. 1994. The Three Sigma Rule. *The American Statistician*, Vol.48, no.2, 88–91.

SRIVASTAVA, A. K, SHALABH. 1997. Consistent estimation for the non-normal ultrastructural model, *Statist. Probab. Lett.* 34 .67–73.

SHANNON C, E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*. 27, 379–423.