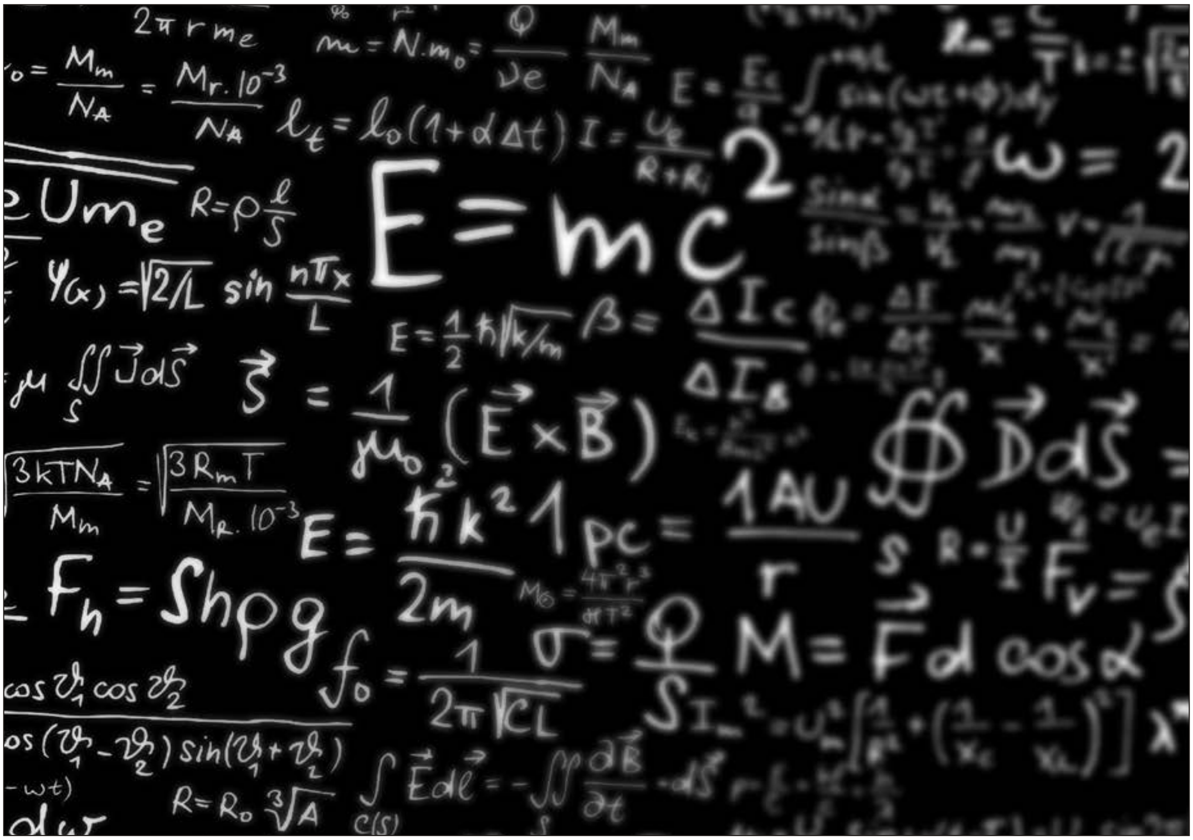

Asian Journal of Mathematics & Statistics

ISSN 1994-5418

ANSInet
an open access publisher
http://ansinet.com

science alert

Asian Journal of Mathematics and Statistics 3 (2): 111-118, 2010
ISSN 1994-5418

Multilevel Linear Models Analysis using Generalized Maximum Entropy

A.D. Al-Nasser, O.M. Eidous and L.M. Mohaidat
Department of Statistics, Faculty of Science, Yarmouk University,
21163, Irbid, Jordan

Abstract: The aimed of this study was to introduce the general multilevel models and discusses the Generalized Maximum Entropy (GME) estimation method that may be used to fit such models. The proposed procedure is applied to the two-level data model. The GME estimates were compared with Goldstein's generalized least squares estimates. The comparisons are made by two criteria; the bias and the efficiency. We find that the estimates of two level's model were substantially and significantly biased using Goldstein's generalized least squares approach. However, the GME estimates are unbiased and consistent, we conclude that the GME approach is a recommended procedure to fit multilevel models. An application to a real data in education is also discussed.

Key words: Multilevel models, generalized maximum entropy, simulation, goldstein's generalized least squares

INTRODUCTION

Multilevel linear models or random coefficients models are a type of mixed model with hierarchical data in away that each group at the higher level is assumed to have different regression slopes as well as different intercepts for purposes of predicting an individual-level of the dependent variable. Random coefficients model is illustrated by Bryk and Raudenbush (1992), Goldstein (1987), Langford (1987) and Raudenbush *et al*. (2005). The two levels model can be expressed in two equations; level 1 and 2 as:

Level 1

$$y_{ij} = B_{0j}+B_{1j}\times X_{ij}+r_{ij} \qquad \begin{array}{l} i = 1,2,....,n_j \\ j = 1,2,....,n \end{array} \tag{1}$$

where, i refers to the level 1 unit and j refers to the level 2 units, y_{ij} is the response variable for level 1 unit i within level 2 unit j, B_{0j} represents random intercept for the level 2 unit j, B_{1j} represents random slope of variable X_i of unit j and r_{ij} represents the residual for unit i within unit j. Also, J is the largest number of levels and n_j is the jth level sample size.

Level 2

$$\begin{array}{l} B_{0j} = \gamma_{00} + \gamma_{01}\times W_j + U_{0j} \\ B_{1j} = \gamma_{10} + \gamma_{11}\times W_j + U_{1j} \end{array} \tag{2}$$

Corresponding Author: A.D. Al-Nasser, Department of Statistics, Faculty of Science, Yarmouk University, 21163, Irbid, Jordan

In level 2 the parameters γ_{00} and γ_{10} are intercepts, γ_{01} and γ_{11} represent slopes predicting B_{0j} and B_{1j}, respectively from an outer variable W_j explanatory variable of level 2. Noting that, W_j should be in matrix form involves a (J+1) row vector of predictors in a block diagonal fashion. Moreover, U_{0j} and U_{1j} are level two random errors (random effects) that assumed to have zero means with an arbitrary variance covariance matrix.

The traditional estimation method used to estimate the parameters of model given in Eq. 1 and 2 is the iterative generalized least squares method; which is a sequential refinement procedure based on Ordinary Least Square (OLS) estimation. The method has been described in detail by Goldstein (1986). For known variance covariance matrix (V) of the level1 residual, then the GLS of the coefficients in level 1 is:

$$\hat{\beta} = (x'V^{-1}x)^{-1}x'V^{-1}Y$$

However, the GLS analysis to estimate level 2 parameters is:

$$\hat{\gamma} = (W'V^{*-1}W)^{-1}W'V^{*-1}Y^{*}$$

Given that, V^* and $V \otimes V$, where, vec is the vector operator and $\otimes$ is the Kronecker product.

GENERALIZED MAXIMUM ENTROPY

The traditional maximum entropy formulation is based on the entropy-information measure which reflects the uncertainty about the occurrence of a collection of events. Shannon (1948) defined the entropy of the distribution (discrete events $\{x_1, x2, ..., x_k\}$ whose probabilities of occurrences are $p_1, p_2, ..., p_k$), 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. Golan *et al.* (1996) proposed an estimator based on the maximum entropy formalism of Jaynes (1957) that they called the Generalized Maximum Entropy (GME) estimator. The idea underling the GME approach in the general linear model can be clarified by considering the following nonlinear relationships:

$$y_i = f(x_i, \beta);\ i = 1,2, \ldots n$$

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 unknown parameters are reparameterized as follow: $\beta = ZP$; 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 $P'_k = [p_{k1}, p_{k2}, p_{k3}, ..., p_{kR}]$. That is:

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

In similar fashion, the disturbance term may be rewritten as and ε = VW, where V is a (n×x nJ1) matrix and W is a nJ1-dimensional vector of weights; that is to say:

$$e_i = \sum_{j=1}^{J1} v_{ij} w_{ij},\ \sum_{j=1}^{J1} w_{ij} = 1,\ \sum_{j=1}^{J1} w_{ij} = 1,\ \ 0 \le w_{ij} \le 1,\ \ i = 1,2,....,n$$

The choice of Z should be uniformly and symmetrically around zero with equally spaced distance discrete points, for example Z = (-c,0,c), c large value. On the other hand, the actual bounds for v_i depend on the observed sample as well as any conceptual or empirical information about the underlying error. However; if such conceptual or empirical information does not exist, then v_i may be specified to be uniformly and symmetrically distributed around zero. Chebychev's inequality may be used as a conservative means of specifying sets of error bounds. For any random variable, X, such that E(X) = 0 and Var(X) = σ^2, the inequality provides, $P(|X|<d\sigma)\ge 1-1/d^2$, d>0; then the chebchyev's error bounds are $v_1 = d\sigma$ and $v_n = d\sigma$. One can use 3σ rules. However, the number of support points for each parameter, R and for the disturbance, J1, may be increased to reflect higher moments or more refined prior knowledge about β and ε, based on Al-Nasser (2003), Al-Nasser (2005), Ciavolino and Al-Nasser (2009) and Golan (2008) it appears that the greatest improvement in precision comes for using R and J1 to be 5 support points.

Now, using the reparameterized unknowns' β = ZP and ε = 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:

$$\text{Maximize } H(P,W) = P'ln(P)-W'ln(W)$$

Subject to:

$$\left.\begin{array}{ll}(i) & y = f(x_i ZP) + VW \\ (ii) & (I_K \otimes 1'_R)\,P = 1K \\ (iii) & (I_n \otimes 1'_J)\ W = 1_n\end{array}\right\} \quad (3)$$

Note that $\otimes$ is the Kronecker product, 1_K is a K-dimensional vector of ones and In(P) = (IN(p_{11}),In(p_{12}),...,In(p_{KR})). The GME system in Eq. 3 is a non-linear programming system that can be solved by applying the Lagrangian method, in which after finding the lagrangian function, the first order conditions are solved.

GENERALIZED MAXIMUM ENTROPY TO RANDOM COEFFICIENT MODEL

In order to estimate the two level random coefficient model by using GME method we rewrite Eq. 1 and 2 by one equation as:

$$y_{ij} = \gamma_{00} + \gamma_{01} * W_j + U_{0j} + (\gamma_{10} + \gamma_{11} * W_j + U_{1j}) * X_{ij} + r_{ij} \quad \begin{array}{l} i = 1,2,....,n_j \\ j = 1,2,....,J \end{array} \quad (4)$$

In the new general model Eq. 4, there are four unknown parameters which should be reparametrized by following the GME principles, that is mean each parameter will be rewritten as a convex combination of a discrete random variable in the following matrix form:

$$\gamma_{00} = AP, \text{ where, } 1'_R\, P = 1,\ \gamma_{10} = ZQ, \text{ where, } 1'_B\, Q = 1$$

$$\gamma_{01} = CN, \text{ where, } 1'_R\, N = 1,\ \gamma_{11} = ZQ, \text{ where, } 1'_B\, G = 1$$

Also, in this model; there are three error terms that should be reparametrized in a similar manner:

$U_0 = V^0T$, where, $1'n_j = (I\ n_j \otimes 1'_E)T$; $U_1 = V^1F$; where, $1'n_j = (I\ n_j \otimes 1'_D)F$ and $R = VO$; where, $1'n_j = (In_j \otimes 1'_M)O$;

Using these reparameterization expressions, then the model can be rewritten as:

$$Y = AP+XZQ+WCN+X(WDG+V^1F+V^0T)+VO$$

Therefore, the GME nonlinear programming system:

Maximize:

$$H(P, Q, N, G, T, F, O) = P'ln(P)-Q'ln(Q)-N'ln(N)-G'ln(G)-T'ln(T)F'ln(F)-O'ln(F)$$

Subject to:
(1) $Y = AP+XZQ+WCN+X(WDG+V^1F+V^0T)+V0$
(2) $1'_R\, P = 1$, (3) $1'_BQ = 1$, (4) $1'_KN = 1$ (5) $1'_SG = 1$
(6) $1'_{n_j} \otimes 1'_E)T$, (7) $1'_{n_j} \otimes 1'_D)F$, (8) $1'_{n_j} = (I_{n_j} \otimes 1'_M)0$

Note that $\otimes$ is the Kronecker product. Then this nonlinear programming system can be solved numerically. However, the final estimators will be obtained by the following formulas:

$$\hat{\gamma}_{00} = A\hat{P};\ \hat{\gamma}_{10} = Z\hat{Q};\ \hat{\gamma}_{01} = C\hat{N} \text{ and } \hat{\gamma}_{11} = D\hat{G}$$

SIMULATION STUDY

A simple Monte Carlo simulation study is considered to study the performance of the parameter estimation using GME. For the purposes of the simulation study we considered the following balanced random slope model:

$$\begin{aligned} y_{ij} &= \beta_j x_{ij} + \varepsilon_{ij} \qquad & i = 1, 2, ..., n \\ \beta_j &= \gamma_0 + \gamma_1 w_j + u_j \qquad & j = 1, 2, ..., J \end{aligned}$$

Then the compound model:

$$y_{ij} = (\gamma_0 + \gamma_1 w_j + u_j)x_{ij} + \varepsilon_{ij}$$

Table 1: Monte carlo comparisons between OLS and GME for Random coefficient model

	$\hat{\gamma}_0$			$\hat{\gamma}_1$		
Sample size	Bias (OLS)	Bias (GME)	Eff	Bias (OLS)	Bias (GME)	Eff
10	0.0287	0.0110	1.75	0.0162	0.0021	1.31
20	0.0281	0.0094	1.52	0.0123	0.0015	1.30
30	0.0270	0.0095	1.58	-0.0113	0.0015	1.19
50	0.0205	0.0082	1.62	0.0222	0.0010	1.24

Then the simulation study performed under the following assumptions:

- Generate 1000 random sample of size n = 10, 20, 30 and 50 and number of intercepts J = 2
- The error ~ N(0, 1), X~Exp 1 and set $\beta_1 = \beta_2 = 1$
- The error u ~ N(0, 1), W~U(0,1) and set $\gamma_0 = 1$, $\gamma_1 = 1.5$
- For GME estimator, we initial three support values for parameter in the interval [-10, 0, 10] and three support values for the error term selected in the interval [- 3 S, 0, 3S] where, S is the standard deviation of the dependent variable y
- The simulation results for estimating fixed effect parameters, are given in Table 1

The simulated bias and the efficiency are computed based on the following formulas:

$$\text{Bias} = \frac{\sum_{i=1}^{1000} (\hat{\gamma}_i - \gamma)}{1000} \quad \text{and} \quad \text{Eff} = \frac{\text{MSE(OLS)}}{\text{MSE(GME)}}$$

Under the simulation assumptions, the results in Table 1 indicate the superiority of GME estimation method over the OLS. It can be noted that, for all sample sizes the GME estimators are more accurate and more efficient than their counter part based on the OLS estimation method.

AN APPLICATION TO A REAL DATA

The data of the High School and Beyond (HSB) study is used. These data consist of a total sample of 7,185 students who are nested within 160 schools; 90 public and 70 private. Between 14 and 67 students were assessed from each school, with a median of number of 47 students assessed. The outcome of interest is a student-level measure of math achievement. The first predictor was a continuous measure of student socioeconomic status (SES). The second predictor was a dichotomous measure of school sector in which a value of 0 reflected a public school and a value of 1 reflected a private school (49% of schools were private). The final predictor was a continuous measure of disciplinary climate of the school in which higher values reflected greater disciplinary problems. Since we need the average intercept and average slope, then each school has its own regression model and this means the intercept and the slope vary within schools.

The Model

The random coefficient model with two levels is used to represents the relationships between the variables, where level-1 is the student's level and consists of two variables:

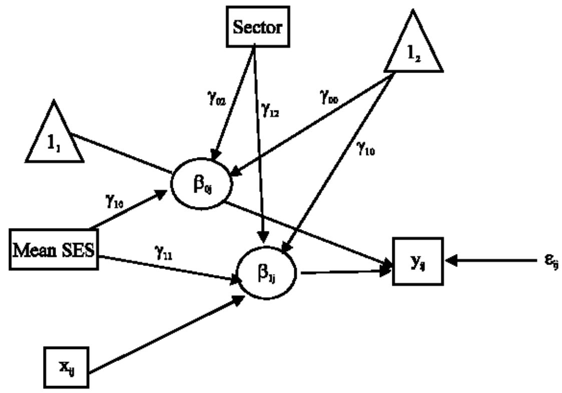

Fig. 1: Math achievement model

- **SES:** Socio-economic status
- **Mathach:** Math achievement

$$\text{Level}-1 \quad y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij} \tag{5}$$

where, y_{ij} is the outcome scores for student i in school j, x_{ij} are the values on the SES for student i in school j. Each school's distribution of math achievement is characterized by two parameters: the intercept, β_{0j} and the slope β_{1j}

The intercept and slope parameters, β_{0j} and β_{1j} are vary across schools in the level-2 model which consists of two variables:

- **Sector:** 1 = Private, 0 = Public
- **Mean SES:** Mean of the SES values for the students in this school who are included in the level-1

Then the school-level model can be written as:

$$\text{Level}-2 \quad \begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{Mean SES})_j + \gamma_{02}(\text{Sector})_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{Mean SES})_j + \gamma_{12}(\text{Sector})_j + u_{1j} \end{aligned} \tag{6}$$

Where:

γ_{00} = Overall intercept

γ_{01} = The main effect of Mean SES,

γ_{02} = The main effect of sector

γ_{10} = The main effect of SES

γ_{12} = Two cross level with interactions involving sector with student SES

γ_{11} = Mean SES with student SES. Moreover

u_{0j} and u_{1j} = Random errors

The path diagram for the model is shows in Fig. 1.

The unified equation model; which represents the combinations of the Eq. 5 and 6 can be written as:

Table 2: Results for educational research model: OLS and GME estimates

	OLS		GME	
Model	Coefficient	SE	Coefficient	SE
For School means				
Intercept ($\hat{\gamma}_{00}$)	12.10	0.20	12.63	0.21
Mean SES ($\hat{\gamma}_{01}$)	5.33	0.37	5.08	0.24
Sector ($\hat{\gamma}_{02}$)	1.23	0.31	3.11	0.24
For SES-achievement slopes				
Intercept ($\hat{\gamma}_{10}$)	2.94	0.16	2.94	0.11
Mean SES ($\hat{\gamma}_{11}$)	1.03	0.31	2.95	0.21
Sector ($\hat{\gamma}_{12}$)	-1.64	0.25	0.81	0.22

$$y_{ij} = \gamma_{00} + \gamma_{01}(\text{Mean SES})_j + \gamma_{02}(\text{Sector})_j + \gamma_{10}\, x_{ij} + \gamma_{11}(\text{Mean SES})_j x_{ij} + \gamma_{12}(\text{Sector})_j x_{ij} + u_{0j} + u_{1j}x_{ij} + \varepsilon_{ij} \quad (7)$$

We estimate γ_{01} to study whether high-SES differ from low-SES schools in means achievement (controlling for sector). Similarly, we estimate γ_{02} to learn whether private schools differ from public schools in terms of the mean achievement once Mean SES is controlled. These two estimates will clarify whether the Mean SES is significantly predicted the intercept or the school's slope, respectively. While by estimating γ_{11} we discover whether high-SES schools differ from low-SES schools in terms of the strength of association between student SES and achievement within them (controlling for sector). Also, we estimate γ_{12} to examine whether the private differ from the public schools in terms of the strength of association between student SES and achievement.

Now, in order to estimate the parameters by using GME estimation method; we need to reparametrize the unknowns and the error terms in Eq. 7. Then the GME system will be a nonlinear programming problem given as follows:

Maximize:

$$H(P_{00}, P_{01}, P_{02}, P_{10}, P_{11}, P_{12}, w_0, w_1, w) = -P'_{00}ln(P_{00}) - P'_{01}ln(P_{01}) - P'_{02}ln(P_{02}) - P'_{11}ln(P_{11}) - P'_{12}ln(P_{12}) - w'_0 ln(w_0) - w'_1 ln(w_1) - w''ln(w)$$

Subject to:

$$Y = Z_{00}P_{00} + (\text{MeanSES})\, Z_{01}P_{01} + \text{Sector}\, Z_{02}P_{02} + X(Z_{01}P_{01} + (\text{MeanSES})\, Z_{11}P_{11} + \text{Sector}\, Z_{12}P_{12} + v_1 w_1) + v_0 w_0 + vw$$

$$1'_R P_{00} = 1;\ 1'_R P_{01} = 1;\ 1'_R P_{00} = 1;\ 1'_R P_{02} = 1;\ 1'_R P_{11} = 1;\ 1'_R P_{12} = 1;$$
$$(I_n \otimes 1'_M)w = 1n;\ (I_n \otimes 1'_M)w_0 = 1n;\ (I_n \otimes 1'_M)w_1 = 1_n$$

Hereafter, we use the IMSL/Fortran in order to estimate the unknown parameters. The estimation results are shown in Table 2.

The estimations results indicate that both methods gave almost similar effect on the predictors, but the GME estimators have smaller standard error. In general, it can be noted that Mean SES is positively related to school mean math achievement, $\hat{\gamma}_{00}$ = 5.33 (0.37) (GLS) and 5.08 (0.24) (GME). Also, Private schools have higher mean achievement than public schools, controlling for the effect of Mean SES, $\hat{\gamma}_{02}$ = 1.23 (GLS) and 3.11 (GME). With regard to the slopes, there is a tendency for schools of high Mean-SES to have larger slopes than do schools with low Mean SES, $\hat{\gamma}_{11}$ = 1.03 (GLS) and 2.95 (GME). The only difference between the two methods is that the GLS estimates indicated that private schools have weaker SES slopes $\hat{\gamma}_{12}$ = -1.64, while the GME estimates indicated that private schools have positive SES slopes $\hat{\gamma}_{12}$ = 0.81.

CONCLUSION

This study proposed the GME estimation method in context of parameter estimation of random coefficient models. By comparing the GME estimates with their counterpart based on Goldstein's estimators, the simulation results demonstrated that GME estimates are superior and often closer to the parameter than the GLS estimates. For all sample sizes used in the simulation study, there is an advantage for using the GME estimator. Also, the real data analysis also supports the robustness of the GME estimation method since the GME estimators have smaller standard errors than the GLS estimators. The conclusions that are suggested by the analysis of the given example are that; the SES positively related to the Math achievement and the private schools have better achievement than the public schools. Moreover, even the GLS method has good advantages from the computational and interpretational point of view; it found that the GME method gives more precise estimates. Consequently, the GME estimator can be recommended as an alternative method for estimating the two level random coefficients parameters.

ACKNOWLEDGMENTS

Author would like to thank Luigi D'Ambra and Enrico Ciavolino the MTISD2008 Program Committee and a referee for their valuable suggestions which improved the contents of this study.

REFERENCES

Al-Nasser, A., 2003. Customer satisfaction measurement models: Generalized maximum entropy approach. Pak. J. Stat., 19: 213-226.

Al-Nasser, A.D., 2005. Entropy type estimator to simple linear measurement error models. Aust. J. Stat., 34: 283-294.

Bryk, A.S. and S.W. Raudenbush, 1992. Hierarchical Linear Model: Application and Data Analysis Method. Sage Publications, New Bury Park, CA.

Ciavolino, E. and A.D. Al-Nasser, 2009. Comparing generalized maximum entropy and partial least squares methods for structural equation models. J. Nonparam. Stat., 21: 1017-1036.

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

Golan, A., 2008. Information and entropy econometrics-a review and synthesis. Foundat. Trends Econ., 2: 1-145.

Goldstein, H.I., 1986. Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika, 73: 43-56.

Goldstein, H.I. 1987. Multilevel Model in Educational and Social Research. Oxford University Press, London.

Jaynes, E.T., 1957. Information and statistical mechanics(I,II). Phys. Rev., 108: 171-190.

Langford, N.T., 1987. A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. Biometrika, 74: 817-827.

Raudenbush, S., A. Bryk and R. Condon, 2005. HLM: Hierarchical Linear and Non-linear Modeling. Scientific Software International, Chicago, IL.

Shannon, C.E., 1948. A mathematical theory of communications. Bell Syst. Technol. J., 27: 379-423.