

Determination of Support Vector Boundaries in Generalized Maximum Entropy for Multilevel Models

Serpil Kılıç Depren¹ & Özer Depren²

¹ Department of Statistics, Yıldız Technical University, İstanbul, Turkey

² Customer Experience and Idea Management Department, Yapı Kredi Bank, İstanbul, Turkey

Correspondence: Serpil Kılıç Depren, Department of Statistics, Faculty of Arts and Science, Yıldız Technical University, Davutpasa Campus, 34220, Esenler-Istanbul, Turkey. E-mail: serkilig@yildiz.edu.tr

Received: August 9, 2016 Accepted: August 28, 2016 Online Published: October 17, 2016

doi:10.5539/ijsp.v5n6p10

URL: <http://dx.doi.org/10.5539/ijsp.v5n6p10>

Abstract

Generalized Maximum Entropy (GME) approach is one of the alternative estimation methods for Regression Analysis. GME approach is superior to other classical approaches in terms of parameter estimation accuracy when some or none of the assumptions of classical approaches are violated. However, determining bounds of parameter support vectors is one of the open parts of this approach when researchers have no prior information about the parameters. If support vectors cannot be determined correctly, parameters estimations will not be obtained correctly. There are some theoretical studies about GME for different datasets in the literature, but there are fewer studies about how to determine parameter support vectors. To obtain robust parameter estimations in GME, we introduced a new iterative procedure for determining parameter support vectors bounds for multilevel dataset. In this study, the new iterative procedure was applied for multi-level random intercept model and the new procedure was tested both simulation study and the real life data. The Classical and the new procedures of GME estimations were compared to Generalized Least Square Estimations in terms of Root Mean Square Error (RMSE) statistics. As a result, the estimations of the new approach provided lower RMSE values than classical methods.

Keywords: generalized maximum entropy, multilevel models, support vector bounds

1. Introduction

Shannon was defined the term "Entropy" as a measure of uncertainty in communication theory in 1948 and the basic principle of Generalized Maximum Entropy (GME) is based on Jaynes' Maximum Entropy Principle (Shannon, 1948; Jaynes, 1957). Golan, Judge and Miller generalized this principle for regression framework (Golan, Judge, & Miller, 1996). In this approach, Golan et al. maximized Shannon's entropy formula under model consistency constraints. GME approach requires fewer assumptions than the classical methods and it has been used an alternative approach for both classical linear and nonlinear estimation models (Golan, Judge, & Miller, 1996). The most important part of GME is re-parameterization of regression coefficients and the error term processes. The main topic of GME approach is determination of support vector boundaries when researchers have no prior information about the parameters. Thus, this approach has been receiving increasing attention in the statistics literature.

In this study, Random Intercept Model and GME Approach were combined. The main purpose of this study is to determine parameter support vector boundaries which allow to obtain more consistent parameters than classical methods for multilevel data. Therefore, the determination of error vector boundaries will not be taken into consideration in this study. Thus, the error vector boundaries were determined according to the literature by Pukelsheim (3σ rule). (Pukelsheim, 1994).

The outline of this paper organized as follows. The literature review was given in Section 2. We introduced Multilevel Modeling and briefly describe the GME estimation process in Section 3. In Section 4, we suggested the new procedure for determination of parameter support vector boundaries in Multilevel Random Intercept Models and we presented both real-life and simulation study. Also we compared the results obtained from different methods. Section 5 is a brief conclusion of the study.

2. Literature Review

In the literature, the term "Entropy" (and also known as "Maximum Entropy") has been widely used in many disciplines such as thermodynamics, communication, education and statistics. In previous studies, researches were proposed new

approaches which could be used rather than classical approaches and compared their results to different methods with simulation studies (Al-Nasser, 2014; Al-Rawwash & Al-Nasser, 2011). Also, there were several attempts to answer the question of “How parameter and the error term support vectors could be determined in the GME approach?” (Henderson, Golan & Seabold, 2015; Ciavolino & Calcagni, 2014; Golan & Gzly, 2012; Fernandez-Vazquez, Mayor-Fernandez & Rodriguez-Valez, 2008; Caputo & Paris, 2008), especially when researchers have no prior information about distribution of the parameters. The Maximum Entropy principle was also combined with Longitudinal Analysis, Data Envelopment Analysis, Multilevel Models, Regression Analysis and Logistic Regression Analysis in statistical area in order to compare GME results to the results of classical approach (Al-Nasser & Al-Atrash, 2011; Al-Nasser et al., 2010; Gastón & García-Vinas, 2011; Donoso et al., 2011). Furthermore, alternative solutions were proposed when data have the problem of multicollinearity and heteroscedasticity (Akdeniz et al., 2011).

Abellan, Baker and Coolen studied on the Nonparametric Predictive Inference (NPI) model for the multilevel data and proved that the lower and upper probabilities of the NPI could be obtained only by using the singleton probabilities (Abellan, Baker, & Coolen, 2011).

The performance of the GME estimator in both large and small samples was studied to assess the efficiency of the results in terms of estimation accuracy (Mittelhammer, Cardell, & Marsh, 2013; Gastón & García, 2011).

GME Approach was combined with Hierarchical Cumulative Logit Model in the study of Donoso, Grange and González in 2011. They compared the results of Hierarchical Cumulative Logit Model with Maximum Likelihood estimation using Monte Carlo simulations (Donoso, Grange, & González, 2011). In conclusion, they showed that the simulations produced reduced-bias in the estimates of the subjective value of time.

An alternative solution for the problem of multicollinearity was suggested by Akdeniz, Çabuk and Güler. Appropriate constraints were added to the classical Generalized Maximum Entropy Approach according to the characteristics of the relationship among independent variables and the results were compared with OLS (Akdeniz, Çabuk, & Güler, 2011).

The mathematical properties of Entropy, Maximum Entropy, Minimum Cross Entropy and Maximum Entropy Leuven Method were also explained in detail in Altaylıgil's study in 2008. The performance of the parameter estimations of this method was tested with OLS, Generalized OLS and Ridge by using Monte Carlo simulations (Altaylıgil, 2008).

In all of these studies, new algorithm or methodology comparison between classical methods and GME were provided, but there were not many study about how to determine parameter support vectors (especially bounds of support vectors), which is the missing part of this approach especially when researchers have no prior information about parameters distributions.

3. Methods

Multilevel Modelling is a generalization of linear regression models and such can be used for a variety of purposes, including prediction, data reduction and causal inference from clustered or hierarchical datasets (Gelman, 2006; Raudenbush et al., 2005; Raudenbush & Bryk, 2002). Also, GME, which is based on optimization technique, is one of the alternative solution approaches of Linear and Non-Linear Models. In this approach, all of the unknown parameters (β) and the error term (ϵ) are re-parameterized by using finite-dimensional known support vectors z and v (Golan, Judge & Miller, 1996). Similar to the coefficient re-parameterization process, ϵ could be re-parameterized as a finite and discrete random variable with $2 \leq J \leq \infty$ possible outcomes (Golan, Judge, & Miller, 1996). Adaptation of the re-parameterization process to Random Intercept Models is given in the next section.

3.1 Adaptation of GME Approach to Random Intercept Models

The study of Al-Nasser in 2010, GME approach was adapted to Multilevel Random Coefficient Model (Al-Nasser et al., 2010). In this study, GME approach was adapted to Random Intercept Model which was illustrated by Raudenbush and Bryk (2002) and Raudenbush et al. (2005) and a new iterative approach was proposed for determination of parameter support vectors for Random Intercept Model.

Two-level Random Intercept Model can be expressed as two equations.

$$\text{Level - 1: } y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij} \quad j = 1, 2, \dots, n \quad i = 1, 2, \dots, n \quad (1)$$

In Equation 1, i refer to level-1 units while j refers to level-2 units. Y_{ij} refers to response variable for level-1 unit i within level-2 unit j . β_{0j} is the random intercept for level-2 unit j while β_{1j} is the random slope of X_i of unit j . r_{ij} is the residual term for unit i within unit j .

$$\begin{aligned} \text{Level - 2: } \beta_{0j} &= \gamma_{00} + \gamma_{01}W_j + U_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}W_j \end{aligned} \quad (2)$$

In Equation 2, γ_{00} and γ_{10} are the intercepts. γ_{01} and γ_{11} represent slopes predicting β_{0j} and β_{1j} respectively. Furthermore, U_{0j} is the level-2 random errors. Finally, Equation 1 and 2 could be written as in Equation 3.

$$y_{ij} = \gamma_{00} + \gamma_{01}W_j + U_{0j} + (\gamma_{10} + \gamma_{11}W_j)X_{ij} + r_{ij} \tag{3}$$

In Equation 3, there are four unknown parameters and two error terms. They will be re-parameterized in order to use Generalized Maximum Entropy Approach (Al-Nasser et al. 2010).

$$\begin{aligned} \gamma_{00} &= \sum_{r=1}^R a_r p_r && \text{where } p_r \in (0,1) \text{ and } \sum_{r=1}^R p_r = 1 \\ \gamma_{10} &= \sum_{b=1}^B z_b q_b && \text{where } q_b \in (0,1) \text{ and } \sum_{b=1}^B q_b = 1 \\ \gamma_{01} &= \sum_{k=1}^K c_k N_k && \text{where } N_k \in (0,1) \text{ and } \sum_{k=1}^K N_k = 1 \\ \gamma_{11} &= \sum_{s=1}^S d_s G_s && \text{where } G_s \in (0,1) \text{ and } \sum_{s=1}^S G_s = 1 \\ U_{0j} &= \sum_{i=1}^E v_{ij} T_{ij} && \text{where } T_{ij} \in (0,1) \text{ and } \sum_{i=1}^E T_{ij} = 1 \\ r_{ij} &= \sum_{l=1}^M v_{lij}^* O_{lij} && \text{where } O_{lij} \in (0,1) \text{ and } \sum_{l=1}^M O_{lij} = 1 \end{aligned} \tag{4}$$

where $j=1,2,\dots,J$ and $I=1,2,\dots,n_j$.

After re-parameterization process, the model could be written as below.

$$\begin{aligned} y_{ij} &= \sum_{r=1}^R a_r p_r + (\sum_{b=1}^B z_b q_b)x_{ij} + (\sum_{k=1}^K c_k N_k)w_j + \\ &+ (\sum_{s=1}^S d_s G_s)w_j x_{ij} + \sum_{i=1}^E v_{ij} T_{ij} + \sum_{l=1}^M v_{lij}^* O_{lij} \end{aligned} \tag{5}$$

Therefore, the GME model for the Two-Level Random Intercept Model can be expressed by the following nonlinear programming system:

$$\text{Maximize } H(p, q, N, G, T, O) = -\sum p \ln(p) - \sum q \ln(q) - \sum N \ln(N) - \sum G \ln(G) - \sum T \ln(T) - \sum O \ln(O) \tag{6}$$

Subject to:

$$\begin{aligned} y_{ij} &= \sum_{r=1}^R a_r p_r + (\sum_{b=1}^B z_b q_b)x_{ij} + (\sum_{k=1}^K c_k N_k)w_j + (\sum_{s=1}^S d_s G_s)w_j x_{ij} + \\ &+ \sum_{i=1}^E v_{ij} T_{ij} + \sum_{l=1}^M v_{lij}^* O_{lij} \\ &\sum_{r=1}^R p_r = 1 \\ &\sum_{b=1}^B q_b = 1 \\ &\sum_{k=1}^K N_k = 1 \\ &\sum_{s=1}^S G_s = 1 \\ &\sum_{i=1}^E T_{ij} = 1 \\ &\sum_{l=1}^M O_{lij} = 1 \end{aligned} \tag{7}$$

The new optimization problem for Random Intercept Model is solved by using Lagrangian Method.

3.2 The Procedure for the Determination of Parameter Support Boundaries

In the literature, there are two main rules for determining support vector boundaries which are based on prior knowledge about parameters and the error term (Golan, Judge, & Miller, 1996).

- Support vector boundaries might be determined according to the prior information of parameters and the error term.
- They might also be determined extensively enough to include population parameters around zero and Pukelsheim's 3σ rule for the error term when the information about the parameters and the error term does not exist.

In the first rule, for example, if it is assumed that the mean, minimum and maximum values of β_1 are 1, 0.5 and 2, respectively; the parameter support vector should be $z_1=[0.5 \ 1 \ 2]$ for $M=3$ and $z_1=[0.5 \ 0.75 \ 1 \ 1.5 \ 2]$ for $M=5$. However, in the literature, it was not clearly identified how accurate support vector boundaries were determined when the researchers had no prior information about the parameters and the error term. As a result of these literature findings, the number of discrete points of parameters and the error term support vectors were selected as five (Golan, Judge, & Miller, 1996; Al-Nasser, 2011).

This study aimed to obtain prior information using the information of the current dataset. For this aim, the following steps, which were called as repeated sampling with replacement in Depren's study, were adopted for multilevel dataset (Depren, 2014):

1. A new sample (n and $2n$ sample size) was created by using repeated sampling with replacement (n is the total number of observation).

2. Restricted Maximum Likelihood approach was used to obtain prior parameters without checking whether the data met the required assumptions, such as multicollinearity, autocorrelation and homoscedasticity or not.
3. The first two steps were repeated (t) 50, 100, 500, 1000, 1500 and 2000 times to obtain 50, 100, 500, 1000, 1500 and 2000 different values (parameter matrix) in each repetition.
4. All the obtained parameter estimations were sorted in an ascending order.
5. %5 of the top and the bottom values of the parameters (outliers) were extracted from the parameter matrix.
6. The support vector boundaries of each parameter were determined according to Equation (8).

$$z' = \left[\text{TrimMean} - \frac{\sigma}{4} \quad \text{TrimMean} - \frac{\sigma}{8} \quad \text{TrimMean} \quad \text{TrimMean} + \frac{\sigma}{8} \quad \text{TrimMean} + \frac{\sigma}{4} \right] \quad (8)$$

Since sample size (n), repeats (t) and parameter support vectors (z) are the parameters of the procedure to be identified, the procedure is much more complex than standard GME or other estimation techniques. However, simple code can be written in SAS, R or in other package programs to run the procedure. Thus, researchers can overcome this complexity in business life.

The results obtained from classical and new approaches were compared by using RMSE (Miller, 2002; Timm, 2002).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2} \quad (9)$$

4. Application

The new procedure was tested in both simulated and real-life datasets. In this study, the following questions to be answered were which sample size should be chosen (n or 2n) and how many repeats (t) should be run for sampling with replacement procedure.

4.1 Background of Simulated Data

The simulation study was performed under following assumptions:

1. Generate 1000, 2000, 3000, 5000 and 10000 random sample of size n and 2n with repeated sampling with replacement technique.
2. The errors are distributed as $r_{ij} \sim Normal(0,1)$ and $U_{0j} \sim Normal(0,1)$
3. The variables are distributed as $X_{ij} \sim Normal(7,1)$ and $W_j \sim Uniform(0,1)$
4. The coefficients were set $\gamma_{00} = 1$, $\gamma_{01} = 2$, $\gamma_{10} = 1.5$ and $\gamma_{11} = 1$.
5. Five-point support vectors for parameters and the error term were used for GME estimator (Al-Nasser 2011).
6. For determination of parameter support vector boundaries, two different alternatives were tested.
 - i. Parameter support vector bounds were determined extensively enough to include the population parameters around zero and these bounds were narrowed down in every iteration for each sample.
 - 1.iteration $z' = [-1000 \ -500 \ 0 \ 500 \ 1000]$
 - 2.iteration $z' = [-100 \ -50 \ 0 \ 50 \ 100]$
 - 3.iteration $z' = [-10 \ -5 \ 0 \ 5 \ 10]$
 - ii. Parameter support vector boundaries were determined according process which was explained in Section 3.2.
7. The error support bounds were determined by Pukelsheim's 3σ rule. $e = [-3s \ -1.5s \ 0 \ 1.5s \ 3s]$ where s is the standard deviation of the dependent variable.
8. Simulation results were compared using RMSE (Miller 2002; Timm 2002).

4.2 Simulation Study

In order to identify n and t for sampling with the replacement procedure, Table 1 was prepared. Model outputs are shown by n and t values for all datasets.

Table 1. Outputs of the alternative approach for simulated dataset

	Intercept	Coefficient of X	Coefficient of W	RMSE
Sample Size: n	-3.83602	2.1975	9.0183	1.5374
t: 1000	(3.23E-06)	(4.46E-07)	(1.80E-07)	
Sample Size: n	-3.83586	2.19752	9.0178	1.5374
t: 2000	(3.02E-06)	(4.11E-07)	(1.83E-07)	
Sample Size: n	-3.83502	2.19736	9.0191	1.5374
t: 3000	(2.99E-06)	(4.13E-07)	(1.80E-07)	
Sample Size: n	-3.8361	2.1976	9.01768	1.5374
t: 5000	(3.08E-06)	(4.23E-07)	(1.85E-07)	
Sample Size: n	-3.83704	2.19772	9.01822	1.5374
t: 10000	(3.16E-06)	(4.27E-07)	(1.82E-07)	
Sample Size: 2n	-3.83604	2.19746	9.01914	1.5374
t: 1000	(1.49E-06)	(2.06E-07)	(9.33E-08)	
Sample Size: 2n	-3.83772	2.19772	9.01894	1.5374
t: 2000	(1.53E-06)	(2.08E-07)	(9.10E-08)	
Sample Size: 2n	-3.83542	2.19744	9.0178	1.5374
t: 3000	(1.49E-06)	(2.06E-07)	(8.94E-08)	
Sample Size: 2n	-3.83598	2.1975	9.01826	1.5374
t: 5000	(1.53E-06)	(2.10E-07)	(9.01E-08)	
Sample Size: 2n	-3.83736	2.1977	9.01804	1.5374
t: 10000	(1.58E-06)	(2.14E-07)	(8.78E-08)	

():Standard Deviation

Once the results of the different scenarios were compared according to RMSE, there was no significant difference for all scenarios in terms of β coefficients. For this reason, identifying n and t is not important for simulated dataset. In this study, first scenario (2n sample size and 1000 repeats) was chosen for further analysis.

The parameter support vectors, used for all datasets, were shown in Table 2. In this section, Support Vector I, II and III were used for Classical GME estimations and Support Vector IV was used for proposed GME estimation.

Table 2. Parameter support vectors used in analysis

Support Vectors	Boundaries
Support Vector I (Classical GME)	$\begin{bmatrix} Constant \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} -10 & -5 & 0 & 5 & 10 \\ -10 & -5 & 0 & 5 & 10 \\ -10 & -5 & 0 & 5 & 10 \\ -10 & -5 & 0 & 5 & 10 \end{bmatrix}$
Support Vector II (Classical GME)	$\begin{bmatrix} Constant \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} -100 & -50 & 0 & 50 & 100 \\ -100 & -50 & 0 & 50 & 100 \\ -100 & -50 & 0 & 50 & 100 \\ -100 & -50 & 0 & 50 & 100 \end{bmatrix}$
Support Vector III (Classical GME)	$\begin{bmatrix} Constant \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} -1000 & -500 & 0 & 500 & 1000 \\ -1000 & -500 & 0 & 500 & 1000 \\ -1000 & -500 & 0 & 500 & 1000 \\ -1000 & -500 & 0 & 500 & 1000 \end{bmatrix}$
Support Vector IV (Proposed GME)	$z' = \left[TrimMean - \frac{\sigma}{4} \quad TrimMean - \frac{\sigma}{8} \quad TrimMean \quad TrimMean + \frac{\sigma}{8} \quad TrimMean + \frac{\sigma}{4} \right]$

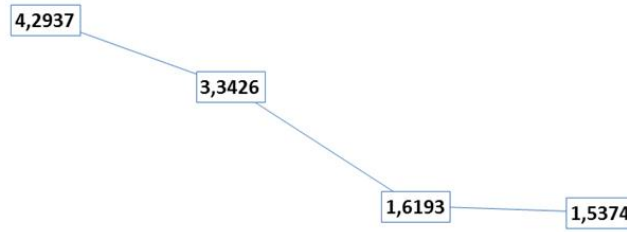


Figure 1. RMSE statistics of the classical and alternative approaches for simulated dataset

As shown in the Figure 1, as the parameter support vector bounds were widened, RMSE decreased. The new estimation (iterative) technique (Support Vector IV) produced better RMSE statistics.

Coefficients and standard errors were given in Table 3. In the new iterative approach (Support Vector IV), standard errors were estimated relatively small.

Table 3. Coefficients and standard errors of different parameter support vectors

	Support Vector I	Support Vector II	Support Vector III	Support Vector IV
Intercept	0.35571 (7.25E-04)	0.44843 (1.08E-03)	-0.23014 (2.06E-02)	-3.83604 (1.49E-06)
X	2.45960 (4.78E-03)	2.23619 (4.10E-03)	1.73687 (4.39E-03)	2.19746 (2.06E-07)
W	0.28050 (5.76E-04)	2.53905 (1.56E-02)	8.56674 (2.75E-02)	9.01914 (9.33E-08)

():Standard Deviation

4.3 An Application to Real Life Data

New procedure was tested with the data of the Programme for International Study Assessment (PISA) conducted in 2009. Data consists of a total sample of 515.985 students who are nested within 1.535 schools (OECD, 2009). Similar to the study of Kılıç et al. (2012), Turkey and neighbouring countries of Turkey, which are Bulgaria, Greece, Azerbaijan, Russian Federation, Israel, Serbia, Romania and Jordan, were included in this study in order to examine learning strategies accounted for mathematics achievement. Thus, the results of two studies could be compared in terms of estimation accuracy. In this study, the study of Kılıç et al. (2012) is named as referenced study.

Table 4. Mathematics scores of countries

Countries	Mathematics Score
Russian Federation	468
Greece	466
Israel	447
Turkey	445
Serbia	442
Azerbaijan	431
Bulgaria	428
Romania	427
Jordan	387
OECD Average	496

PISA mathematics test score of 42.417 15-year-old students were analysed. Three-level random coefficient model was used to model differences across countries and across schools. In this study, mathematics achievement was considered as a dependent variable. At the first level, gender, socio-economic status, elaboration, memorization, control strategy, home

educational resources and cultural possession were considered. School size and student–teacher ratio were considered at the second level and gross domestic product (GDP) was considered at the third level variables. Mathematics achievement score of countries are given in Table 4.

The best performer country is Russian Federation while the worst performer country is Jordan in PISA 2009 study in terms of mathematics achievement. Turkey is at rank 4 among these countries.

First-Level Variables;

1. Gender: Male coded as 1 and Female coded as 0.
2. Socio–Economic and Cultural Status (ESCS): The index of ESCS was derived from three indices: home possessions, higher parental occupation (HISEI) and higher parental education expressed as years of schooling.
3. Cultural Possession (CULTPOSS): It was derived from students' responses to the three items listed below.
 - a) Classic literature, b) Books of property, c) Works of art
4. Home Educational Resources (HEDRES): The PISA 2006 index of home educational resources was derived from students' responses to the some items.
 - a) Desk for study, b) A quiet place to study, c) Your own calculator, d) Books to help with your school work, e) A dictionary
5. Memorization (MEMOR): It was derived from students' responses to the four items measuring preference for memorisation/rehearsal as a learning strategy for mathematics as listed below.
 - a) I go over some problems in mathematics so often that I feel as if I could solve them in my sleep, b) When I study for mathematics, I try to learn the answers to problems off by heart, c) In order to remember the method for solving a mathematics problem, I go through examples again and again, d) To learn mathematics, I try to remember every step in a procedure.
6. Elaboration (ELAB): It was derived from students' responses to the five items measuring preference for elaboration strategy as listed below.
 - a) When I am solving mathematics problems, I often think of new ways to get the answer, b) I think how the mathematics I have learnt can be used in everyday life, c) I try to understand new concepts in mathematics by relating them to things I already know, d) When I am solving mathematics problems, I often think about how the solution might be applied to other interesting questions, e) When learning mathematics, I try to relate the work to things I have learnt in other subjects
7. Control Strategy (CSTRAT): Control learning strategies was derived from students' responses to the five items measuring preference for control as a learning strategy as listed below.
 - a) When I study for a mathematics test, I try to work out what are the most important parts to learn, b) When I study mathematics, I make myself check to see if I remember the work I have already done, c) When I study mathematics, I try to figure out which concepts I still have not understood properly, d) When I cannot understand something in mathematics, I always search for more information to clarify the problem, e) When I study mathematics, I start by working out exactly what I need to learn.

Second-Level Variables;

1. School Size (SCSIZE): Total number of male and female students in a school.
2. Student–Teacher Ratio (STRATIO): The number of students per teacher in a school.

Third-Level Variable;

1. Gross Domestic Product (GDP): Gross Domestic Product of a country. Since GDP has right skewed distribution, Log(GDP) is used in this study.

Descriptive statistics of these variables are given in Table 5. Male-Female ratio was 50%-50%.

Table 5. Descriptive statistics of the variables

	Min	Max	Mean	Standard Deviation
Socio-Economic and Cultural Status	-4.79	3.09	-0.31	1.05
Home Educational Resources	-4.52	2.64	0.11	1.25
Cultural Possession	-2.24	1.82	0.16	1.03
Memorization	-3.02	2.69	0.30	1.03
Elaboration	-2.41	2.76	0.34	1.02
Control Strategy	-3.45	2.50	0.18	1.03
School Size	9.00	3140.00	713.05	448.25
Student-Teacher Ratio	0.51	79.63	13.62	6.47
Log(Gross Domestic Product)	10.40	12.09	11.13	0.57

Output of Restricted Maximum Likelihood (REML) estimation method is given in Table 6.

Table 6. Restricted maximum likelihood estimation

	Coefficient	Standard Deviation
Intercept	123.66	81,63
Gender (Male)	18.21	0,65
Socio-Economic and Cultural Status	10.71	0,39
Home Educational Resources	4.16	0,35
Cultural Possession	4.15	0,35
Memorization	-11.59	0,34
Elaboration	2.76	0,36
Control Strategy	11.56	0,39
School Size	-1.45	0,25
Student-Teacher Ratio	0.02	0,003
Log(Gross Domestic Product)	28.42	7,31
<i>RMSE</i>		1219.5

Table 7 was prepared to identify n and t for sampling with replacement procedure. Model outputs are shown by n and t values for PISA dataset.

Table 7. Coefficient estimates and standard deviation for different sample size and repeats

	Intercept	Gender	Escs	Memor	Elab	Cstrat	Cultposs	Hedres	Stratio	Schsize	LogGDP	RMSE
Sample Size:n t: 1000	131.56	10.29	23.96	-15.10	0.29	17.47	5.49	4.00	-1.43	0.01	29.24	733.68
Sample Size:n t: 2000	131.54	10.30	23.94	-15.13	0.28	17.47	5.51	3.98	-1.43	0.01	29.24	733.68
Sample Size:n t: 3000	131.54	10.29	23.95	-15.11	0.29	17.48	5.50	3.99	-1.43	0.01	29.24	131.54
Sample Size:n t: 5000	131.28	10.32	23.95	-15.12	0.29	17.47	5.49	3.99	-1.43	0.01	29.26	733.68
Sample Size:n t: 10000	131.60	10.29	23.95	-15.12	0.29	17.47	5.50	3.99	-1.43	0.01	29.23	733.68
Sample Size:2n t: 1000	131.23	10.31	23.95	-15.12	0.28	17.48	5.50	3.99	-1.43	0.01	29.26	733.68

Sample Size:2n	131.26	10.33	23.96	-15.12	0.29	17.48	5.49	3.99	-1.43	0.01	29.26	733.68
t: 2000												
Sample Size:2n	131.58	10.29	23.95	-15.13	0.29	17.47	5.50	4.00	-1.43	0.01	29.23	733.68
t: 3000												
Sample Size:2n	131.39	10.31	23.95	-15.12	0.29	17.47	5.50	3.99	-1.43	0.01	29.25	733.68
t: 5000												
Sample Size:2n	131.39	10.31	23.95	-15.12	0.29	17.47	5.50	3.99	-1.43	0.01	29.25	733.68
t: 10000												

Similar to simulation study, the results of the different scenarios were compared and there was no significant difference for all scenarios in terms of β coefficients and RMSE statistics. For this reason, similar to simulation study, determination of n and t is not important for PISA dataset. In this study, first scenario (2n sample size and 1000 repeats) was chosen for further analysis.

Different support vectors were used for each regression model. Thus, the efficiency of the new procedure could be compared by using RMSE statistics. The parameter support vectors are given in Table 8 for PISA dataset.

Table 8. Different support vectors of the new and classical approaches

Support Vectors	Boundaries				
Support Vector I for all coefficients	[-10 -5 0 5 10]				
Support Vector II for all coefficients	[-100 -50 0 50 100]				
Support Vector III for all coefficients	[-1000 -500 0 500 1000]				
Support Vector IV for all coefficients	$TrimMean - \frac{\sigma}{4}$	$TrimMean - \frac{\sigma}{8}$	$TrimMean$	$TrimMean + \frac{\sigma}{8}$	$TrimMean + \frac{\sigma}{4}$
	$k = 1, 2, \dots, K$				

(σ): Standard Deviation

In Figure 2, the RMSE statistics of the new iterative approach and classical approach were shown.

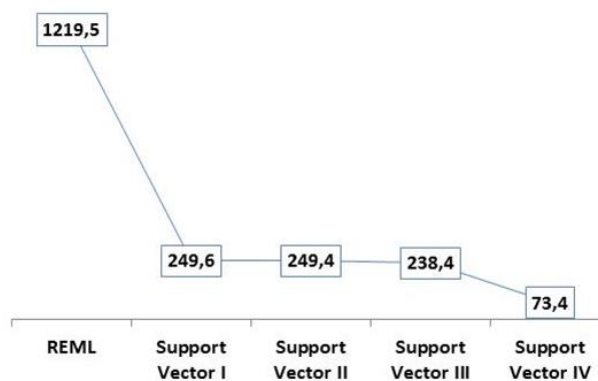


Figure 2. RMSE statistics of the classical and alternative approaches for PISA dataset

As shown in Figure 2, creation of prior information was important to obtain a more robust model in GME. As the parameter support vector bounds were widened, RMSE decreased. The new estimation (iterative) technique (Support

Vector IV) produced lower RMSE and it was significant that the iterative technique required no assumptions regarding distribution or parameters.

Coefficients and standard errors are given in Table 9.

Table 9. Coefficient estimates and standard deviation of the alternative approach for PISA dataset

	REML	Support Vector I	Support Vector II	Support Vector III	Support Vector IV
Intercept	123.66 (81.63)	0.00046 (1.45E-06)	0.00139 (5.81E-06)	0.08998 (4.98E-04)	131.2301 (4.70E-05)
Gender	18.21 (0.65)	0.00022 (7.09E-07)	0.67800 (2.79E-06)	0.043324 (2.39E-04)	10.31170 (2.32E-07)
ESCS	10.71 (0.39)	-0.00014 (4.31E-07)	-0.00057 (3.00E-06)	-0.04168 (2.63E-04)	23.94882 (4.51E-08)
MEMOR	-11.59 (0.34)	0.00015 (4.66E-07)	0.00036 (1.39E-06)	0.020285 (1.17E-04)	-15.11750 (5.38E-08)
ELAB	2.76 (0.36)	0.00016 (5.23E-07)	0.00045 (1.89E-06)	0.027699 (1.61E-04)	0.28162 (6.17E-08)
CSTRAT	11.56 (0.39)	0.00009 (3.01E-07)	0.00019 (7.03E-07)	0.008873 (5.70E-05)	17.48090 (4.70E-08)
CULTPOSS	4.15 (0.35)	0.00008 (2.56E-07)	0.00021 (7.89E-07)	0.012142 (6.70E-05)	5.49562 (2.53E-08)
HEDRES	4.16 (0.35)	0.00007 (2.16E-07)	0.00009 (4.90E-07)	0.001621 (4.20E-05)	3.99492 (1.59E-08)
STRATIO	0.02 (0.003)	0.00698 (2.20E-05)	0.02054 (9.50E-05)	1.288434 (0.00814)	-1.43140 (7.05E-08)
SCHSIZE	-1.45 (0.25)	0.44340 (1.44E-03)	0.44260 (1.44E-03)	0.41276 (0.00139)	0.00884 (5.90E-10)
LogGDP	28.42 (7.31)	0.00507 (1.60E-05)	0.15671 (6.50E-05)	1.011791 (0.00559)	29.26376 (4.04E-06)

():Standard Deviation

In the new iterative approach (Support Vector IV), standard errors were estimated relatively small. Furthermore, RMSE obtained from new approach was smaller than other alternatives.

5. Conclusion

In the literature, there are many different approaches in order to make a robust estimation. The GME estimator is a one of the robust estimators resistant to multicollinearity, heteroscedasticity and the existence of outliers. Although it was not necessary to make strict assumptions about parameters or population distributions, the most important point was to specify discrete supports for the coefficients and the error term in this approach, which had a significant effect on the results obtained.

In this study, a new approach based on iterative process was presented for Random Intercept Models. The Classical GME Approach, using wide parameter support boundaries, was mostly suggested in the literature in case the researcher had no prior information. As opposed to this technique, a new iterative approach is suggested in this paper which will help the researcher to obtain consistent parameters without prior information. Furthermore it was adopted multilevel datasets.

The new approach was tested on both simulated and real-life datasets. The obtained results proved that the suggested approach provides better parameter estimates and lowest RMSE than classical methods. The results produced with support vectors suggested in the literature (Support Vector I, II and III) were not closer to the true regression parameters. However, parameter estimations of proposed approach were relatively closer to the true regression parameters.

References

- Abellan, J., Baker, R. M., & Coolen, F. P. A. (2011). Maximising Entropy on the Nonparametric Predictive Inference Model for Multinomial Data. *European Journal of Operational Research*, 212(1), 112-122.
<http://dx.doi.org/10.1016/j.ejor.2011.01.020>
- Akdeniz, F., Çubuk, A., & Güler, H. (2011). Generalized Maximum Entropy Estimators: Applications to the Portland Cement Dataset. *The Open Statistics & Probability Journal*, 3, 13-20.
<http://dx.doi.org/10.2174/1876527001103010013>
- Al-Nasser, A. D. (2014). Two Steps Generalized Maximum Entropy estimation procedure for fitting linear regression when both covariates are subject to error. *Journal of Applied Statistics*, 41(8), 1708-1720.
<http://dx.doi.org/10.1080/02664763.2014.888544>
- Al-Nasser, A. D. (2011). An Information-Theoretic Alternative to Maximum Likelihood Estimation Method in Ultrastructural Measurement Error Model. *Hacettepe University Bulletin of Natural Sciences and Engineering Series B: Mathematics and Statistics*, 40(3), 469-481.
- Al-Rawwash, M. Y., & Al-Nasser, A. D. (2011). Longitudinal Data Analysis Using Generalized Maximum Entropy Approach. *Jordan Journal of Mathematics and Statistics*, 4(1), 47-60.
- Al-Nasser, A. D., Eidous, O. M., & Mohaidat, L. M. (2010). Multilevel Linear Models Analysis using Generalized Maximum Entropy. *Asian Journal of Mathematics and Statistics*, 3(2), 111-118.
<http://dx.doi.org/10.3923/ajms.2010.111.118>
- Al-Nasser, A. D., & Al-Atrash, A. R. (2011). Information Theoretic Approach for constructing a super Data Envelopment Analysis. *Asian Journal on Quality*, 12(1), 54-66. <http://dx.doi.org/10.1108/15982681111140543>
- Altaylıgil, B. (2008). *Entropi Ölçüsü ve Bazı Ekonometri Uygulamaları (Entropy Measure and Some of Its Econometric Applications)(Doctorate Thesis)*. İstanbul: İstanbul Üniversitesi.
- Caputo, M. R., & Paris, Q. (2008). Comparative statics of the generalized maximum entropy estimator of the general linear model. *European Journal of Operational Research*, 185(1), 195-203.
<http://dx.doi.org/10.1016/j.ejor.2006.12.031>
- Ciavolino, E., & Calcagni, A. (2014). A generalized maximum entropy (GME) approach for crisp-input/fuzzy-output regression model. *Quality & Quantity*, 48(6), 3401-3414. <http://dx.doi.org/10.1007/s11135-013-9963-9>
- Depren, Ö. (2014). *Regresyonda Maksimum Entropi Modellemesi (Maximum Entropy Modelling in Regression)(Doctorate Thesis)*. İstanbul: Marmara Üniversitesi.
- Donoso, P., Grange, L. D., & González, F. (2011). A Maximum Entropy Estimator for the Aggregate Hierarchical Logit Model. *Entropy*, 13(12), 1425-1445. <http://dx.doi.org/10.3390/e13081425>
- Fernandez-Vazquez, E., Mayor-Fernandez, M., & Rodriguez-Valez, J. (2008). Estimating spatial autoregressive models by GME-GCE techniques. *International Regional Science Review*, 32(2), 148-172.
<http://dx.doi.org/10.1177/0160017608326600>
- Gastón, A., & García-Vinas, J. I. (2011). Modelling Species Distributions with Penalised Logistic Regressions: A Comparison With Maximum Entropy Models. *Ecological Modelling*, 222(13), 2037-2041.
<http://dx.doi.org/10.1016/j.ecolmodel.2011.04.015>
- Gelman, A. (2006). Multilevel (Hierarchical) Modeling: What It Can and Cannot Do?. *Technometrics*, 48(3), 432-435.
<http://dx.doi.org/10.1198/004017005000000661>
- Golan, A., & Gzyl, H. (2012). An entropic estimator for linear inverse problems. *Entropy*, 14(5), 892-923.
<http://dx.doi.org/10.3390/e14050892>
- Golan, A., Judge, G. G., & Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with limited data*. New York: John Wiley & Sons.
- Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, 4(106), 620-630.
<http://dx.doi.org/10.1103/physrev.106.620>
- Henderson, H., Golan, A., & Seabold, S. (2015). Incorporating prior information when true priors are unknown: An Information-Theoretic approach for increasing efficiency in estimation. *Economics Letters*, 127, 1-5.
<http://dx.doi.org/10.1016/j.econlet.2014.12.014>
- Kılıç, S., Çene, E., Demir, İ. (2012). Comparison of Learning Strategies for Mathematics Achievement in Turkey with Eight Countries. *Kuram ve Uygulamada Eğitim Bilimleri - Educational Sciences: Theory & Practice*, 12(4),

2585-2598.

Miller, A. J. (2002). *Subset Selection in Regression*. London: Chapman&Hall/CRC.

<http://dx.doi.org/10.1201/9781420035933>

Mittelhammer, R., Cardell, N. S., & Marsh, T. L. (2013). The Data-Constrained Generalized Maximum Entropy Estimator of the GLM: Asymptotic Theory and Inference. *Entropy*, 15(5), 1756-1775.

<http://dx.doi.org/10.3390/e15051756>

Organization for Economic Cooperation and Development (OECD). (2009c). *PISA 2006 Technical Report*, OECD, Paris.

Pukelsheim, F. (1994). The Three Sigma Rule. *The American Statistician*, 48(2), 88-91.

<http://dx.doi.org/10.1080/00031305.1994.10476030>

Raudenbush, S. W., Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. London: Sage Publications.

Raudenbush, S. W., Bryk, A. S., & Condon, R. (2005). *HLM: Hierarchical Linear and Non-Linear Modeling*. Chicago: Scientific Software International.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379-423.

<http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x>

Timm, N. H. (2002). *Applied Multivariate Analysis*. New York: Springer-Verlag. <http://dx.doi.org/10.1007/b98963>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).