Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi, Cilt 24, Sayı 2, 2015, Sayfa 337-348

POVERTY RATE AND ITS DETERMINANTS FOR 12 STATISTICAL REGIONS OF TURKEY: GENERALIZED MAXIMUM ENTROPY APPROACH*

Hüseyin GÜLER**
Fikri AKDENİZ***
Hasan Altan ÇABUK****
Sibel ÖRK ÖZEL*****

ABSTRACT

In this study, poverty rate of Turkey on 12 statistical regions (NUTS – 1 level) and some determinants of this rate is modeled by a linear regression model. Average household size, unemployment rate, high school and university enrollment rates, median income and urbanization rate as determinants of poverty rate are used as explanatory variables of this model. It is observed that the ordinary least squares (OLS) produce unstable estimates since the design matrix X is subject to strong multicollinearity. In order to obtain stabilized parameter estimates, two biased estimation methods known in the literature, namely Ridge regression and generalized maximum entropy (GME), are used. Inequality and sign constraints that are required in the context of economic theory are used for the GME estimator. Estimators are compared by their efficiency with the estimated mean squared error values obtained by the bootstrap method.

Keywords: Generalized maximum entropy, Least squares, Ridge regression, Multicollinearity, Bootstrap.

ÖZET

Bu çalışmada Türkiye'nin NUTS – 1 düzeyinde 12 istatistiki bölgesi için yoksulluk oranı ve bu oranın belirleyicileri doğrusal regresyon modeli ile analiz edilmiştir. Kurulan modelde açıklayıcı değişken olarak yoksulluk oranının belirleyicileri olan ortalama hanehalkı büyüklüğü, işsizlik oranı, lise ve üniversite okullaşma oranları, medyan gelir ile şehirleşme oranı kullanılmıştır. Tasarım matrisi X'in yüksek dereceden çoklu iç ilişkiye sahip olduğu ve bu nedenle en küçük kareler (EKK) tahmin edicinin tutarsız sonuçlar verdiği tespit edilmiştir. Tutarlı tahminler elde edebilmek adına literatürde yer alan ridge regresyon ve genelleştirilmiş maksimum entropi (GME) yanlı tahmin edicileri modele uygulanmıştır. GME tahmin edici için iktisat teorisinin

---

* Bu çalışma TÜBİTAK 2211-A Genel Yurtiçi Doktora Burs Programı kapsamında desteklenmiştir.
** Yrd. Doç. Dr., Çukurova Üniversitesi, İ.İ.B.F., Ekonometri Bölümü, hguler@cu.edu.tr
***Prof.Dr., Çağ Üniversitesi, Fen-Edebiyat Fakültesi, Matematik-Bilgisayar Bölümü, akdeniz@cu.edu.tr
****Prof.Dr., Çukurova Üniversitesi, İ.İ.B.F., Ekonometri Bölümü, haltan@cu.edu.tr
*****Arş.Gör., Çukurova Üniversitesi, İ.İ.B.F., Ekonometri Bölümü, sork@cu.edu.tr

gerektirdiği eşitsizlik ve işaret kısıtları da modele eklenmiştir. Tahmin edicilerin etkinliği bootstrap yöntemiyle tahmin edilen ortalama hata kareleri ile karşılaştırılmıştır.
Anahtar kelimeler: Genelleştirilmiş maksimum entropi, En küçük kareler, Ridge regresyon, Çoklu iç ilişki, Bootstrap

1. Introduction

Parameters of a linear regression model can be estimated with the ordinary least squares (OLS) method. However, assumptions of OLS might be violated in real-world datasets. Deviations from assumptions might arise from sample that is being used. Datasets might be influenced by some errors during sample selection or composing data. In this cases, various problems occur.

One of these problems is the multicollinearity of the dataset. This problem is described as being ill-posed because of non-stationarity or since the number of parameters to be estimated exceeds to number of data points. Alternatively, it is described as being ill-conditioned when the parameter estimates are highly unstable.

The least squares estimators are not biased but their variances and covariances might be inflated in the presence of multicollinearity. In this case, the existence of multicollinearity may result in wider confidence intervals for parameters, may lead estimates with unexpected signs and may affect decisions in hypothesis tests.

Several solutions are proposed in the literature to overcome the multicollinearity problem. One of the solutions is to use biased (but stable) estimators such as the ridge regression estimator proposed by Hoerl and Kennard (1970). They show that the ridge regression estimator is superior to OLS in terms of mean square error (MSE) for a suitably chosen biasing parameter. There is a vast amount of literature about ridge regression estimator following Hoerl and Kennard’s (1970) paper like Sarkar (1992), Akdeniz and Kaçıranlar (1995), Liu (1993), Kibria (2003) and Liu (2003) to name a few.

A more recent solution to the multicollinearity problem is the generalized maximum entropy (GME) estimator proposed by Golan et. al. (1996). In this paper, we examine the GME estimator and compare it with ridge regression and OLS in the sense of mean square error (MSE) criteria on a real dataset. In order to do this, we estimate the parameters of a linear regression model about poverty rate and its determinants in 12 statistical regions of Turkey and we calculate MSE values with the bootstrap method.

The remainder of the paper is organized as follows: Section 2 reviews GME estimation in the linear regression model. In section 3, we consider the poverty dataset for 12 statistical regions of Turkey and estimate the parameters of the regression model with OLS, ridge regression and GME. We also compare these estimators according to the MSE values obtained with bootstrap replications. Section 4 concludes the paper.

2. Generalized Maximum Entropy Estimator

GME estimator is proposed by Golan, Judge and Miller (1996). The maximum entropy (ME) estimator is based on Shannon’s (1948) information entropy concept and Jaynes’ (1957) maximum entropy principle. ME has been used for the solution of pure inverse problems under the cover of information concept. The ME principle suggests that the probability distribution which best represents the current state of knowledge is the one that has the maximum entropy. According to ME, the model with the maximum entropy gives the proper distribution for the problem being investigated. ME is used only for pure inverse problems while GME is used for both pure inverse problems and ill-posed inverse problems.

To explain how GME works, consider the general linear model (GLM)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} \tag{1}$$

where $\boldsymbol{y}$ is a $T \times 1$ vector of sample observations on the dependent variable, $\boldsymbol{X}$ is a $T \times K$ design matrix, $\boldsymbol{\beta}$ is a $K \times 1$ vector of unknown parameters and $\boldsymbol{u}$ is a $T \times 1$ vector of unknown errors. Golan, Judge and Miller (1996) redefines unknown parameters and unknown errors for GME estimation by using compact supports. Each regression coefficient β_k is reparametrized as a discrete random variable with a compact support interval consisting of $2 \leq M < \infty$ possible outcomes (Golan et. al., 1996, p. 86). Thus, we can express each β_k as a convex combination $\beta_k = \boldsymbol{z}'_k \boldsymbol{p}_k$, where $\boldsymbol{z}_k = [z_{k1} \quad \cdots \quad z_{kM}]'$ are called support vectors (hypothesized values for the parameters) and $\boldsymbol{p}_k = [p_{k1} \quad \cdots \quad p_{kM}]'$ are corresponding unknown probabilities. These convex combinations may be written in the matrix form as

$$\boldsymbol{\beta} = \boldsymbol{Zp} = \begin{bmatrix} \boldsymbol{z}'_1 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{z}'_2 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{z}'_k & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{z}'_K \end{bmatrix}_{K \times KM} \begin{bmatrix} \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \\ \vdots \\ \boldsymbol{p}_k \\ \vdots \\ \boldsymbol{p}_K \end{bmatrix}_{KM \times 1} . \tag{2}$$

Similarly, each error term can be written as a convex combination hypothesized values and corresponding unknown probabilities. Let $\boldsymbol{V}$ a $T \times TJ$ matrix of unknown support values for $\boldsymbol{u}$, and $\boldsymbol{w}$ a vector of probability weights $TJ \times 1$ such that $w > 0$, where J $(2 \leq J < \infty)$ is the number of support values of errors. There exist sets of error bounds v_{t1} and v_{tJ} for each u_t so that $Pr\left[v_{t1} < u_t < v_{tJ}\right]$ may be made arbitrarily small (Golan et. al., 1996, p. 87). Then each error term u_t can be written as $u_t = \boldsymbol{v}'_t \boldsymbol{w}_t$, where $\boldsymbol{v}_t = [v_{t1} \quad \cdots \quad v_{tJ}]'$ and $\boldsymbol{w}_t = [w_{t1} \quad \cdots \quad w_{tJ}]'$. These convex combinations may be written in the matrix form as

$$\boldsymbol{u} = \boldsymbol{Vw} = \begin{bmatrix} \boldsymbol{v}'_1 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{v}'_2 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{v}'_t & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{v}'_T \end{bmatrix}_{T \times TJ} \begin{bmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{w}_t \\ \vdots \\ \boldsymbol{w}_T \end{bmatrix}_{TJ \times 1} . \tag{3}$$

Golan, Judge and Miller (1996) recommend using the three-sigma rule of Pukelsheim (1994) to establish bounds on the error components.

Using the reparameterized unknowns $\boldsymbol{\beta} = \boldsymbol{Zp}$ and $\boldsymbol{u} = \boldsymbol{Vw}$ given in (2) and (3), Judge and Golan (1992) rewrite the GLM in (1) as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} = \boldsymbol{XZp} + \boldsymbol{Vw}. \tag{4}$$

The objective of GME is to predict the unknown parameters of (3) using the sets of probabilities p and w. Accordingly, Golan et. al. (1996) expresses GME solution to the linear inverse problem with noise that selects $p, w \gg 0$ as follows:

$$maxH(\boldsymbol{p},\boldsymbol{w}) = -\sum_{k=1}^{K}\sum_{m=1}^{M} p_{km} ln p_{km} - \sum_{t=1}^{T}\sum_{j=1}^{J} w_{tj} ln w_{tj} \quad (5a)$$

subject to:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} = \boldsymbol{XZp} + \boldsymbol{Vw} \quad (5b)$$

$$\sum_{m=1}^{M} p_{km} = 1 \quad , \quad k = 1, \dots, K \quad (5c)$$

$$\sum_{j=1}^{J} w_{tj} = 1 \quad , \quad t = 1, \dots, T. \quad (5d)$$

GME parameter and error estimates are obtained by defining and solving a Lagrangian function for (5a) – (5d) and they are given by

$$\hat{\boldsymbol{\beta}} = \boldsymbol{Z}\hat{\boldsymbol{p}} \quad (6)$$

and

$$\hat{\boldsymbol{u}} = \boldsymbol{V}\hat{\boldsymbol{w}} \quad (7)$$

where

$$\hat{p}_{km} = \frac{exp(-\sum_{t=1}^{T}\hat{\lambda}_t z_{km} x_{tk})}{\Omega_k^p(\hat{\lambda})}, \quad (8)$$

$$\hat{w}_{tj} = \frac{exp(-\hat{\lambda}_t v_{tj})}{\Psi(\hat{\lambda})}, \quad (9)$$

$$\Omega_k^p(\hat{\boldsymbol{\lambda}}) = \sum_{m=1}^{M} exp\left(-\sum_{t=1}^{T}\hat{\lambda}_t z_{km} x_{tk}\right), \quad (10)$$

and

$$\Psi(\hat{\boldsymbol{\lambda}}) = \sum_{j=1}^{J} exp\left(-\hat{\lambda}_t v_{tj}\right). \quad (11)$$

The GME solution to (5a) – (5d) requires solving a non-linear programming system. Golan et. al. (1996) give more details about the estimation procedure. Since this is an iterated procedure, the standard errors of the GME estimates can be estimated with the bootstrap procedure as pointed out by Akdeniz et. al (2011).

3. Comparing Estimators on Poverty Rate and Its Determinants for 12 Statistical Regions of Turkey

In this section, we analyze the poverty rate and its determinants for 12 statistical regions of Turkey in the first level (NUTS-1). We estimate the parameters of the linear regression model for poverty dataset using OLS, ridge, and GME estimators and compare aforementioned estimators according to MSE values obtained with bootstrap. Calculations and comparisons are done with GAUSS 10 codes.

The statistical regions of Turkey in the first level are given in Table 1. We use the model of Campbell and Hill (2001) and Ramanathan (2002) and extend it with the suggestion of Vazquez et. al. (2009). We consider the following linear regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_8 X_{8i} + u_i \;,\; i = 1,2,\dots,12. \quad (12)$$

where the dependent variable is the poverty rate (Y) and explanatory variables are average household size (X_2), unemployment rate (X_3), high school enrollment rate (X_4), university enrollment rate (X_5), median income (X_6), urbanization rate (X_7), and squared urbanization rate (X_8). Vazquez et. al. (2009) have pointed out a U-shape relationship between the level of urbanization and poverty. According to this, poverty decreases when urbanization level increases at the beginning but after a certain point, poverty increases when urbanization level decreases. For this reason, we add squared urbanization rate to the model. In addition to this, Campbell and Hill (2001) points out that poverty rate increases when average household size and unemployment rate increase. Moreover, it is expected that high school enrollment rate, university enrollment rate and median income have a negative impact on poverty rate. We also expect that urbanization rate has a negative effect and squared urbanization rate has a positive effect on poverty rate according to the U-shape relationship between urbanization and poverty given by Vazquez et. al. (2009). Therefore, the prior information about parameter signs are $\beta_2, \beta_3, \beta_8 > 0$ and $\beta_4, \beta_5, \beta_6, \beta_7 < 0$. In addition to sign restrictions, it is also possible to add a magnitude restriction to the model following Campbell and Hill (2001). According to Campbell and Hill (2001), it is expected that the effect of the university enrollment rate to poverty rate is smaller than the effect of the high school enrollment rate. Therefore, we expect that $\beta_5 < \beta_4 < 0$.

Table.1 Statistical Regions of Turkey (NUTS-1 Level)

Code	**Region**
TR1	Istanbul
TR2	West Marmara
TR3	Aegean
TR4	East Marmara
TR5	West Anatolia
TR6	Mediterranean
TR7	Central Anatolia
TR8	West Black Sea
TR9	East Black Sea
TRA	Northeast Anatolia
TRB	Central East Anatolia
TRC	Southeast Anatolia

The dataset is obtained from Turkish Statistical Institute and it covers the year 2011. Descriptive statistics and the correlation matrix of the explanatory variables are given in Tables 2, and 3, respectively. As can be seen form Table 3, the sample correlation coefficients are quite high and many of them are above 80%. We also note that correlations are generally significant at 5% or 1% significant level. Therefore, it is possible to argue that the linear regression model in (12) might be affected by multicollinearity.

OLS estimates, their standard errors and t-statistics along with some model statistics are given in Table 4. It is seen that the model has a high R^2 statistic while the F and most of the t-statistics are insignificant. In addition to this, signs of $\hat{\beta}_2$ and $\hat{\beta}_5$ are in contrast to the prior information and the magnitude restriction is not satisfied for $\hat{\beta}_4$ and $\hat{\beta}_5$. These results indicate that the model is under the influence of multicollinearity. Another diagnostic of multicollinearity is the condition number suggested by Belsley et. al. (1980) and it is defined as $\kappa = \sqrt{\xi_{max}/\xi_{min}}$ where ξ_{max} and ξ_{min} are the largest and smallest eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$. For the model in (12), the condition number $\kappa = 683.331$ indicates that model suffers from multicollinearity and OLS estimates are affected badly.

Table.2 Means and Standard Deviations of Poverty and Its Determinants

Variable	Mean	Standard Deviation
Y	12.578	1.732
X_2	4.033	0.876
X_3	9.267	1.996
X_4	17.405	2.609
X_5	10.221	2.462
X_6	8.333	1.670
X_7	71.028	13.968
X_8	5223.867	2124.395

Table.3 Correlations Between Explanatory Variables

Variables	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_2	1	0,453	-0,661*	-0,578 *	-0,626*	-0,376	-0,362
X_3	0,453	1	0,157	0,100	0,135	0,464	0,453
X_4	-0,661*	0,157	1	0,873**	0,858**	0,812**	0,816**
X_5	-0,578**	0,100	0,873**	1	0,841**	0,805**	0,809**
X_6	-0,626*	0,135	0,858**	0,841**	1	0,835**	0,858**
X_7	-0,376	0,464	0,812**	0,805**	0,835**	1	0,996**
X_8	-0,362	0,453	0,816**	0,809**	0,858**	0,996**	1

*: $p - value < 5\%$, **: $p - value < 1\%$

Table.4 OLS Estimates

Variable	$\hat{\boldsymbol{\beta}}$	Standard Error	$\boldsymbol{t}$
Constant	47.731	40.035	1.192
X_2	-0.637	1.799	-0.354
X_3	0.258	0.459	0.563
X_4	-1.262	0.416	-3.032
X_5	0.905	0.350	2.584
X_6	-0.706	1.107	-0.637
X_7	-0.507	0.744	-0.682

X_8	0.004	0.005	0.718
$R^2 = 0.867$	$\hat{\sigma} = 1.047$	$F = 3.727$	$p - value = 0.111$

In order to overcome the multicollinearity problem, we apply the ridge regression estimator defined by Hoerl and Kennard (1970):

$$\widehat{\boldsymbol{\beta}}(k) = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{13}$$

where $k > 0$ is the biasing parameter. Various methods are proposed in the literature for the optimum choice of k. We use two different suggestions given in the literature. One of them is proposed by Hoerl et. al. (1975) and it is given as

$$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\beta}}} \tag{14}$$

where p is the number of parameters, $\hat{\sigma}^2$ and $\widehat{\boldsymbol{\beta}}$ are the least square estimates of error variance and model parameters, respectively. Hoerl et. al. (1975) show that ridge regression estimator is superior to OLS in the sense of MSE for the k value given in (14). Alternatively, we also use the k value suggested by Lawless and Wang (1976) which is given by

$$\hat{k}_{LW} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \xi_i \hat{\alpha}_i^2} \tag{15}$$

where $\xi_1 > \xi_2 > \cdots > \xi_p$ are the eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$ and $\hat{\alpha}_i$ $(i = 1, \dots, p)$ are the least squares estimates of the canonical model parameters.

Ridge estimates for $\hat{k}_{HKB}$ and $\hat{k}_{LW}$ are given in Table 5. It is seen that the signs of ridge estimates are in contrast to the prior information except for $\hat{\beta}_2$ and $\hat{\beta}_4$. We also observe that magnitude restriction ($\hat{\beta}_5 < \hat{\beta}_4$) is not satisfied ridge estimates. Since there is no sign restriction on the ridge estimator it is not assured that all estimates have the expected signs. However, these results show that the ridge estimator is not as suitable as it is expected to solve the multicollinearity problem for this dataset.

Table.5 Ridge Estimates

Variable	$\widehat{\boldsymbol{\beta}}(\hat{k}_{HKB})$	$\widehat{\boldsymbol{\beta}}(\hat{k}_{LW})$
Constant	7.253	6.281
X_2	1.083	1.124
X_3	-0.104	-0.113
X_4	-0.979	-0.972
X_5	0.727	0.722
X_6	0.317	0.341
X_7	0.235	0.253
X_8	-0.001	-0.002

The second estimator we use to overcome the multicollinearity problem is GME. Two different GME estimates are calculated in this paper: In the first one, GME is applied without any restrictions (unrestricted GME) and in the second we use sign and inequality restrictions on parameters according to prior information (restricted GME).

Parameter and error supports for the unrestricted GME are obtained from OLS estimates. The lower and upper bounds of parameter supports are determined with $\hat{\beta}_k \pm 3 \times se(\hat{\beta}_k)$ where $\hat{\beta}_k$ is the OLS estimate of β_k and $se(\hat{\beta}_k)$ is the standard error of the estimate. In this method, prior means of parameter supports are equal to OLS estimates of related parameters. Table 6 gives parameter supports for unrestricted GME.

For restricted GME, we consider the sign restrictions $\beta_2, \beta_3, \beta_8 > 0$ and $\beta_4, \beta_5, \beta_6, \beta_7 < 0$ by assigning nonnegative and nonpositive supports, respectively. In order to account the magnitude restriction $\beta_5 < \beta_4 < 0$, we consider the following reparametrization

$$\begin{bmatrix} \beta_4 \\ \beta_5 \end{bmatrix} = \boldsymbol{Z}^* \begin{bmatrix} \boldsymbol{p}_4 \\ \boldsymbol{p}_5 \end{bmatrix} = \begin{bmatrix} \boldsymbol{z}_4' & 0 \\ \boldsymbol{z}_4' & \boldsymbol{z}_5' \end{bmatrix} \begin{bmatrix} \boldsymbol{p}_4 \\ \boldsymbol{p}_5 \end{bmatrix} \tag{16}$$

where $\boldsymbol{Z}^*$ is a submatrix of support points for β_4 and β_5, and $\boldsymbol{p}_4$ and $\boldsymbol{p}_5$ represent the unknown probabilities. Since the elements of $\boldsymbol{z}_4'$ and $\boldsymbol{z}_5'$ are nonpositive and $\boldsymbol{p}_4$ and $\boldsymbol{p}_5$ are probabilities, it is easy to verify that $\beta_5 = \boldsymbol{z}_4'\boldsymbol{p}_4 + \boldsymbol{z}_5'\boldsymbol{p}_5 < \beta_4 = \boldsymbol{z}_4'\boldsymbol{p}_4 < 0$. Parameter supports for restricted GME is given in Table 7.

The error support for GME estimates are determined by following Golan et. al. (1996) and Pukelsheim (1994) and the lower and upper bounds of error supports are determined with $0 \pm 3\hat{\sigma}$.

Table.6 Parameter Supports for Unrestricted GME

Variable	Parameter Support	Prior Mean
Constant	$\boldsymbol{z}_1 = [-72.375 \quad -12.322 \quad 47.731 \quad 107.784 \quad 167.837]'$	47.731
X_2	$\boldsymbol{z}_2 = [-6.033 \quad -3.335 \quad -0.637 \quad 2.060 \quad 4.758]'$	-0.637
X_3	$\boldsymbol{z}_3 = [-1.118 \quad -0.430 \quad 0.258 \quad 0.946 \quad 1.634]'$	0.258
X_4	$\boldsymbol{z}_4 = [-2.511 \quad -1.887 \quad 1.262 \quad -0.638 \quad -0.013]'$	-1.262
X_5	$\boldsymbol{z}_5 = [-0.146 \quad 0.380 \quad 0.905 \quad 1.430 \quad 1.955]'$	0.905
X_6	$\boldsymbol{z}_6 = [-4.028 \quad -2.367 \quad -0.706 \quad 0.955 \quad 2.617]'$	-0.706
X_7	$\boldsymbol{z}_7 = [-2.739 \quad -1.623 \quad -0.507 \quad 0.609 \quad 1.724]'$	-0.507
X_8	$\boldsymbol{z}_8 = [-0.012 \quad -0.004 \quad 0.004 \quad 0.012 \quad 0.020]'$	0.004

Table.7 Parameter Supports for Restricted GME

Variable	Parameter Support	Prior Mean
Constant	$\boldsymbol{z}_1 = [-80 \quad -40 \quad 0 \quad 40 \quad 80]'$	0
X_2	$\boldsymbol{z}_2 = [0 \quad 2.5 \quad 5 \quad 7.5 \quad 10]'$	5
X_3	$\boldsymbol{z}_3 = [0 \quad 0.5 \quad 1 \quad 1.5 \quad 2]'$	1
X_4	$\boldsymbol{z}_4 = [-2 \quad -1.5 \quad -1 \quad -0.5 \quad 0]'$	-1
X_5	$\boldsymbol{z}_5 = [-2 \quad -1.5 \quad -1 \quad -0.5 \quad 0]'$	$\hat{\beta}_4 - 1$

X_6	$\boldsymbol{z}_6 = [-4 \quad -3 \quad -2 \quad -1 \quad 0]'$	-2
X_7	$\boldsymbol{z}_7 = [-2 \quad -1.5 \quad -1 \quad -0.5 \quad 0]'$	-1
X_8	$\boldsymbol{z}_8 = [0 \quad 0.5 \quad 1 \quad 1.5 \quad 2]'$	1

Table 8 gives unrestricted and restricted GME parameter estimates with the 95% confidence intervals for supports given in Tables 6 and 7, respectively. Confidence intervals are obtained with the bootstrap method proposed by Efron (1979). Campbell and Hill (2001) and Çabuk and Akdeniz (2007) suggest bootstrap to obtain confidence intervals for GME and Akdeniz et. al. (2011) suggest this method to obtain MSE values for GME. The details of the bootstrap replications are as follows: Let $\widetilde{\boldsymbol{\beta}}$ be an estimate of the coefficient vector $\boldsymbol{\beta}$, $\tilde{\sigma}^2$ denotes the estimated residual variance and $\widetilde{\boldsymbol{u}}$ denotes the standardized estimated residual vector. For each replication j, $\widetilde{\boldsymbol{u}}_j^*$, a random sample taken with replacement from $\widetilde{\boldsymbol{u}}$, is chosen and multiplied by $\tilde{\sigma}$. Then the bootstrap sample of $\boldsymbol{Y}$ for replication j is obtained by $\boldsymbol{Y}_j^* = \boldsymbol{X}\widetilde{\boldsymbol{\beta}} + \widetilde{\boldsymbol{u}}_j^*$. For this bootstrap sample $\boldsymbol{Y}_j^*$ and design matrix $\boldsymbol{X}$, $\widetilde{\boldsymbol{\beta}}_j^*$, the estimate of $\boldsymbol{\beta}$ for replication j, is obtained. This procedure is repeated n times to obtain the empirical sampling distribution of $\widetilde{\boldsymbol{\beta}}$. The 2.5th and 97.5th percentiles of the empirical sampling distribution gives 95% bootstrap confidence interval for $\boldsymbol{\beta}$. Similar to this, MSE of $\widetilde{\boldsymbol{\beta}}$ can be computed from the empirical sampling distribution as

$$MSE(\widetilde{\boldsymbol{\beta}}) = trace\{cov(\widetilde{\boldsymbol{\beta}})\} + bias(\widetilde{\boldsymbol{\beta}})'bias(\widetilde{\boldsymbol{\beta}}) \quad (17)$$

where $cov(\widetilde{\boldsymbol{\beta}})$ and $bias(\widetilde{\boldsymbol{\beta}})$ are the estimated covariance matrix and bias vector from the empirical sampling distribution, respectively. Since the true value of $\boldsymbol{\beta}$ is unknown, $bias(\widetilde{\boldsymbol{\beta}})$ can be estimated with $\widehat{bias}(\widetilde{\boldsymbol{\beta}}) = \hat{E}(\widetilde{\boldsymbol{\beta}}) - \widehat{\boldsymbol{\beta}}$ where $\hat{E}(\widetilde{\boldsymbol{\beta}})$ is the average vector of the empirical sampling distribution and $\widehat{\boldsymbol{\beta}}$ is the least squares estimate $\boldsymbol{\beta}$. This procedure is repeated for 400 times to obtain confidence intervals and MSE values for GME estimators. We also apply this method to obtain MSE values of OLS and ridge estimators and report them with the MSE estimates for GME in Table 9.

It is seen from Table 8 that the unrestricted GME estimates for $\hat{\beta}_2$ and $\hat{\beta}_5$ doesn't meet the expected signs. In addition to this, it is possible to note that $\hat{\beta}_5 > \hat{\beta}_4$ which is in contrast to the prior information about the magnitudes of $\hat{\beta}_4$ and $\hat{\beta}_5$. However, the restricted GME satisfies the sign and magnitude restrictions about the parameters. 95% confidence intervals for unrestricted and restricted GME show that all coefficients are significant at 5% confidence level. We also observe that confidence intervals of restricted GME are wider than the confidence intervals of unrestricted GME, which leads to the result that restricted GME has larger standard errors.

Table.8 GME Estimates

	Unrestricted GME			**Restricted GME**		
		95% Confidence Interval			**95% Confidence Interval**	
Variable	**Estimate**			**Estimate**		
Constant	47.933	41.369	53.002	38.146	30.193	53.115
X_2	-0.647	-1.134	-0.099	1.374	0.149	3.050

X_3	0.260	0.117	0.440	0.401	0.188	0.830
X_4	-1.263	-1.449	-1.130	-0.340	-0.670	-0.116
X_5	0.902	0.719	1.013	-0.629	-1.381	-0.305
X_6	-0.710	-1.269	-0.332	-1.055	-1.860	-0.411
X_7	-0.510	-0.661	-0.288	-0.654	-0.981	-0.437
X_8	0.004	0.002	0.005	0.276	0.004	0.277

Bootstrap MSE estimates for OLS, ridge and GME are given in Table 9. According to this table, it is possible to note that the best estimator for the poverty dataset is GME. Both unrestricted and restricted versions of GME have considerably smaller MSE values compared with OLS and ridge. It is also seen that unrestricted GME has a smaller MSE than the restricted one. Since prior mean of the unrestricted GME is equal to least squares estimates, we expect that the bias estimate of unrestricted GME is smaller than restricted GME's. We've also observed that the restricted GME has larger standard errors than the unrestricted GME's. Combining these two might explain the larger MSE value of restricted GME versus the unrestricted one. We also note that ridge estimators are outperformed by OLS, which is an unexpected result under multicollinearity. This might be explained with the fact that ridge estimates doesn't have the expected signs and they are far more different than the OLS estimates (especially for the constant term). Differences between ridge and OLS estimates make the bootstrap bias estimate larger, hence leads to higher MSE values for ridge than OLS.

Table.9 Bootstrap MSE Estimates

Estimator	**Bootstrap MSE Estimate**
$\hat{\boldsymbol{\beta}}$	1245,359
$\hat{\boldsymbol{\beta}}(\hat{k}_{HKB})$	2182,614
$\hat{\boldsymbol{\beta}}(\hat{k}_{LW})$	2213,540
Unrestricted GME	9,495
Restricted GME	65,293

Despite the larger MSE value of restricted GME estimator compared with unrestricted GME, we might argue that restricted GME is the best estimator in this study since it also satisfies the sign and magnitude restrictions about the parameters. MSE is a general but not the best criteria to choose among estimators. In an applied research, it is expected that parameter estimates do not violate strict assumptions of an explicitly stated theoretical model. In the light of this, we state that the restricted GME overcomes the multicollinearity problem and leads to good parameter estimates which are justifiable with the economic theory.

4. Conclusions

In this study, we consider the linear regression model subject to multicollinearity problem. In the presence of multicollinearity, the least squares estimator may produce unreliable results. Biased estimators such as ridge and

generalized maximum entropy (GME) can be used to overcome this problem. In this paper, we evaluate GME estimator and compare it with OLS and ridge regression estimators on a real dataset. We estimate parameters of a linear regression model to examine the poverty rate and its determinants in 12 statistical regions of Turkey. Our results show that the dataset is prone to high levels of multicollinearity problem and the signs and magnitudes of OLS are distorted due to the problem. Sign and magnitude distortions still persist for the ridge regression and unrestricted GME estimator while a solution to this problem is obtained with restricted GME. We also compare aforementioned estimators with the mean square error (MSE) criteria. MSE values are estimated with bootstrap and they show that both unrestricted and restricted GME are superior to OLS and ridge regression while OLS dominates ridge regression estimator. Even though unrestricted GME is better than the restricted one in the sense of MSE, our results show that restricted GME leads to good parameter estimates which are justifiable with the economic theory. Therefore, the restricted GME estimator is chosen as the best estimator for the model being considered. Our results show that the restricted GME estimator might produce reliable parameter estimates that are consistent with the economic theory. As a general result, we suggest researchers to apply different estimators when the model assumptions are not satisfied and choose the best one that fits the model of interest.

REFERENCES

Akdeniz, F., Çabuk, A., & Güler, H. (2011). Generalized maximum entropy estimators: applications to the Portland Cement dataset. *The Open Statistics and Probability Journal*, 3: 13-20.

Akdeniz, F., & Kaçıranlar, S. (1995). On the almost unbiased generalized Liu estimator and unbiased estimation of the bias and mse. *Communications in Statistics - Theory and Methods*, 24(7): 1789-1797.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics*, Wiley, New York.

Campbell, R. C., & Hill, C. R. (2001). Maximum entropy estimation in economic models with linear inequality restrictions. *Working paper*.

Çabuk, A. & Akdeniz, F. (2007). İçilişki ve genelleştirilmiş maksimum entropi tahmin edicileri. *Journal of Statistical Research*, 5(2): 1-19.

Efron, B. (1979). Bootstrap methods: another look at the jacknife. *The Annals of Statistics*, 7(1): 1-26.

Golam, K. B. M. (2003). Performance of Some New Ridge Regression Estimators. *Communications in Statistics: Simulation & Computation*, 32(2): 419-435.

Golan, A., Judge, G., & Miller, D. (1996). *Maximum entropy econometrics: robust estimation with limited data*, John Wiley & Sons, New York.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1): 55-67.

Hoerl, A. E., Kennard, R.W., & Baldwin, K. F. (1975). Ridge regression: some simulation. *Communication in Statistics*, 4: 105-123.

Jaynes, E. T. (1957). Information theory and statistical mechanics II. *Physical Review*, 108(2): 171-190.

Judge, G., & Golan, A. (1992). Recovering information in the case of ill-posed inverse problems with noise. *Working paper, University of California, Berkeley, Department of Agricultural and Resource Economics*.

Lawless, J. F., & Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics: Theory and Methods*, 14: 1589-1604.

Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics: Theory and Methods*, 22(2): 393-402.

Liu, K. (2003). Using Liu-type estimator to combat collinearity. *Communications in Statistics: Theory and Methods*, 32(5): 1009-1020.

Pukelsheim, F. (1994). The three sigma rule. *The American Statistician*, 48(2): 88-91.

Ramanathan, R. (2002). *Introductory econometrics with applications*, Harcourt College Publishers, Fort Worth.

Sarkar, N. (1992). A new estimator combining the ridge regression and the restricted least squares methods of estimation, *Communications in Statistics: Theory and Methods*, 21(7): 1987-2000.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3): 379-423.

Vazquez, J. M., Panudulkitti, P., & Timofeev, A. (2009). Urbanization and the poverty level. *International Studies Program Working Paper 9-14 (updated*), Georgia State University.