Generalized Maximum Entropy Estimators: Applications to the Portland Cement Dataset

Fikri Akdeniz[a,*] Altan Çabuk[b] and Hüseyin Güler[b]

[a]*Department of Statistics, Çukurova University, 01330 Adana, Turkey*

[b]*Department of Econometrics, Çukurova University, 01330 Adana, Turkey*

Abstract: Consider the linear regression model $y = X\beta + u$ in the usual notation. In many applications the design matrix X is frequently subject to severe multicollinearity. In this paper an alternative estimation methodology, maximum entropy is given and used to estimate the parameters in a linear regression model when the basic data are ill-conditioned. We described the generalized maximum entropy (GME) estimator, imposing sign restrictions of parameters and imposing cross parameter restrictions for GME. Mean squared error (*mse*) values of the estimators are estimated by the bootstrap method. We compared the generalized maximum entropy (GME) estimator, least squares and inequality restricted least squares (IRLS) estimator on the widely analyzed dataset on Portland cement.

1. INTRODUCTION

In the econometric literature a distinction is usually made between the types of problem that can affect the classical regression model. A problem is described as being ill-posed because of non-stationarity or because the number of parameters to be estimated exceeds the number of data points. Alternatively, a problem is described as being ill-conditioned when the parameter estimates are highly unstable. This frequently occurs when the data generating process is badly designed, or the data are generated non-experimentally. One problem in this case is collinearity of the data. The problem of multicollinearity and its statistical consequences on a linear regression model are very well-known in statistics. The multicollinearity is defined as the existence of nearly linear dependency among column vectors of the design matrix X in the linear model $y = X\beta + \varepsilon$. The best way of explaining the existence and structure of multicollinearity is to look at the eigenvalues and eigenvectors of the matrix $X'X$. The existence of multicollinearity may result in wide confidence intervals for individual parameters (unstable estimates), may give estimates with wrong signs and may affect our decision in a hypothesis testing. Severe multicollinearity may make the estimates so unstable that they become practically useless. In the statistics literature there are several standard approaches for dealing with the problem of collinearity.

This paper examines the generalized maximum entropy (GME) estimator in the linear model. The maximum entropy estimator is based on Shanon's (information) entropy concept [1] and maximum entropy principle [2].

To date Golan *et al.* [3] have considered applications of GME when the basic data are ill-conditioned and suggest that GME is a feasible alternative method of estimation when faced by ill-conditioned linear inverse problems Campbell and Hill [4] carry out an extensive Monte Carlo experiment in which the effect of the degree of collinearity on the MSE of both GME and restricted GME Campbell and Hill [5] consider the support spaces and effects of collinearity in a Monte Carlo experiment.

Wu [6] generalized the GME method to the weighted GME (W-GME) and proposed a data driven method to select the weights in the entropy objective function. Since GME estimation requires us to specify bounds for parameters, we discuss how to specify the GME parameter and error support matrices. We vary the GME parameter and error support matrices and examine the sensitivity of the GME estimates to the prior information imposed. Applications of restricted maximum entropy estimation can be found in [7] and [8].

Mishra [9] compared the restricted Liu estimates of regression coeffecients with those obtained by applying the maximum entropy Leuven (MEL) family of estimators on the widely analyzed dataset on Portland cement. Using a new development in the method of comparative statics, the sensitivity of the resulting coefficient and error estimates to the prior information is investigated in [10].

The remainder of the paper is organized as follows. Section 2 reviews GME estimation in the linear regression model. In Section 3, to illustrate the theoretical results we consider a dataset on Portland cement. We estimate a regression model, using GME and compare the GME estimates to least squares estimates, inequality restricted least squares (IRLS) estimates, We also describe how to impose inequality restrictions through the GME parameter support matrix.

*Address correspondence to this author at the Department of Statistics, University of Çukurova, 01330 Adana, Turkey; Tel: 90 322 3386548; Fax: 90 322 3386070; E-mail: fikriakdeniz@gmail.com

2. GENERALIZED MAXIMUM ENTROPY ESTIMATION IN THE GENERAL LINEAR MODEL

In this section, we briefly review the principle of the generalized maximum entropy estimator for the linear models

2.1. The GME Estimator

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

$$y_t = \sum_{k=1}^{K} X_{tk}\beta_k + \varepsilon_t, t = 1, 2, ..., N \tag{1}$$

or in matrix form as

$$y = X\beta + e \tag{2}$$

where y is an $N \times 1$ vector of sample observations on the dependent variable, X is an $N \times K$ design matrix, β is a $K \times 1$ vector of unknown parameters and e is an $N \times 1$ vector of unknown errors. Our objective is to estimate the vector β from the sample observations y.

The GME estimator requires a number of support values supplied subjectively and exogenously by the researcher. The estimates as well as their standard errors depend on those support values. Lacking such theoretical knowledge, we usually assume the support space is symmetric around zero with large range.

Golan *et al.* [3] recommend using the "three-sigma rule" of [11] to establish bounds on the error components: the lower bound is:

$$v = -3\hat{\sigma}_y \text{ and the upper bound is } v' = +3\hat{\sigma}_y, \text{ where } \hat{\sigma}_y$$

is the empirical standard deviation of the sample y. For example if $J=5$, then $v = (-3\hat{\sigma}_y, -1.5\hat{\sigma}_y, 0, 1.5\hat{\sigma}_y, 3\hat{\sigma}_y)$ [12].

We briefly summarize how we would obtain a GME estimate of equation (2) [3] proceed by "reparameterizing" β to obtain a problem for which the β values are uniquely determined. This is done by defining a support for β and set of weights which are often called probabilities over the support [3] generalize the maximum entropy methodology and reparameterize the linear model such that the unknown parameters and the unknown errors are in the form of probabilities. Using the reparameterized unknowns, $\beta = Zp$ and $e = Vw$, we rewrite the GLM, equation (2), as

$$y = X\beta + e = XZp + Vw \tag{3}$$

where Z is a $K \times KM$ matrix of support points, and p is a $KM \times 1$ vector of unknown weights such that $p_{km} > 0$ and $p_k' i_M = 1$ for all k where $i_M' = [1,1,...,1]$ and M is the number of support points. Each regression coefficient β_k is reparameterized as a discrete random variable with a compact support interval consisting of $2 \leq M < \infty$ possible outcomes, where a support interval is defined as a closed and bounded interval of the real line. We can express each β_k as a convex combination $\beta_k = z_k' p_k$, where $z_k = [z_{k1}, ..., z_{kM}]'$ and

$p_k = [p_{k1}, ..., p_{kM}]'$ (see [10]). Similarly V is a $N \times NJ$ matrix of known support values for e and w is a vector of probability weights $NJ \times 1$ such that $w > 0$ and $w_t' i_J = 1$ for each t where $i_J' = [1,1,...,1]$, where J is the number of support values chosen for each error w_t.

We jointly estimate the unknown parameters and errors by solving the constrained optimization problem:

$$\text{Max } H(p, w) = -p' \ln(p) - w' \ln(w) \tag{4}$$

subject to

$$y = X\beta + e = XZp + Vw, \ (I_K \otimes i_M')p = i_K, \ (I_N \otimes i_J')w = i_N \tag{5}$$

where $\otimes$ is the Kronecker product. For this problem, we can apply Lagrange method. The lagrangian equation takes the form

$$L = -p'\ln(p) - w'\ln(w) + \lambda'(y - XZp - Vw)$$

$$+ \gamma'[i_K - (I_K \otimes i_M')p] + \delta'[i_N - (I_N \otimes i_J')w] \tag{6}$$

Solving the first order conditions, the GME parameter and error estimates are given by

$$\hat{\beta}_{GME} = Z\hat{p} \tag{7}$$

and

$$\hat{e}_{GME} = V\hat{w} \tag{8}$$

where $\hat{p}$ and $\hat{w}$ are the estimated probability vectors.

3. APPLICATIONS OF GENERALIZED MAXIMUM ENTROPY ESTIMATORS TO THE PORTLAND CEMENT DATASET

We now consider in this section a dataset on Portland cement originally due to Woods *et al.*, [13]. We assemble our data as follows:

$$X = (x_1; x_2, x_3, x_4) = \begin{bmatrix} 7 & 26 & 6 & 60 \\ 1 & 29 & 15 & 52 \\ 11 & 56 & 8 & 20 \\ 11 & 31 & 8 & 47 \\ 7 & 52 & 6 & 33 \\ 11 & 55 & 9 & 22 \\ 3 & 71 & 17 & 6 \\ 1 & 31 & 22 & 44 \\ 2 & 54 & 18 & 22 \\ 21 & 47 & 4 & 26 \\ 1 & 40 & 23 & 34 \\ 11 & 66 & 9 & 12 \\ 10 & 68 & 8 & 12 \end{bmatrix}, \sum_{i=1}^{4} x_i = \begin{bmatrix} 99 \\ 97 \\ 95 \\ 97 \\ 98 \\ 97 \\ 97 \\ 98 \\ 96 \\ 98 \\ 98 \\ 98 \\ 98 \end{bmatrix}, y = \begin{bmatrix} 78.5 \\ 74.3 \\ 104.3 \\ 87.6 \\ 95.9 \\ 109.2 \\ 102.7 \\ 72.5 \\ 93.1 \\ 115.9 \\ 83.8 \\ 113.3 \\ 109.4 \end{bmatrix}$$

These data come from an experimental investigation of the heat evolved during the setting and hardening of Portland cements of varied composition and the dependence of this heat on the percentages of four compounds in the clinkers from which the cement was produced. In this example, the dependent variable y is defined as heat evolved in calories per gram of cement. The four compounds considered by [13] are tricalcium aluminate (X_1), tricalcium silicate (X_2), tetracalcium alumino ferrite (X_3), β - dicalcium silicate (X_4). The four columns of the 13x4 matrix X comprise the data on x_1, x_2, x_3, x_4 respectively.

Consider the following homogeneous linear regression model:

$$y = X\beta + u = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u . \quad (9)$$

The ordinary least squares estimates of β and their estimated standard errors (in parentheses) together with the associated t-statistics with n-p =13-4=9 degrees of freedom and two-sided critical values (in parentheses) are:

$$\tilde{\beta}_{OLS} = \begin{bmatrix} 2.1931 \\ 1.1533 \\ 0.7585 \\ 0.4863 \end{bmatrix}, \begin{bmatrix} (0.1853) \\ (0.0479) \\ (0.1595) \\ (0.0414) \end{bmatrix}, \begin{bmatrix} 11.84 \\ 24.06 \\ 4.76 \\ 11.74 \end{bmatrix}, \begin{bmatrix} (<0.0001) \\ (<0.0001) \\ (0.0010) \\ (<0.0001) \end{bmatrix}$$

where $\tilde{\sigma}^2_{OLS} = 5.845461$ and the $m\hat{s}e$ $(\tilde{\beta}_{OLS}) = 0.0638$. We note that all of the t-statistics are highly significant.

The condition number of X is $\kappa(X) = 20.5846$ so X may be considered as "well-conditioned". We now, however, following Hald ([14], pp.648-649) and Gorman and Toman ([15], pp. 35-36), Daniel and Wood ([16], pp: 89) and [17] add a column of ones to the matrix X and fit an inhomogeneous linear regression model with intercept to the data. Now, consider the following regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u . \quad (10)$$

It is seen that the row totals of the 13×4 matrix X (without the column of ones) are all approximately equal to 100 (the compounds are all measured in percentages). We still have n=13 observations but there are now p=5 unknown regression coefficients. The 13x5 matrix X now has singular values $\theta_1 = 211.3675$, $\theta_2 = 77.2361$,

$\theta_3 = 28.4597$, $\theta_4 = 10.2674$, $\theta_5 = 0.0349$. The condition number of X is now $\kappa(X) = \theta_1 / \theta_5 = 6056.37$, which is exceptionally large and so X may be considered as being quite "ill-conditioned". With an intercept the associated inhomogeneous linear model certainly posses "high multicollinearity" with possible effects on the OLS estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$. The ordinary least squares estimates of β and their estimated Standard errors (in parenthesis) together with the associated t-statistics with n-p =13-4=9 degrees of freedom and two-sided critical values (in parenthesis) are:

$$\hat{\beta}_{OLS} = \begin{bmatrix} 62.4054 \\ 1.5511 \\ 0.5102 \\ 0.1019 \\ -0.1441 \end{bmatrix} , \begin{bmatrix} (70.0710) \\ (0.7448) \\ (0.7238) \\ (0.7547) \\ (0.7091) \end{bmatrix} , \begin{bmatrix} 0.89 \\ 2.08 \\ 0.70 \\ 0.14 \\ -0.20 \end{bmatrix} , \begin{bmatrix} (0.3991) \\ (0.0708) \\ (0.5009) \\ (0.8959) \\ (0.8441) \end{bmatrix}$$

We note that not one of these t-statistics is significant (at the 5% level). Thus, the estimated regression function is:

$$\hat{y} = 62.4054 + 1.551 x_1 + 0.5102 x_2 + 0.1019 x_3 - 0.1441 x_4 ,$$

$\hat{\sigma}^2_{OLS} = 5.982955$ and $R^2 = 0.9824$, which implies that 98.24% of the total variation has been explained by the regressors.

Following [17], we may improve our estimates by adding to the inhomogeneous model with intercept the linear restriction $\beta_3 = \beta_2 - \beta_1$ or $R\beta = r = 0$ where $R = (0, 1, -1, 1, 0)$. We test the linear hypothesis $H_o : R\beta = r = 0$ in the unrestricted linear model (10). The test statistic for testing the null-hypothesis H_o against $H_1 : R\beta \neq r = 0$, given our observations, is $F = \dfrac{(R\hat{\beta}_{OLS} - r)'(RS^{-1}R')^{-1}(R\hat{\beta}_{OLS} - r)}{\hat{\sigma}^2} = 1.92.$

Since $F_{0.05,1,8} = 5.32$, H_o is not rejected at the 5% significance level. We have the following estimates for σ^2 :

(a) unrestricted inhomogeneous model: $\hat{\sigma}^2_{OLS} = 5.982955$

(b) restricted inhomogeneous model with the restriction $\beta_3 = \beta_2 - \beta_1$: $\hat{\sigma}^2_c = 6.59199$.

Table **1** gives summary statistics for the data. The sample coefficient of variation is defined as $CV = s / \bar{x}$, where s is

Table 1. Summary Statistics for Cement Data (n=13 Observations)

Variable	Mean	Min.	Max.	Standard Deviation	CV= $s/\bar{x}$
y	95.423	72.500	115.900	15.044	0.158
x_1	7.462	1.000	21.000	5.882	0.788
x_2	48.154	26.000	71.000	15.561	0.323
x_3	11.769	4.000	23.000	6.405	0.544
x_4	30.000	6.000	60.000	16.738	0.558

Table 2 Sample Correlation Matrix

	y	x_1	x_2	x_3	x_4
y	1.000	0.731	0.816	-0.535	-0.821
x_1	0.731	1.000	0.229	-0.824	-0.245
x_2	0.816	0.229	1.000	-0.139	-0.973
x_3	-0.535	-0.824	-0.139	1.000	0.030
x_4	-0.821	-0.245	-0.973	0.030	1.000

Table 3. Parameter Support for GME-S3 and GME-S4

Parameter	Parameter Support (M=5)	Parameter Support (M=7)	Prior Mean
β_0 (constant)	$z_0' = \{-75 \ -37.5 \ 0 \ 37.5 \ 75\}$	$z_0' = \{-75 \ -50 \ -25 \ 0 \ 25 \ 50 \ 75\}$	0
β_1 x_1	$z_1' = \{-4 \ -2 \ 0 \ 2 \ 4\}$	$z_1' = \{-4 \ -2.67 \ -1.33 \ 0 \ 1.33 \ 2.67 \ 4\}$	0
β_2 x_2	$z_2' = \{-2 \ -1 \ 0 \ 1 \ 2\}$	$z_2' = \{-2 \ -1.33 \ -0.66 \ 0 \ 0.66 \ 1.33 \ 2\}$	0
β_3 x_3	$z_3' = \{-2 \ -1 \ 0 \ 1 \ 2\}$	$z_3' = \{-2 \ -1.33 \ -0.66 \ 0 \ 0.66 \ 1.33 \ 2\}$	0
β_4 x_4	$z_4' = \{-2 \ -1 \ 0 \ 1 \ 2\}$	$z_4' = \{-2 \ -1.33 \ -0.66 \ 0 \ 0.66 \ 1.33 \ 2\}$	0

the sample standard deviation and $\overline{x}$ is the absolute value of the sample mean.

Table **2** gives the sample correlation matrix.

As can be seen from Table **2** the sample correlation coefficients are: $r_{x_1x_2} = 0,229$, $r_{x_1x_3} = -0,824$, $r_{x_1x_4} = -0,245$, $r_{x_2x_3} = -0.139$, $r_{x_2x_4} = -0.973$ and $r_{x_3x_4} = 0.030$ [18] notes correlation and collinearity are not the same. It is possible to have data that are subject to collinearity when the correlation between pairs of explanatory variables is low.

3.1. Prior Mean of Each Parameter Is Zero

For both the coefficient and error supports, we need to select the number of points, M and J, respectively.

GME- First consider the parameter support. We choose M=5 support points for each parameter. First consider the parameter support. Here we specify wide bounds of $\left[-75,75\right]$ (for the intercept term, $\left[-4,4\right]$ for β_1 and…) $\left[-2,2\right]$ for the other coefficients. The supports are symmetric about zero so prior mean of each parameter is zero. Table **3** gives the parameter support for GME.

We also need the error support for the GME estimates. We initially construct the GME estimator with error bounds $\pm 3\sigma$. However, since σ is not known we must replace it with a consistent estimator. We considered two possible estimates for σ: 1) $\hat{\sigma} = 2.446$ from the *OLS* regression, 2) The sample standard deviation of y is $s = 15.044$. We used the more conservative value s=15.044 of the sample standard deviation of *y*. Using the sample standard deviation of *y*, 3σ and 4σ - rule results are:

For 3σ and M=5 $v' = \{-46 \ -23 \ 0 \ 23 \ 46\}$

For 4σ and M=5 $v' = \{-60 \ -30 \ 0 \ 30 \ 60\}$

For 3σ and M=7 $v' = \{-46 \ -30.67 \ -15.33 \ 0 \ 15.33 \ 30.67 \ 46\}$

For 4σ and M=7 $v' = \{-60 \ -40 \ -20 \ 0 \ 20 \ 40 \ 60\}$

Table **4** gives point estimates for the cement data using OLS, and GME estimators, where S3 and S4 refer to the use of a 3σ and 4σ - rule respectively.

We have obtained that the results are not improved by using more than 5 support points. The results show that the GME estimates differ much from OLS in terms of the signs and magnitudes of the estimates. The GME estimates for β_0 and β_1 are smaller than the OLS estimate, while the GME estimates for β_2, β_3 and β_4 are larger in magnitude than the OLS estimates.

Using Table **4**, we have

$$\hat{\beta}'_{OLS}\hat{\beta}_{OLS} = 3897.08 > \hat{\beta}'_{GME1S3M7}\hat{\beta}_{GME1S3M7} = 703.119$$
$$> \hat{\beta}'_{GME1S3M5}\hat{\beta}_{GME1S3M5} = 699.790 > \hat{\beta}'_{GME1S4M7}\hat{\beta}_{GME1S4M7} = 686.198$$
$$> \hat{\beta}'_{GME1S4M5}\hat{\beta}_{GME1S4M5} = 682.648 .$$

On the other hand, mean squared error (*mse*) values of the estimators are estimated by the bootstrap method. The idea behind the bootstrap is as follows: Let $\tilde{\beta}$ be an estimate of the coefficient vector β, $\tilde{\sigma}^2$ denotes the estimated residual variance and $\tilde{e}$ denotes the standardized estimated residual

Table 4. OLS, and GME Estimates (n=13 Observations)

Variable and Parameter	OLS	M=5 GME-S3	M=5 GME-S4	M=7 GME-S3	M=7 GME-S4
constant β_0	62.405	26.398 (0.388)	26.077 (0.366)	26.461 (0.380)	26.145 (0.360)
x_1 β_1	1.551	1.368 (0.191)	1.247 (0.193)	1.369 (0.190)	1.248 (0.192)
x_2 β_2	0.510	0.972 (0.027)	0.978 (0.024)	0.970 (0.027)	0.976 (0.024)
x_3 β_3	0.102	0.162 (0.039)	0.147 (0.029)	0.163 (0.039)	0.148 (0.029)
x_4 β_4	-0.144	0.305 (0.025)	0.324 (0.023)	0.304 (0.025)	0.324 (0.023)

Bootstrap standard error estimates are given in parenthesis.

Table 5. Parameter Support for R1GME (Sign Restrictions Only)

Variable and Parameter	Parameter Support	Prior Mean
constant β_0	$z_0' = \{0.0 \quad 30.0 \quad 60.0 \quad 90.0 \quad 120.0\}$	60.0
x_1 β_1	$z_1' = \{0 \quad 0.5 \quad 1 \quad 1.5 \quad 2\}$	1
x_2 β_2	$z_2' = \{0 \quad 0.5 \quad 1 \quad 1.5 \quad 2\}$	1
x_3 β_3	$z_3' = \{0 \quad 0.1 \quad 0.2 \quad 0.3 \quad 0.4\}$	0.2
x_4 β_4	$z_4' = \{0.0 \quad 0.2 \quad 0.4 \quad 0.6 \quad 0.8\}$	0.4

vector, which is obtained using $\tilde{\beta}$ in (2). For the j-th bootstrap replication, $\tilde{e}_j^*$, a random sample of $\tilde{e}$ with replacement is taken and multiplied by $\tilde{\sigma}$. Using this $\tilde{e}_j^*$, $\tilde{\beta}$ and the X matrix; y_j^* is computed from the equation (2). For this bootstrap sample y_j^* and design matrix *X*, $\tilde{\beta}_j^*$, j-th bootstrap estimate of β is computed. This procedure is replicated for 10000 times to obtain the empirical sampling distribution of $\tilde{\beta}$. After then, the variance and the expected value of each element in $\tilde{\beta}$ is estimated with this sampling distribution. Scalar mse of $\tilde{\beta}$ is estimated using bootstrap variances and expected values in $\mathrm{mse}(\tilde{\beta}) = \mathrm{trace}\{\mathrm{cov}(\tilde{\beta})\} + \mathrm{bias}(\tilde{\beta})'\,\mathrm{bias}(\tilde{\beta})$. On the other hand, since the true value of β is not known, we used OLS estimate of β to estimate bias as $\hat{\mathrm{bias}}(\tilde{\beta}) = \hat{E}(\tilde{\beta}) - \tilde{\beta}_{OLS}$, where $\hat{E}(\tilde{\beta})$ is the bootstrap estimate of the expected value. Estimated mse values are as follows

$$\hat{\mathrm{mse}}(\hat{\beta}_{OLS}) = 4402.698 \; > \; \hat{\mathrm{mse}}(\hat{\beta}_{GME1S4M5}) = 1326.741 \; >$$

$$\hat{\mathrm{mse}}(\hat{\beta}_{GME1S4M7}) = 1320.869 > \; \hat{\mathrm{mse}}(\hat{\beta}_{GME1S3M5}) = 1306.551 \; >$$

$$\hat{\mathrm{mse}}(\hat{\beta}_{GME1S3M7}) = 1300.118 .$$

Golan *et al.* [3] show that the GME estimator has a lower mean squared error than the OLS, Ridge and RLS estimators in several sampling experiments, particularly when the data exhibit a high degree of collinearity.

We now modify our parameter support to account for the expected signs of the coefficients.

3.2. Imposing Sign Restrictions of Parameters for GME

R1GME- If we expect the coefficients to have positive effect, we modify the parameter support such that the prior mean of their coefficients is positive. We impose correct sign restrictions for all the parameters. The prior mean is simply the value the parameters are shrunk toward, not a binding restriction. When we restrict the parameter estimates to be positive we obviously must specify a positive prior mean. For the R1GME estimator we specify the parameter support for β_1 as $\begin{bmatrix} 0 & 0.5 & 1 & 1.5 & 2 \end{bmatrix}$ which has a prior mean 1. We choose *M*=5 support points for each parameter. Table **5** gives the parameter support for R1GME. When we restrict the parameter estimates to be positive we must specify a positive prior mean.

We again specify error supports using $\pm 3\sigma$ and $\pm 4\sigma$.,

3.3. Imposing Cross Parameter Restrictions for GME

R2GME-We now estimate the Hald data (cement data) model with sign restrictions plus the additional restriction that $\beta_1 \geq \beta_4 > 0$. This is just a made up restriction to illustrate

Table 6 Estimates with a Single Cross-Parameter Restriction ($\beta_0 > 0, \beta_1 > 0, \beta_2 > 0, \beta_3 > 0, \beta_4 > 0$)

Variable and Parameter	OLS	IRLS	R1GME-S3	R1GME-S4
Constant β_0	62.405	48.194	34.852 (1.961)	36.696 (1.856)
x_1 β_1	1.551	1.696	1.075 (0.044)	1.033 (0.029)
x_2 β_2	0.510	0.657	0.871 (0.032)	0.844 (0.026)
x_3 β_3	0.102	0.250	0.193 (0.002)	0.194 (0.001)
x_4 β_4	-0.144	0.000	0.319 (0.013)	0.334 (0.012)

Bootstrap standard error estimates are given in parenthesis.

Table 7. Parameter Support for R2GME (Single Restriction)

Variable and Parameter	Parameter Support	Prior Mean
constant β_0	$z_0' = \{0.0 \quad 30.0 \quad 60.0 \quad 90.0 \quad 120.0\}$	60.0
x_1 β_1	$z_1' = \{0 \quad 0.3 \quad 0.6 \ 0.9 \ 1.2\}$	$\hat{\beta}_4 + 0.6$
x_2 β_2	$z_2' = \{0 \quad 0.5 \quad 1.0 \ 1.5 \ 2.0\}$	1
x_3 β_3	$z_3' = \{0 \quad 0.1 \quad 0.2 \quad 0.3 \quad 0.4\}$	0.2
x_4 β_4	$z_4' = \{0 \ 0.2 \ 0.4 \quad 0.6 \ 0.8\}$	0.4

how to impose cross restrictions. We include this restriction to illustrate the use of a non-block diagonal parameter support matrix. We specify the restricted GME support matrix using

$$\begin{bmatrix} \beta_1 \\ \beta_4 \end{bmatrix} = Z.\begin{bmatrix} p_1 \\ p_4 \end{bmatrix} = \begin{bmatrix} z_1' & z_4' \\ 0 & z_4' \end{bmatrix}.\begin{bmatrix} p_1 \\ p_4 \end{bmatrix}$$

where Z is the $2 \times 2M$ sub matrix of support points for β_1 and β_4, and p_1 and p_4 represent the unknown probabilities associated with support points for tricalcium aluminate (X_1) and β - dicalcium silicate (X_4), respectively. For this set of restrictions, we specify the parameter supports for R2GME as follows:

R2GME: Both $\hat{\beta}_1$ and $\hat{\beta}_4$ are constrained to be positive because of the support of β_4. The estimate for β_1, $\hat{\beta}_1 = \hat{\beta}_4 + z_1' p_1 \geq \hat{\beta}_4$. The prior mean of the estimate for β_1 is $\hat{\beta}_4 + 0.6$

We again specify error supports using $\pm 3\sigma$ and $\pm 4\sigma$.

Finally, we include an example imposing multiple inequality restrictions through the parameter support matrix. In addition to the parameter sign restrictions, we impose the

restrictions $\beta_1 \geq \beta_4$ and $\beta_2 \geq \beta_1 - \beta_3$. Using the matrix notation we have:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

To impose these restrictions, we specify the R3GME support matrix as ($\beta_0 > 0, \beta_1 > 0, \beta_2 > 0, \beta_3 > 0, \beta_4 > 0$)

$$\beta = Zp = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} z_0' & 0 & 0 & 0 & 0 \\ 0 & z_1' & 0 & 0 & z_4' \\ 0 & z_1' & z_2' & -z_3' & z_4' \\ 0 & 0 & 0 & z_3' & 0 \\ 0 & 0 & 0 & 0 & z_4' \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix}$$

The restrictions are satisfied as

$$\hat{\beta}_1 = z_1' p_1 + z_4' p_4 = \hat{\beta}_4 + z_1' p_1 \geq \hat{\beta}_4,$$

$$\hat{\beta}_2 = z_1' p_1 + z_2' p_2 - z_3' p_3 + z_4' p_4 = (\hat{\beta}_1 - z_4' p_4) + \geq \hat{\beta}_1 - \hat{\beta}_3.$$
$$z_2' p_2 - \hat{\beta}_3 + z_4' p_4 = \hat{\beta}_1 - \hat{\beta}_3 + z_2' p_2$$

Table 8. Estimates with a Single Cross-Parameter Restriction

Variable and Parameter	OLS	IRLS	R2GME-S3	R2GME-S4
constant β_0	62.405	48.194	34.744 (2.023)	36.786 (1.873)
x_1 β_1	1.551	1.696	0.966 (0.030)	0.956 (0.021)
x_2 β_2	0.510	0.657	0.883 (0.035)	0.850 (0.028)
x_3 β_3	0.102	0.250	0.192 (0.002)	0.194 (0.002)
x_4 β_4	-0.144	0.000	0.331 (0.015)	0.340 (0.013)

Bootstrap standard error estimates are given in parenthesis.

Table 9. Parameter Support for R3GME (Multiple Restrictions)

Variable and Parameter	Parameter Support	Prior mean
constant β_0	$z'_0 = \{0 \quad 30 \quad 60 \quad 90 \quad 120\}$	60
x_1 β_1	$z'_1 = \{0 \quad 0.3 \quad 0.6 \quad 0.9 \quad 1.2\}$	$\hat{\beta}_4 + 0.6$
x_2 β_2	$z'_2 = \{0 \quad 0.5 \quad 1.0 \quad 1.5 \quad 2.0\}$	$\hat{\beta}_1 - \hat{\beta}_3 + 1$
x_3 β_3	$z'_3 = \{0 \quad 0.1 \quad 0.2 \quad 0.3 \quad 0.4\}$	0.2
x_4 β_4	$z'_4 = \{0 \quad 0.2 \quad 0.4 \quad 0.6 \quad 0.8\}$	0.4

Table 10. Estimates with Multiple Inequality Restriction (N=13 Observations)

Variable and Parameter	OLS	IRLS	R3GME-S3	R3GME-S4
constant β_0	62.405	0	25.531 (1.380)	25.617 (1.116)
x_1 β_1	1.551	1.911	0.643 (0.026)	0.633 (0.018)
x_2 β_2	0.510	1.202	1.182 (0.049)	1.211 (0.045)
x_3 β_3	0.102	0.710	0.160 (0.004)	0.164 (0.003)
x_4 β_4	-0.144	0.491	0.272 (0.011)	0.265 (0.008)

For this set of restrictions, we specify the parameter supports for R3GME as follows:

R3GME- Both $\hat{\beta}_1$ and $\hat{\beta}_4$ are constrained to be positive due to the support for β_4. The prior mean of the estimate for β_1 is $\hat{\beta}_4 + 0.6$. Table **9** gives the parameter support for R3GME.

Table **10** gives the OLS, IRLS and R3GME estimates for the parameters.

4. CONCLUSIONS

The Generalized Maximum Entropy estimator is a robust estimator that is resistant to multicollinearity. This paper applies maximum entropy estimation to the Portland cement dataset. We discuss how to specify the parameter and error support matrices for the GME estimator. In our model, the results show that the GME estimates differ much from OLS in terms of the signs and magnitudes of the estimates. We illustrate the GME optimization method, through the GME

parameter support matrix when there are linear inequality restrictions and multiple restrictions.

We have seen that estimation is not improved by choosing more than about five support points. Varying the width of the error bounds has a larger impact on the estimates. For the Portland cement data example, we conclude that error bounds of $\pm 4\hat{\sigma}$ are preferred over error bounds of $\pm 3\hat{\sigma}$. The GME estimator is a shrinkage estimator where the parameter estimates are shrunk towards the prior mean, which is based on nonsample information. We demonstrate how to impose restrictions on the GME estimator using a simple example of sign restrictions and new parameter support matrices.

REFERENCES

[1] C. E. Shannon. "A mathematical theory of communication", *Bell. Syst. Tech. J.,* vol. 27, pp. 379-423, 1948.

[2] E. T. Jaynes. "Information theory and statistical mechanics", *Phys. Rev.,* vol. 106, pp. 620-630, 1957.

[3] A. Golan, G. Judge, and D. Miller, *Maximum Entropy Econometrics*, John Wiley: New York, 1996.

[4] R. C. Campbell, and R. C. Hill, "A monte carlo study of the effect of design characteristics on the inequality restricted maximum entropy estimator", *Rev. App. Econ.,* vol. 1*,* pp. 53-84, 2005.

[5] R. C. Campbell, and R.C. Hill, "Imposing parameter inequality restrictions using the principle of maximum entropy", *J. Stat. Comput. Simul.,* vol. 76, pp. 985-1000, 2006.

[6] X. Wu, "A weighted generalized maximum entropy estimator with a data-driven Weight", *Entropy,* vol. 11, pp. 917-930, 2009.

[7] I. Fraser, "An application of maximum entropy estimation: The demand for meat in the United Kingdom", *App. Econ.*, vol. 32, pp. 45-59, 2000.

[8] E. Z. Shen, and J.M. Perlof, "Maximum entropy and Bayesian approaches to the ratio problem", *J. Appl. Econ.*, vol. 104, pp. 289-313, 2001.

[9] H. K. Mishra, " Estimation under multicollinearity application of restricted Liu and maximum entropy estimators to the Portland cement dataset," MPRA Paper 1809, University Library of Munich, Germany, 2004.

[10] M. R. Caputo, and Q. Paris, "Comparative statics of the generalized maximum entropy estimator of the general linear model", *EJOR.,* vol. *185* , pp. 195-203, 2008.

[11] F. Pukelsheim, "The three sigma rule", *Am. Statistician.*, vol. 48, pp. 88-91, 1994.

[12] A. Golan, and J.M. Perloff, "Comparison of maximum entropy and higher-order entropy Estimators", *J. Econ.*, vol. 107, pp. 195-211, 2002.

[13] H. Woods, H.H. Steinor, and H.R. Starke, "Effect of composition of Portland cement on heat evolved during hardening", *Ind. Eng. Chem. Res.*, vol. 24, pp. 1207-1214, 1932.

[14] A. Hald. *Statistical Theory with Engineering Applications*. John Wiley, New York: 1952.

[15] J. W. Gorman, and R.J. Toman, "Selection of variables for fitting equations to data". *Technometrics*, vol. 8, pp. 27-51, 1966.

[16] C. Daniel, and F.S. Wood, *Fitting Equations to Data: Computer Analysis of Multifactor Data*. New York: John Wiley, 1980.

[17] S. Kaçıranlar, S. Sakallıoğlu, F. Akdeniz, G.P.H. Styan, and H.J. Werner, "A new biased estimator in linear regression and a detailed analysis of the widely-analysed data set on Portland cement" *Sankhya.*, vol. 61, Series B, 3, pp. 443-459, 1999.

[18] D. A. Belsley, *Conditioning Diagnostics: Collinearity and Weak Data in Regression;* John Wiley: New York, 1991.

.