Maximum entropy estimation in economic models with linear inequality restrictions

Randall C. Campbell*, R. Carter Hill

Department of Economics, Louisiana State University, Baton Rouge, LA 70803,USA

Abstract

In this paper, we use maximum entropy to estimate the parameters in an economic model. We demonstrate the use of the generalized maximum entropy (GME) estimator, describe how to specify the GME parameter support matrix, and examine the sensitivity of GME estimates to the parameter and error bounds. We impose binding inequality restrictions through the GME parameter support matrix and develop a more general parameter support matrix that allows us to impose a broader set of restrictions than is possible under the traditional formulation. Bootstrapping is used to obtain confidence intervals and examine the precision of the GME estimator.

JEL Classification: C13; C14; C20; C49; C51

Keywords: Maximum Entropy; Generalized Maximum Entropy; Linear Inequality Restrictions; Bootstrapping

* Corresponding author. Tel.: 1-225-578-3782; fax: 1-225-578-3807.
E-mail address: camprr@msn.com

1. Introduction

This paper examines the generalized maximum entropy (GME) estimator in the general linear model (GLM). Since GME estimation requires us to specify bounds for the parameters, we present an economic application and 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. The GME estimates are compared to least squares estimates, both with and without inequality restrictions placed on the parameters. Finally, we use the bootstrap to obtain confidence intervals and examine the precision of the GME estimator.

We use the GME estimator developed by Golan, Judge, and Miller (1996, pp. 85-89) [hereinafter GJM]. GJM show that the GME estimator has lower risk than both the OLS and IRLS estimators in several sampling experiments (GJM, pp. 133-144), particularly when the data exhibit a high degree of collinearity. GJM specify a block diagonal parameter support matrix for the GME estimator, which allows us to impose single parameter restrictions such as $\beta_i > 0$. Applications of single parameter restrictions on the GME estimator may be found in Fraser (2000) and Shen and Perloff (2001). We impose binding single parameter restrictions through the parameter support matrix and, in addition, we specify a more general parameter support matrix that is not block diagonal and which allows us to impose multiple parameter restrictions such as $\beta_i > \beta_j$ and $\beta_1 + \beta_2 + \beta_3 < 1$. Specifying a non-block diagonal support parameter matrix provides a relatively simple way to impose several restrictions that might be encountered in practice.

We show that GME is a feasible approach to estimating linear regression models. All of our GME estimates take the same signs and have roughly the same magnitude as OLS and IRLS estimates. In addition, our bootstrap results show that the sampling precision is better for the GME estimator than for the OLS and IRLS estimators.

Section 2 discusses GME estimation in the linear regression model. In Section 3, we describe how to impose inequality restrictions through the GME parameter support matrix and present a non-block diagonal support matrix that allows us to impose restrictions that are not possible under the traditional support matrix. In Section 4, we estimate an economic model using GME, both with and without binding parameter inequality restrictions, and compare the GME estimates to least squares estimates. Section 5 presents GME confidence intervals obtained using a bootstrap. Section 6 concludes the paper.

2. Generalized maximum entropy estimation in the general linear model

GJM (1996, Ch. 6) use GME to jointly estimate the unknown parameters and the unknown errors in the GLM.[1] We write the GLM in matrix form as

$$y = X\beta + e, \tag{1}$$

where y is an $N \times 1$ vector of sample observations on the dependent variable, X is an $N \times K$ matrix of explanatory variables, e is an $N \times 1$ vector of unknown errors, and β is a $K \times 1$ vector of unknown parameters.

Jaynes (1957a, 1957b) shows that maximum entropy allows us to estimate the unknown probabilities in a discrete probability distribution.[2] GJM 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. We specify a set of support points for each unknown parameter and error and use maximum entropy to estimate the unknown probabilities associated with the support points. Hence we must assume that both the unknown parameters and the unknown errors may be bounded *a priori*. Let z_{k1} be the smallest possible value of β_k and z_{k2} be the largest possible value of β_k. Then, for each parameter, β_k, there exists $p_k \in [0,1]$ such that

$$\beta_k = p_k z_{k1} + (1 - p_k) z_{k2} = \begin{bmatrix} z_{k1} & z_{k2} \end{bmatrix} \begin{bmatrix} p_k \\ 1 - p_k \end{bmatrix}. \tag{2}$$

The parameter support is based on prior information or economic theory. For example, we might specify boundaries of $z_{k1} = 0$ and $z_{k2} = 1$ when estimating the marginal propensity to consume.

---

[1] Our GME estimator corresponds to GJM's GME-D estimator on p. 86.

[2] The ME distribution is the most uniform distribution compatible with the prior information.

However, specifying the largest and smallest possible values for each variable is not an easy task since economic theory does not usually provide this information.[3]

Define a matrix consisting of $M \geq 2$ support points for each parameter, which may or may not be symmetric about zero and which bound the unknown parameters. Let z_k be the $M \times 1$ support vector for the k^{th} parameter and let p_k be the associated $M \times 1$ vector of probabilities or weights on these support points. We write the unknown parameter vector, β, as

$$\beta = Zp = \begin{bmatrix} z_1' & 0 & \cdots & 0 \\ 0 & z_2' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_K' \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \end{bmatrix}, \tag{3}$$

where β is a $K \times 1$ vector of unknown parameters, 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. This is the traditional GME parameter support matrix, which is block diagonal so the support points for any parameter do not directly impact the other parameter estimates.

Similarly, for the unknown errors, let v_{i1} be the smallest possible value of e_i and v_{i2} be the largest possible value of e_i. For each random error, e_i, there exists $w \in [0,1]$ such that

$$e_i = w_i v_{i1} + (1 - w_i) v_{i2} = \begin{bmatrix} v_{i1} & v_{i2} \end{bmatrix} \begin{bmatrix} w_i \\ 1 - w_i \end{bmatrix}. \tag{4}$$

Placing boundaries on the unknown errors is difficult in practice. Following Pukelsheim (1994), GJM suggest setting the error bounds as $v_{i1} = -3\sigma$ and $v_{i2} = 3\sigma$, where σ is the standard deviation of e. To use this rule we must either know or estimate the value of σ.

Define a set of $J \geq 2$ support points for each error, which are symmetric about zero and which bound the unknown errors. Let v_i be the $J \times 1$ support vector for the i^{th} error and let w_i be the associated $J \times 1$ vector of weights on these support points. We write the unknown error vector as

[3] When we do not have good prior information about a parameter we specify a wide set of parameter bounds centered about zero. GJM (1996, p. 138) discuss this point and conclude that the consequences of specifying a wide parameter support are small in terms of risk measures.

$$e = Vw = \begin{bmatrix} v_1' & 0 & \cdots & 0 \\ 0 & v_2' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_N' \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}, \tag{5}$$

where e is an $N \times 1$ vector of random errors, V is an $N \times NJ$ matrix of support points, and w is an $NJ \times 1$ vector of unknown weights such that $w_{ij} > 0$ and $w_i' i_J = 1$ for all i. We write the reparameterized model in matrix form as

$$y = XZp + Vw, \tag{6}$$

where y, X, Z, and V are known and we estimate the unknown p and w vectors using maximum entropy. 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.

Jaynes (1957a) shows that entropy is additive for independent sources of uncertainty. Assuming the unknown weights on the parameter and the error supports for the GLM are independent, we jointly estimate the unknown parameters and errors by solving the constrained optimization problem

$$\max H(p, w) = -p' \ln(p) - w' \ln(w) \tag{9}$$

subject to

$$y = XZp + Vw \tag{10}$$

$$(I_K \otimes i_M')p = i_K \tag{11}$$

$$(I_N \otimes i_J')w = i_N, \tag{12}$$

where $\otimes$ is the Kronecker product. Equation (10) is a data constraint and equations (11) and (12) are additivity constraints, which require that the probabilities sum to one for each of the K parameters and each of the N errors.

The solutions to the GME constrained optimization problem are

$$\hat{p}_{km} = \frac{\exp(z_{km} x_k' \hat{\lambda})}{\sum_{m=1}^{M} \exp(z_{km} x_k' \hat{\lambda})} \tag{13}$$

and

$$\hat{w}_{nj} = \frac{\exp(v_{nj} \hat{\lambda}_n)}{\sum_{j=1}^{J} \exp(v_{nj} \hat{\lambda}_n)}, \tag{14}$$

where x_k is the $N \times 1$ vector of observations for the k^{th} explanatory variable and λ is an $N \times 1$ vector of Lagrange multipliers for the data constraint. Thus, the GME parameter estimates are a function of the Lagrange multipliers for the data constraint, the support points placed on the parameters *a priori*, and the sample data. The GME error estimates are a function of the Lagrange multipliers for the data constraint and the support points placed on the errors *a priori*.

Pre-multiplying the GME data constraint (6) by X' yields

$$X'y = X'XZp + X'Vw. \tag{15}$$

Substituting the optimal probabilities, $\hat{p}$, and error weights, $\hat{w}$, we obtain

$$X'y = X'XZ\hat{p} + X'V\hat{w} = X'X\hat{\beta} + X'\hat{e}.$$

The GME parameter estimates are given by

$$\hat{\beta}_{GME} = (X'X)^{-1} X'y - (X'X)^{-1} X'\hat{e} = (X'X)^{-1} X'(y - \hat{e}). \tag{16}$$

Thus, GME minimizes the SSE for a fitted regression line that passes through the mean of $y - \hat{e}$ rather than through the mean of y. As $\hat{e} \to 0$ (narrower error bounds), the GME estimator grows closer to the OLS estimator. As $\hat{e} \to y$ (wider error bounds), the GME estimator goes to zero.[4]

In the linear regression problem, the GME estimator is a shrinkage estimator similar to the Stein-like and empirical Bayes estimators described, for example, by Judge, Hill, and Bock (1990). GME selects the most uniform probability distribution compatible with the constraints, which are

[4] Assuming the parameter support is symmetric about zero. The GME estimator is actually shrunk toward its prior mean, which may or may not be zero, as the error bounds grow large.

based on prior information. GME shrinks the parameter estimates towards the expected value of the parameter support, which is specified *a priori*. The expected value of the parameter support is equal to the sum of the support points multiplied by the associated prior distribution, and is known as the prior mean of the unknown parameters. For example, suppose we specify a parameter support that is symmetric about zero. If the prior probability distribution is uniform the prior mean of the parameter support is equal to zero (since $\hat{\beta}_k = z_k' \hat{p}_k$).

3. GME estimation of an economic model

In this section, we estimate an economic model of poverty rates and their determinants. Applications of GME estimation in linear regression models can be found in Fraser (2000), Shen and Perloff (2001), Preckel (2001), and Miller and Plantinga (1999). Golan, Judge, and Perloff (1997) and Golan, Perloff, and Shen (2001) use GME to estimate a censored regression model. While these papers apply GME estimation to various economics problems, this paper provides a general discussion of how one might select the GME parameter and error supports. We examine the sensitivity of GME estimates to the chosen parameter and error supports. We find that GME estimates vary a little in terms of magnitude, but that the signs of the parameter estimates do not change as we vary the prior information. This is generally consistent with other research, although Fraser's results exhibit a surprisingly high degree of variation in response to relatively small changes in the parameter and error supports.

Our data set is taken from Ramanthan (2002, Data 7-6, p. 653) and consists of poverty rates and their determinants across California counties. The data set contains both 1980 Census data and 1990 Census data for 58 counties, a total of 116 observations. The dependent variable in our model is percentage of families with income below the poverty level (POV_t). The explanatory variables are average household size ($HHSZ_t$), percentage unemployment rate ($UNEMP_t$), percentage of population age 25 and over with high school degree only (HS_t), percentage of population age 25 and over that completed 4 or more years of college ($COLL_t$), median household income

($MEDINC_t$), and a dummy variable ($D90_t$) that equals one for the 1990 Census and zero for the 1980 Census. We estimate the following model

$$POV_t = \beta_1 + HHSZ_t \cdot \beta_2 + UNEMP_t \cdot \beta_3 + HS_t \cdot \beta_4 + COLL_t \cdot \beta_5$$
$$+ MEDINC_t \cdot \beta_6 + D90_t \cdot \beta_7, \qquad t = 1, \ldots, T \tag{17}$$

Table 1 gives summary statistics for the poverty data. The sample coefficient of variation is defined as $CV_x = s(x)/\bar{x}$, where $s(x)$ is the sample standard deviation of x and $\bar{x}$ is the sample mean of x.

Table 1. Summary Statistics for Poverty Data (N =116 Observations)

Variable	Mean	Min.	Max.	Standard Deviation	Coefficient of Variation
POV	9.51	3.00	20.80	3.32	0.3486
HHSZ	2.92	2.29	3.73	0.31	0.1063
UNEMP	9.62	3.50	21.30	3.63	0.3775
HS	56.78	41.30	68.50	5.96	0.1050
COLL	17.72	9.00	44.00	7.08	0.3994
MEDINC	27.29	13.52	59.15	10.23	0.3748
D90	0.50	0.00	1.00	0.50	1.0043

Using OLS, the estimated regression function (with standard errors in parentheses) is:

$$\underset{}{P\hat{O}V} = \underset{(5.53)}{21.659} + \underset{(1.16)}{1.804}\,HHSZ + \underset{(0.06)}{0.076}\,UNEMP - \underset{(0.04)}{0.201}\,HS + \underset{(0.05)}{0.021}\,COLL - \underset{(0.05)}{0.416}\,MEDINC + \underset{(1.04)}{8.504}\,D90$$

$R^2 = 0.746$, $\quad \hat{\sigma}^2 = 1.717$, $\quad F(6,109) = 53.307$

All the estimates take the expected signs except the coefficient on *COLL* since we expect the percentage of college-educated individuals to have a negative effect on poverty rates. We will impose this restriction in Section 4.

We now estimate the model using GME. Because we must specify support matrices for the unknown parameters and errors, there is no single set of GME estimates. As shown in equations (13) and (14), the GME estimates depend on the supports. We specify different parameter and error supports to examine the sensitivity of the GME estimates to the specification of priors. First, consider the parameter support. For this problem, the dependent variable is a percentage so each parameter must be between –100 and 100. Because the effect on the poverty rate of a unit change in any one variable is certainly much smaller then 100% we impose somewhat narrower bounds. We specify three models, denoted GME1, GME2, and GME3 as follows:

- GME1 – Here we specify wide bounds of [-50,50] for the intercept and relatively wide bounds of [-20,20] for the other coefficients. The supports are symmetric about zero so the prior mean of each parameter is zero. Here we are assuming that we have very little prior information about each parameter so we specify a relatively wide support with a prior mean of zero. Table 2 gives the parameter support for GME1.

Table 2. Parameter Support for GME1

Parameter	Parameter Support	Prior Mean
β_1 (constant)	$z_1' = \{-50 \quad -25 \quad 0 \quad 25 \quad 50\}$	0
$\beta_2 - \beta_7$	$z_k' = \{-20 \quad -10 \quad 0 \quad 10 \quad 20\}, \quad k = 2,\ldots,7$	0

- GME2 – We expect that a one percent change in *UNEMP*, *HS*, or *COLL* will not change the poverty rate by more than one or two percent in either direction, so we specify a narrow support for the coefficients of these variables. We also specify narrower supports for the coefficients of *HHSZ* and *MEDINC*. In general, we may specify wider bounds to indicate either a lack of good prior information or an expectation that the coefficient may be large. All parameters again have a prior mean of zero. Table 3 gives the parameter support for GME2.

Table 3. Parameter Support for GME2

Parameter	Parameter Support	Prior Mean
β_1 (constant)	$z_1' = \{-50 \quad -25 \quad 0 \quad 25 \quad 50\}$	0
β_2 (hhsz)	$z_2' = \{-10 \quad -5 \quad 0 \quad 5 \quad 10\}$	0
β_3 (unemp)	$z_3' = \{-2 \quad -1 \quad 0 \quad 1 \quad 2\}$	0
β_4 (hs)	$z_4' = \{-2 \quad -1 \quad 0 \quad 1 \quad 2\}$	0
β_5 (coll)	$z_5' = \{-2 \quad -1 \quad 0 \quad 1 \quad 2\}$	0
β_6 (medinc)	$z_6' = \{-10 \quad -5 \quad 0 \quad 5 \quad 10\}$	0

β_7 (d90)	$z_7' = \{-20 \quad -10 \quad 0 \quad 10 \quad 20\}$	0

- GME3 – We now modify our parameter support to account for the expected signs of the coefficients. We have no prior expectations about the signs of the coefficients for the intercept and *D90*. Since we expect *HHS*Z and *UNEMP* to have a positive effect on the poverty rate we modify the parameter support such that the prior mean of their coefficients is positive. Likewise, since we expect *HS, COLL,* and *MEDINC* to be inversely related to poverty rates we specify the parameter support such that the prior mean of their coefficients is negative. Table 4 gives the parameter support for GME3.

Table 4. Parameter Support for GME3

Parameter	Parameter Support	Prior Mean
β_1 (constant)	$z_1' = \{-50 \quad -25 \quad 0 \quad 25 \quad 50\}$	0
β_2 (hhsz)	$z_2' = \{-5 \quad 0 \quad 5 \quad 10 \quad 15\}$	5
β_3 (unemp)	$z_3' = \{-1 \quad 0 \quad 1 \quad 2 \quad 3\}$	1
β_4 (hs)	$z_4' = \{-3 \quad -2 \quad -1 \quad 0 \quad 1\}$	-1
β_5 (coll)	$z_5' = \{-3 \quad -2 \quad -1 \quad 0 \quad 1\}$	-1
β_6 (medinc)	$z_6' = \{-15 \quad -10 \quad -5 \quad 0 \quad 5\}$	-5
β_7 (d90)	$z_7' = \{-20 \quad -10 \quad 0 \quad 10 \quad 20\}$	0

In practice, we would choose a specification like GME3 that incorporates our prior beliefs about the magnitude and signs of each coefficient. Note that we do not constrain any of the coefficients to take a specific sign. The prior mean is simply the value the parameters are shrunk toward, not a binding restriction. We choose $M = 5$ support points for each parameter since GJM find that estimation is not improved by choosing more than about five support points.

We also vary the error support for our GME estimates. Following GJM, we initially construct the GME estimator with error bounds of $\pm 3\sigma$. However, since σ is unknown we must replace it with an estimate. We considered two possible estimates for σ: 1) $\hat{\sigma}$ from the OLS regression, which equals 1.72, and 2) the sample standard deviation of y, which equals 3.32. We obtained much better results using the more conservative value of the sample standard deviation of y. In fact, some of our programs did not converge when we used the smaller estimate for σ in specifying our error bounds.

Using the sample standard deviation of y, the 3σ - rule results in an error support of $\{-10 \quad -5 \quad 0 \quad 5 \quad 10\}$. As with the parameter support, we choose $J=5$ support points for each error. We also specify a wider set of error bounds, which yields parameter estimates that are shrunk more towards their prior mean. We follow a 4σ - rule for the second set of estimates and our error support is $\{-13 \quad -6.5 \quad 0 \quad 6.5 \quad 13\}$. We obtain GME1, GME2, and GME3 estimates using each error support and we refer to the estimates as GME1S3, GME2S3, GME3S3, GME1S4, GME2S4, and GME3S4, with S3 and S4 indicating the use of a 3σ or 4σ rule, respectively. Table 5 gives point estimates for the poverty data using OLS and our six different GME estimators.

Table 5. OLS and GME Estimates for Poverty Data (*N* =116 Observations)

			S3			S4		
Variable		OLS	GME1	GME2	GME3	GME1	GME2	GME3
POV	β_1	21.659	16.678	18.363	14.796	15.700	17.908	13.444
HHSZ	β_2	1.804	2.411	2.001	2.895	2.521	1.962	3.134
UNEMP	β_3	0.076	0.136	0.138	0.144	0.128	0.131	0.144
HS	β_4	-0.201	-0.168	-0.175	-0.164	-0.157	-0.166	-0.155
COLL	β_5	0.021	0.055	0.046	0.055	0.038	0.027	0.034
MEDINC	β_6	-0.416	-0.399	-0.393	-0.397	-0.385	-0.375	-0.378
D90	β_7	8.504	8.097	7.834	8.269	8.021	7.666	8.178

The results show that the GME estimates do not differ much from OLS in terms of the signs and magnitudes of the estimates. The signs of the coefficients are the same for all the alternative estimators. The GME estimates for β_1, β_4, β_6, and β_7 are smaller in magnitude than the OLS estimates while the GME estimates for β_2, β_3, and β_5 are larger than the OLS estimates. These results are consistent across all of our GME estimators, although it is not clear why GME estimates are larger than OLS for some variables and smaller than OLS for other variables.

The GME estimates do not vary a great deal as we change the parameter supports. Thus, the cost of using an uninformative prior (as in GME1) is small, which is consistent with the results obtained by GJM. As expected, when we specify wider error bounds (GME1S4, GME2S4, and GME3S4), the coefficients are generally shrunk more towards their prior means. In this case, we are placing more weight on the errors and allowing the probabilities associated with the parameter support to be more uniform.

The results indicate that for a single sample of data, GME estimates are reasonably close to OLS estimates. In addition, the GME estimates do not change much as we change the parameter support. An uninformative parameter support produces results that are generally consistent with both OLS estimates and with GME estimates obtained from a more informative parameter support. This is important because the GME estimator has been criticized on the grounds that it is not always easy to place bounds on the parameters. Section 5 examines the precision of the GME estimator through the use of a bootstrap.

4. Linear inequality restrictions

An economic researcher may have sign or other information about the parameters that can be expressed as a linear inequality restriction. Imposing this nonsample information on the least squares estimator yields the inequality restricted least squares (IRLS) estimator, which is biased, but dominates the OLS estimator, under a squared error loss measure, as long as the restrictions are true (Judge et al., 1988, pp. 822-825). Using the parameter support matrix we impose binding linear inequality restrictions on the GME estimator.

4.1 Imposing binding linear inequality restrictions on the GME estimator

Because each parameter must be bounded, the GME estimator always has inequality restrictions placed on the parameters. However, the bounds do not generally reflect specific prior information such as sign or other restrictions. We discuss how to impose sign and other restrictions through the parameter support matrix and we modify the parameter support matrix in a way that allows us to impose additional restrictions that might be encountered in practice. Below we discuss how different types of linear inequality restrictions are imposed through the GME parameter support matrix.

1) $\beta_1 > 0$

If we have nonsample information that $\beta_1 > 0$ we specify the support vector for β_1 to take only positive values such as $z_1' = [0 \quad 5 \quad 10 \quad 15]$, where z_1 is the $M \times 1$ parameter support vector for β_1. In this case the GME estimate is given by

$$\hat{\beta}_1 = 0\hat{p}_{11} + 5\hat{p}_{12} + 10\hat{p}_{13} + 15\hat{p}_{14} > 0\,, \tag{18}$$

since $\hat{p}_{1m} \geq 0$ for all M support points.

2) $\beta_1 > \beta_2$

To impose cross-parameter inequality restrictions in GME we specify a more general parameter support matrix that is not block diagonal. For the restriction $\beta_1 > \beta_2$, we specify the GME parameter support matrix as

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = Z^* \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} z_1' & z_2' \\ 0 & z_2' \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}, \tag{19}$$

where Z^* is the $2 \times 2M$ sub-matrix of support points for β_1 and β_2, and p_1 and p_2 represent the unknown probabilities associated with the support points for these parameters. Specify $z_1' = [0 \quad 5 \quad 10 \quad 15]$ (or any z_1' such that $z_{1m} \geq 0$ for all m) and the GME estimate for β_1 is[5]

$$\hat{\beta}_1 = 0\hat{p}_{11} + 5\hat{p}_{12} + 10\hat{p}_{13} + 15\hat{p}_{14} + \hat{\beta}_2 > \hat{\beta}_2\,. \tag{20}$$

We obtain the GME solution the same way as with the block diagonal matrix. However, the equation for the optimal probabilities is slightly more complicated when the parameter support matrix is not block diagonal. The solution for the optimal GME probabilities is now given by

[5] Note that the support for β_2 can include any set of values. Also, we can obtain the same restriction by specifying the support for β_2 to be strictly negative and letting the support for β_1 include any set of values.

$$\hat{p}_{km} = \frac{\exp(z_{km}x_1'\hat{\lambda} + z_{km}x_2'\hat{\lambda} + \ldots + z_{km}x_K'\hat{\lambda})}{\sum_{m=1}^{M}\exp(z_{km}x_1'\hat{\lambda} + z_{km}x_2'\hat{\lambda} + \ldots + z_{km}x_K'\hat{\lambda})} \tag{21}$$

where x_i is the $N \times 1$ vector of observations for the i^{th} explanatory variable ($i = 1,..., K$). Under the block diagonal parameter support matrix, all the cross-product terms drop out (those where $i \neq k$).

3) $\beta_1 + \beta_2 > c$, where c is any constant

In this case, we obtain the GME parameter estimates using

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = Z^* \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} z_1' & -z_2' \\ 0 & z_2' \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \end{bmatrix},$$

where Z^* is the $2 \times 2M$ sub-matrix of support points for β_1 and β_2, and p_1 and p_2 represent the unknown probabilities associated with the support points for these parameters. Specify $z_1' = [c \quad c+5 \quad c+10 \quad c+15]$ (or any z_1' such that $z_{1m} \geq c$ for all m) and the GME estimate for β_1 is

$$\hat{\beta}_1 = z_1'p_1 - z_2'p_2 = z_1'p_1 - \hat{\beta}_2,$$

which implies that

$$\hat{\beta}_1 + \hat{\beta}_2 = z_1'p_1 > c. \tag{22}$$

4) $\beta_1 + \beta_2 > \beta_3$

Finally, we impose a restriction involving three parameters, such as $\beta_1 + \beta_2 > \beta_3$. We specify the parameter support matrix as

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = Z^* \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} z_1' & -z_2' & z_3' \\ 0 & z_2' & 0 \\ 0 & 0 & z_3' \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix},$$

where Z^* is the $3\times 3M$ sub-matrix of support points for β_1, β_2, and β_3; p_1, p_2, and p_3 represent the unknown probabilities associated with the support points for these parameters. We specify z_1' such that all its elements are positive and obtain

$$\hat{\beta}_1 = z_1' p_1 - z_2' p_2 + z_3' p_3 = z_1' p_1 - \hat{\beta}_2 + \hat{\beta}_3 ,$$

which implies that

$$\hat{\beta}_1 + \hat{\beta}_2 = z_1' p_1 + \hat{\beta}_3 > \hat{\beta}_3 . \qquad (23)$$

4.2 Imposing non-binding inequality information

GJM (1996, pp. 140-142) consider the cost of imposing incorrect sign information about a parameter, but do not impose binding restrictions. They estimate a linear regression model using the generalized cross-entropy (GCE) estimator, which is used to specify discrete prior distributions that are not uniform. GJM specify parameter sign information by placing more prior weight on either the positive or negative parameter support points. They specify a parameter support given by $Z_k = [-10, 10]$ with prior weights of $[.375 \quad .625]$ and $[.625 \quad .375]$ and prior means equal to 2.5 and –2.5, respectively. These sign restrictions are not binding however since the parameter estimate is free to take any value between –10 and 10. We imposed this type of non-binding prior information on our GME3 estimator in section 3. For example, we specified a prior mean of –1 for coefficient on *COLL*, but the parameter estimate still came out positive.

In several sampling experiments, GJM find that risk is only slightly lower when the prior means are specified to take the correct signs compared to when they are specified to take incorrect signs. This is consistent with our results, which showed the parameter estimates did not change much in response to non-binding parameter restrictions.

4.3 GME estimation of an economic model with binding inequality restrictions

We now estimate the poverty rate model with binding inequality restrictions. This type of binding restriction has been applied by Fraser (2000) who imposes the restriction that own-price elasticities must be negative in a meat demand model and by Shen and Perloff (2001) who impose the restriction that the speed of adjustment parameter in a cobweb model be positive and less than one. In our model, the only coefficient that took an unexpected sign was the positive coefficient on *COLL* since we expect the percentage of college-educated adults to have a negative impact on poverty rates. For our first set of restricted estimates, we constrain the coefficient on *COLL* to be negative.

Suppose we wanted to impose the stronger restriction that the percentage of college-educated adults to has a larger negative impact on the poverty rate than percentage of high school educated adults. To illustrate the use of our non-block diagonal parameter support matrix we obtain a second set of restricted estimates in which we constrain $\beta_5 < \beta_4 < 0$.

For both sets of restricted estimates, we compare the IRLS estimates to the restricted GME (RGME) estimates. We consider the relatively wide bounds (representing little prior information) of GME1 and the narrower bounds of GME2. We do not re-estimate the GME3 model since it is just GME2 with non-binding restrictions. In each model, we constrain the coefficients for *HHSZ* and *UNEMP* to be positive and the coefficients for *HS, COLL,* and *MEDINC* to be negative. We specify the parameter supports for RGME1 and RGME2 as follows:

- RGME1 – The support for RGME1 maintains the wide bounds, representing little information about the magnitude of the parameters, as in GME1. Note that imposing sign restrictions changes the prior mean of the variables. For example, the support for *MEDINC* has the same lower bound as in GME1, but by removing the positive values we have changed the prior mean from 0 to -10. This can have a potentially large impact on our parameter estimates since they are shrunk towards the prior mean. Table 6 gives the parameter support for RGME1.

Table 6. Parameter Support for RGME1 (sign restrictions only)

Parameter	Parameter Support	Prior Mean
β_1 (constant)	$z_1' = \{-50 \quad -25 \quad 0 \quad 25 \quad 50\}$	0
β_2 (hhsz)	$z_2' = \{0 \quad 5 \quad 10 \quad 15 \quad 20\}$	10
β_3 (unemp)	$z_3' = \{0 \quad 5 \quad 10 \quad 15 \quad 20\}$	10
β_4 (hs)	$z_4' = \{-20 \quad -15 \quad -10 \quad -5 \quad 0\}$	-10
β_5 (coll)	$z_5' = \{-20 \quad -15 \quad -10 \quad -5 \quad 0\}$	-10
β_6 (medinc)	$z_6' = \{-20 \quad -15 \quad -10 \quad -5 \quad 0\}$	-10
β_7 (d90)	$z_7' = \{-20 \quad -10 \quad 0 \quad 10 \quad 20\}$	0

- RGME2 – Here we specify narrower bounds representing better nonsample information. The prior means of the parameters are smaller for RMGE2 than for RGME1. Table 7 gives the parameter support for RGME2.

Table 7. Parameter Support for RGME2 (sign restrictions only)

Parameter	Parameter Support	Prior Mean
β_1 (constant)	$z_1' = \{-50 \quad -25 \quad 0 \quad 25 \quad 50\}$	0
β_2 (hhsz)	$z_2' = \{0 \quad 2.5 \quad 5 \quad 7.5 \quad 10\}$	5
β_3 (unemp)	$z_3' = \{0 \quad 0.5 \quad 1 \quad 1.5 \quad 2\}$	1
β_4 (hs)	$z_4' = \{-2 \quad -1.5 \quad -1 \quad -0.5 \quad 0\}$	-1
β_5 (coll)	$z_5' = \{-2 \quad -1.5 \quad -1 \quad -0.5 \quad 0\}$	-1
β_6 (medinc)	$z_6' = \{-10 \quad -7.5 \quad -5 \quad -2.5 \quad 0\}$	-5
β_7 (d90)	$z_7' = \{-20 \quad -10 \quad 0 \quad 10 \quad 20\}$	0

We again specify error supports using bounds of $\pm 3\sigma$ and $\pm 4\sigma$. The restricted GME estimators are labeled RGME1S3, RGME2S3, RGME1S4, and RGME2S4, where S3 and S4 refer to the use of a 3σ and 4σ rule, respectively. Table 8 gives point estimates for the poverty data using IRLS and our four different RGME estimators.

Table 8. IRLS and RGME Estimates for Poverty Data with Sign Restrictions

Variable		OLS	IRLS	S3 RGME1	S3 RGME2	S4 RGME1	S4 RGME2
POV	β_1	21.659	23.139	12.765	16.223	9.289	15.103
HHSZ	β_2	1.804	1.518	3.672	2.862	4.471	3.190
UNEMP	β_3	0.076	0.072	0.113	0.184	0.110	0.214
HS	β_4	-0.201	-0.210	-0.160	-0.193	-0.144	-0.199
COLL	β_5	0.021	0	0	-0.032	0	-0.066
MEDINC	β_6	-0.416	-0.400	-0.363	-0.324	-0.366	-0.293
D90	β_7	8.504	8.189	8.171	7.295	8.647	7.105

The results show that signs of the parameter estimates are the same for IRLS and RGME. The RGME estimates are larger in magnitude than the GME estimates reported in Table 5. Recall that in our unrestricted GME1 and GME2 specification all of the parameter estimates had a prior mean of zero. Imposing binding sign restrictions in GME not only restricts the values the parameter estimates can take, but also effects the prior mean and therefore the magnitude of the parameter estimates. The change in the prior means can have a fairly large impact on the GME parameter estimates.

In the original model, the only coefficient that took a sign opposite our expectations was the coefficient on *COLL*. With IRLS, restricted coefficients that take the wrong sign in OLS will be equal to zero. Our results show that our RGME1 estimates, based on wide parameter bounds representing little prior information, are also equal to zero. However, the RGME2 estimates for the coefficient on *COLL* are negative. This is appealing since a coefficient estimate of zero is not consistent with the restriction that it be negative. The IRLS estimator and our RGME1 estimator simply eliminate from the model any variables whose coefficient estimates take incorrect signs. The RGME2 estimates are consistent with our belief that the percent of college-educated individuals has a negative effect on poverty rates.

We now estimate the poverty rate model with the sign restrictions plus the additional restriction that $\beta_5 < \beta_4$. We include this restriction to illustrate the use of a non-block diagonal parameter support matrix. We specify the GME support matrix using

$$\begin{bmatrix} \beta_4 \\ \beta_5 \end{bmatrix} = Z^* \begin{bmatrix} p_4 \\ p_5 \end{bmatrix} = \begin{bmatrix} z_4' & 0 \\ z_4' & z_5' \end{bmatrix} \cdot \begin{bmatrix} p_4 \\ p_5 \end{bmatrix},$$

where Z^* is the $2\times 2M$ sub-matrix of support points for β_4 and β_5, and p_4 and p_5 represent the unknown probabilities associated with the support points for *HS* and *COLL,* respectively. We specify the parameter supports for RGME1 and RGME2 as follows:

- RGME1 – Both *HS* and *COLL* are constrained to be negative due to the support for *HS* (z_4'). Note that the coefficient for *COLL*, $\hat{\beta}_5 = \hat{\beta}_4 + z_5' p_5 \le \hat{\beta}_4$. The prior mean of the coefficient for *COLL* is $\hat{\beta}_4 - 0.1$.

Table 9. Parameter Support for RGME1 (sign and other restrictions)

Parameter	Parameter Support	Prior Mean
β_1 (constant)	$z_1' = \{-50 \quad -25 \quad 0 \quad 25 \quad 50\}$	0
β_2 (hhsz)	$z_2' = \{0 \quad 5 \quad 10 \quad 15 \quad 20\}$	10
β_3 (unemp)	$z_3' = \{0 \quad 5 \quad 10 \quad 15 \quad 20\}$	10
β_4 (hs)	$z_4' = \{-20 \quad -15 \quad -10 \quad -5 \quad 0\}$	-10
β_5 (coll)	$z_5' = \{-0.20 \quad -0.15 \quad -0.10 \quad -0.05 \quad 0\}$	$\hat{\beta}_4 - 0.1$
β_6 (medinc)	$z_6' = \{-20 \quad -15 \quad -10 \quad -5 \quad 0\}$	-10
β_7 (d90)	$z_7' = \{-20 \quad -10 \quad 0 \quad 10 \quad 20\}$	0

- RGME2 – Here the parameter supports are exactly as described in Table 7 for RGME2 with sign restrictions only. The only change comes in the specification of the support matrix, which now includes the cross-product term between β_4 and β_5. The prior mean of *COLL* is $\hat{\beta}_4 - 1$.

Table 10. Parameter Support for RGME2 (sign and other restrictions)

Parameter	Parameter Support	Prior Mean
β_1 (constant)	$z_1' = \{-50 \quad -25 \quad 0 \quad 25 \quad 50\}$	0
β_2 (hhsz)	$z_2' = \{0 \quad 2.5 \quad 5 \quad 7.5 \quad 10\}$	5
β_3 (unemp)	$z_3' = \{0 \quad 0.5 \quad 1 \quad 1.5 \quad 2\}$	1
β_4 (hs)	$z_4' = \{-2 \quad -1.5 \quad -1 \quad -0.5 \quad 0\}$	-1
β_5 (coll)	$z_5' = \{-2 \quad -1.5 \quad -1 \quad -0.5 \quad 0\}$	$\hat{\beta}_4 - 1$

β_6 (medinc)	$z_6' = \{-10 \quad -7.5 \quad -5 \quad -2.5 \quad 0\}$	-5
β_7 (d90)	$z_7' = \{-20 \quad -10 \quad 0 \quad 10 \quad 20\}$	0

We again specify error supports using bounds of $\pm 3\sigma$ and $\pm 4\sigma$. The restricted GME estimators are labeled RGME1S3, RGME2S3, RGME1S4, and RGME2S4, where S3 and S4 refer to the use of a 3σ and 4σ rule, respectively. Table 11 summarizes the different GME estimators we use in the paper with references to the tables they are used for. Table 12 gives point estimates for the poverty data using IRLS and our four different RGME estimators.

Table 11. Summary of GME estimators used

Estimator	Parameter Support	Error Bounds	Location
GME1S3	Wide parameter bounds	$[-3\sigma, 3\sigma]$	Tables 2, 5
GME2S3	Narrow parameter bounds	$[-3\sigma, 3\sigma]$	Tables 3, 5
GME3S3	Narrow parameter bounds with non-binding restrictions	$[-3\sigma, 3\sigma]$	Tables 4, 5
GME1S4	Wide parameter bounds	$[-4\sigma, 4\sigma]$	Tables 2, 5
GME2S4	Narrow parameter bounds	$[-4\sigma, 4\sigma]$	Tables 3, 5
GME3S4	Narrow parameter bounds with non-binding restrictions	$[-4\sigma, 4\sigma]$	Tables 4, 5
RGME1S3	Wide parameter bounds with binding restrictions	$[-3\sigma, 3\sigma]$	Tables 6, 8, 9, 12
RGME2S3	Narrow parameter bounds with binding restrictions	$[-3\sigma, 3\sigma]$	Tables 7, 8, 10, 12
RGME1S4	Wide parameter bounds with binding restrictions	$[-4\sigma, 4\sigma]$	Tables 6, 8, 9, 12
RGME2S4	Narrow parameter bounds with binding restrictions	$[-4\sigma, 4\sigma]$	Tables 7, 8, 10, 12

Table 12. IRLS and RGME Estimates for Poverty Data with Sign and Other Restrictions

				S3		S4	
Variable		OLS	IRLS	R1GME1	R1GME2	R1GME1	R1GME2
POV	β_1	21.659	16.509	2.712	8.004	1.247	7.543
HHSZ	β_2	1.804	2.096	4.489	3.419	5.006	3.603
UNEMP	β_3	0.076	0.045	0.115	0.189	0.092	0.215
HS	β_4	-0.201	-0.123	-0.034	-0.085	-0.031	-0.093
COLL	β_5	0.021	-0.123	-0.087	-0.088	-0.102	-0.115
MEDINC	β_6	-0.416	-0.285	-0.262	-0.253	-0.266	-0.234
D90	β_7	8.504	6.742	6.802	6.303	7.340	6.287

With the additional restrictions, RGME and IRLS parameter estimates again take the same signs. IRLS has a corner solution to the restriction with $\hat{\beta}_5 = \hat{\beta}_4$ while the alternative RGME estimators all have $\hat{\beta}_5 < \hat{\beta}_4$ as specified by the restriction. All of our GME programs were written using the GAUSS constrained optimization module. The programs are available on our websites at http://www.bus.lsu.edu/academics/economics/faculty/chill/main.html or http://www.bus.lsu.edu/academics/economics/faculty/rcampbell/main.html.

5. GME Interval Estimates

In this section, we use a bootstrap to obtain interval estimates for the GME estimator. In several sampling experiments, GJM find that the GME estimator has a smaller variance than the OLS estimator. Thus, although the GME estimator is biased, GME has lower empirical risk than OLS due to the small variability of the GME estimator.

The bootstrap is a method for estimating standard errors by resampling the original data. Freedman and Peters (1984a) and Freedman and Peters (1984b) describe the use of the bootstrap in regression models. Horowitz (1997) presents a bootstrap method for computing confidence intervals where t-statistics are obtained from the resampled data and interval estimates are computed as $\hat{\beta} \pm t_c^* se(\hat{\beta})$, where t_c^* is the bootstrap t-statistic and $se(\hat{\beta})$ is the asymptotic standard error of the estimator. Since we do not know the asymptotic distribution of the GME estimator, we use the percentile method described by Mooney and Duval (1993) to obtain confidence intervals and examine the precision of the GME estimator.

We construct our confidence intervals for GME and IRLS by resampling from our original data and estimating the model $T = 400$ times. We then order the resulting estimates and find the values corresponding to the 2.5% or 10th value and the 97.5% or 390th value. For OLS we computed confidence intervals as $b_k \pm t_c se(b_k)$. Table 13 gives interval estimates for the poverty data with no restrictions on the parameter estimates. The interval estimates show a few interesting things about the GME estimator. First, GME interval estimates are generally narrower than the OLS interval estimates, indicating a higher degree of precision. The distribution of the GME estimator is roughly symmetric about the mean as shown in Charts 1 and 2, which give the empirical distributions of $\hat{\beta}_2$ and $\hat{\beta}_4$ for the GME1S4 estimator. GME estimates for the other parameters follow similar distributions.

Table 13. OLS and GME Interval Estimates for Poverty Data (*N* =116 Observations)

Variable		OLS	S3 GME1	S3 GME2	S3 GME3
POV	β_1	[10.698, 32.619]	[9.054, 22.147]	[12.357, 23.057]	[8.157, 19.711]
HHSZ	β_2	[-0.498, 4.107]	[1.248, 3.929]	[1.051, 3.123]	[1.955, 4.189]
UNEMP	β_3	[-0.041, 0.193]	[0, 0.275]	[0.006, 0.274]	[0.012, 0.272]
HS	β_4	[-0.279, -0.124]	[-0.218, -0.113]	[-0.221, -0.124]	[-0.211, -0.113]
COLL	β_5	[-0.069, 0.112]	[-0.047, 0.182]	[-0.055, 0.165]	[-0.044, 0.178]
MEDINC	β_6	[-0.508, -0.323]	[-0.515, -0.301]	[-0.503, -0.298]	[-0.509, -0.299]
D90	β_7	[6.437, 10.572]	[5.666, 10.412]	[5.573, 9.965]	[6.004, 10.447]
Variable			S4 GME1	S4 GME2	S4 GME3
POV	β_1		[9.591, 19.636]	[13.367, 21.147]	[8.385, 17.051]
HHSZ	β_2		[1.567, 3.684]	[1.242, 2.759]	[2.385, 4.061]
UNEMP	β_3		[0.014, 0.269]	[0.027, 0.262]	[0.037, 0.272]
HS	β_4		[-0.200, -0.107]	[-0.204, -0.122]	[-0.196, -0.109]
COLL	β_5		[-0.050, 0.158]	[-0.059, 0.139]	[-0.053, 0.144]
MEDINC	β_6		[-0.495, -0.290]	[-0.481, -0.286]	[-0.485, -0.285]
D90	β_7		[5.918, 10.067]	[5.632, 9.539]	[6.100, 10.054]

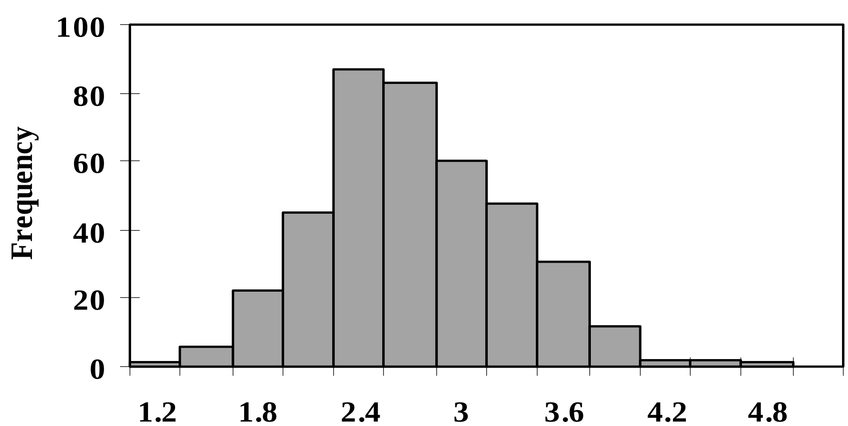

Chart 1. Empirical Distribution of GME1S4 Estimator for β_2 (T = 400 Observations)

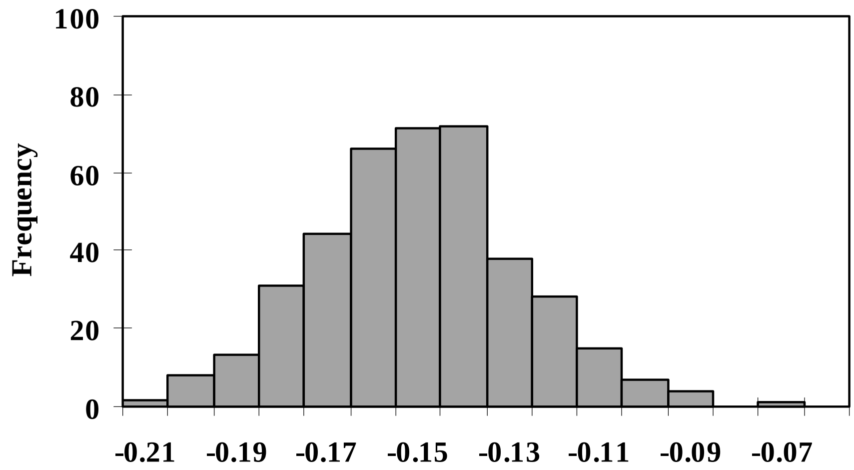

Chart 2. Empirical Distribution of GME1S4 Estimator for β_4 (T = 400 Observations)

Comparing GME1, GME2, and GME3 for a given set of error bounds we observe little difference in the width of the confidence interval. The intervals shift due to differences in the prior mean and the width of the parameter support, but the width of the interval remains fairly constant. However, the intervals become narrower as we increase the error bounds from $\pm 3\sigma$ to $\pm 4\sigma$. Increasing the error bounds shrinks the estimates towards their prior mean and reduces the variability of the parameter estimates. For our problem, it appears that using bounds of $\pm 4\sigma$ leads to better estimates than bounds of $\pm 3\sigma$. However, this may vary from problem to problem. Also, this does not imply that using even wider bounds such as $\pm 5\sigma$ would provide even better estimates. As we make the error bounds infinitely wide the variability goes to zero and the parameter estimates are equal to the prior mean. Table 14 gives interval estimates for the poverty data with sign restrictions placed on the parameters.

Table 14. OLS and GME Interval Estimates for Poverty Data with Sign Restrictions

Variable		IRLS	S3 RGME1	S3 RGME2
POV	β_1	[16.277, 30.641]	[6.953, 16.890]	[12.415, 19.131]
HHSZ	β_2	[0, 3.281]	[2.855, 4.905]	[2.369, 3.500]
UNEMP	β_3	[0, 0.180]	[0.005, 0.229]	[0.101, 0.279]
HS	β_4	[-0.264, -0.163]	[-0.201, -0.112]	[-0.235, -0.155]
COLL	β_5	[-0.063, 0]	[-0.022, 0]	[-0.096, -0.005]
MEDINC	β_6	[-0.472, -0.317]	[-0.448, -0.285]	[-0.403, -0.247]
D90	β_7	[6.119, 9.724]	[6.247, 10.259]	[5.479, 9.110]
Variable			S4 RGME1	S4 RGME2
POV	β_1		[3.985, 13.111]	[11.952, 17.696]
HHSZ	β_2		[3.755, 5.461]	[2.807, 3.674]
UNEMP	β_3		[0.022, 0.222]	[0.143, 0.304]
HS	β_4		[-0.184, -0.095]	[-0.242, -0.161]
COLL	β_5		[-0.028, 0]	[-0.126, -0.030]
MEDINC	β_6		[-0.437, -0.295]	[-0.374, -0.215]
D90	β_7		[6.819, 10.203]	[5.428, 8.697]

The results are consistent with the results for the unrestricted estimates. The restrictions cause all of the confidence intervals to become narrower and the distributions for some of the parameter estimates to be truncated. We again observe smaller intervals as we increase the error bounds. In the restricted case we also observe a large shift in the interval when we increase the error bounds since the parameter estimates are moving towards their prior mean, which is not equal to zero for most of the parameters.

6. Conclusions

This paper applies maximum entropy estimation in an economic model of poverty rates. We discuss how to specify the parameter and error support matrices for the GME estimator. In our model, the GME estimates take the same signs and are roughly the same magnitude as OLS estimates. We find that varying the width of the parameter support does not affect the GME estimates very much. Therefore, a researcher with little prior information could specify a wide parameter support that is symmetric about zero and obtain estimates that are reasonably close to

OLS estimates. Varying the width of the error bounds has a larger impact on the estimates. For the poverty rate example, we conclude that error bounds of $\pm 4\hat{\sigma}$ are preferred over error bounds of $\pm 3\hat{\sigma}$, where $\hat{\sigma}$ is the sample standard deviation of y.

We use a bootstrap to develop confidence intervals for GME. We find that the GME estimator has a narrower confidence interval than the OLS estimator does, which is consistent with GJM. The confidence intervals for the GME estimator become smaller, indicating greater precision in the estimates, as we increase the error bounds. However, the wider the error bounds are set the more important it is that we obtain good nonsample information for specifying the parameter support. The GME estimator is a shrinkage estimator where the parameter estimates are shrunk towards the prior mean, which is based on nonsample information. As we increase the degree of shrinkage towards the prior mean we need to make sure that the prior mean is based on good nonsample information.

Finally, we develop a more general parameter support matrix that allows us to impose a broader set of parameter restrictions than are possible under the traditional support matrix for GME estimation. We demonstrate how to impose restrictions on the GME estimator using a simple example of sign restrictions and another more complicated example using our new parameter support matrix. In both cases, the restrictions are relatively simple to impose and the restricted GME estimator works very well in our model. One important feature of RGME estimation is that it does not restrict us to corner solutions as the IRLS estimator does. For example, when we impose that restriction that $\beta_2 < 0$ the IRLS estimate will be equal to OLS if the OLS estimate is negative and to zero otherwise. We find that the RGME estimate is often negative even when OLS and GME estimates are positive. Because GME estimation relies on prior information, our new support matrix is an important contribution since it allows us a simple way to impose prior information that is often encountered in empirical research.

References

Fraser, I., 2000. An application of maximum entropy estimation: the demand for meat in the United Kingdom. Applied Economics 32, 45-59.

Freedman, D.A, Peters, S.C., 1984a. Bootstrapping a regression equation: some empirical results. Journal of the American Statistical Association, 79, 97-106.

Freedman, D.A., Peters, S.C., 1984b. Bootstrapping an econometric model: some empirical results. Journal of Business and Economic Statistics, 2, 150-158.

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

Golan, A., Judge, G., Perloff, J.M., 1997. Recovering information from censored and ordered multinomial response data. Journal of Econometrics, 79, 23-51.

Golan, A., Perloff, J.M., Shen, E.Z., 2001. Estimating a demand system with nonnegativity constraints: Mexican meat demand. The Review of Economics and Statistics, 83, 541-550.

Griffiths, W.E., Hill, R.C., Judge, G.G., 1993. Learning and practicing econometrics (John Wiley and Sons, New York).

Horowitz, J.L., 1997. Bootstrap methods in econometrics: theory and numerical performance. In: Krpes, D.M., Wallis, K.F. (Eds.), Advances in economics and econometrics: theory and applications, Seventh World Congress, Volume III. University Press, Cambridge, pp. 188-222.

Jaynes, E.T., 1957a. Information theory and statistical mechanics. Physical Review 106, 620-630.

Jaynes, E.T., 1957b. Information theory and statistical mechanics II. Physical Review 108, 171-190.

Judge, G.G., Hill, R.C., Bock, M.E., 1990. An adaptive empirical bayes estimator of the multivariate
normal mean under quadratic loss. Journal of Econometrics 44, 189-213.

Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H., Lee, T., 1988. Introduction to the theory and practice of econometrics 2nd ed. (John Wiley and Sons, New York).

Miller, D.J., Plantinga, A.J., 1999. Modeling land use decisions with aggregate data. American Journal of Agricultural Economics, 81, 180-194.

Mooney, C.Z., Duval, R.D., 1993. Bootstrapping: a nonparametric approach to statistical inference. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-095. Newbury Park, CA: Sage.

Preckel, P.V., 2001. Least squares and entropy: a penalty function perspective. American Journal of Agricultural Economics, 83, 366-377.

Pukelsheim. F., 1994. The three sigma rule. American Statistician 48, 88-91.

Shen, E.Z., Perloff, J.M., 2001. Maximum entropy and Bayesian approaches to the ratio problem. Journal of Econometrics, 104, 289-313.

Ramanathan, R., 2002. Introductory econometrics with applications (Harcourt College Publishers,

Fort Worth).