1. Data

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

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

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:

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

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

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.

2. Programs and Results

We now estimate the model using GME. Because we must specify support matrices for the unknown parameters and errors, there is not a unique GME estimator. We specify different parameter and error supports to examine the sensitivity of the GME estimates to the specification of priors. For the given programs we consider error bounds of $\pm 3\sigma$. 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. Using the sample standard deviation of y, the 3σ - rule results in an error support of $\{-10 \quad -5 \quad 0 \quad 5 \quad 10\}$.

2.1 GME2S3 – No Restrictions

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 specify slightly wider 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 have a prior mean of zero. Table 2 gives the parameter support for GME2S3.

Table 2. Parameter Support for GME2S3 (no restrictions)

Parameter	Parameter Support	Prior Mean
β_1 (constant)	$z_1' = \{-50 \quad -25 \quad 0 \quad 25 \quad 50\}$	0
β_2 (hhsz)	$z_2' = \{-10 \quad -5 \quad 0 \quad 5 \quad 10\}$	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

Table 3 gives point estimates for the poverty data using OLS and six different GME estimators. The program GME2S3 yields the results in column 3. GME1 and GME3 refer to different parameter supports while estimates under S4 were obtained using error bounds of $\pm 4\sigma$.

Table 3. OLS and GME Estimates for Poverty Data (*N* =116 Observations)

			S3			S4		
Variable		OLS	GME1	GME2	GME3	GME1	GME2	GME3
Constant	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

2.2 R1GME2S3 – Parameter sign restrictions only

In the R1GME2S3 program we constraint each coefficient to take its expected sign. We constrain the coefficients for *HHSZ* and *UNEMP* to be positive and the coefficients for *HS, COLL,* and *MEDINC* to be negative. From the unrestricted estimates in GME2S3 we observe that the only 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. Table 4 gives the parameter support for R1GME2S3. Note that the prior means or the values that the parameters are shrunk towards is no longer zero for each parameter.

Table 4. Parameter Support for R1GME2S3 (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 R1GME1S3, R1GME2S3, R1GME1S4, and R1GME2S4, where S3 and S4 refer to the use of a 3σ and 4σ rule, respectively. Table 5 gives point estimates for the poverty data using IRLS and our four different RGME estimators.

Table 5. IRLS and RGME Estimates for Poverty Data with Sign Restrictions

				S3		S4	
Variable		OLS	IRLS	R1GME1	R1GME2	R1GME1	R1GME2
Constant	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

2.3 R2GME2S3 – Parameter sign and other restrictions

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 $\beta_5 < \beta_4 < 0$.

The R2GME2S3 program illustrates 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. Table 6 gives the parameter supports for R2GME2S3.

Table 6. 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 R2GME1S3, R2GME2S3, R2GME1S4, and R2GME2S4, where S3 and S4 refer to the use of a 3σ and 4σ rule, respectively. Table 7 gives point estimates for the poverty data using IRLS and our four different RGME estimators.

Table 7. IRLS and RGME Estimates for Poverty Data with Sign and Other Restrictions

				S3		S4	
Variable		OLS	IRLS	R2GME1	R2GME2	R2GME1	R2GME2
Constant	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

3. GME Interval Estimates

In this section, we use a bootstrap to obtain interval estimates for the GME estimator. In several sampling experiments, Golan, Judge, and Miller (1996) 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 8 gives interval estimates for the poverty data with no restrictions on the parameter estimates. The GME interval estimates are generally narrower than the OLS interval estimates, indicating a higher degree of precision. Table 9 gives interval estimates for the poverty data with sign restrictions placed on the parameters.

Table 8. OLS and GME Interval Estimates for Poverty Data (N =116 Observations)

Variable		OLS	S3 GME1	S3 GME2	S3 GME3
Constant	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
Constant	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]

Table 9. OLS and GME Interval Estimates for Poverty Data with Sign Restrictions

Variable		IRLS	S3 RGME1	S3 RGME2
Constant	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
Constant	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]

References

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).

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.

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.

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