A Monte Carlo study of the effect of design characteristics on the inequality restricted maximum entropy estimator

Randall C. Campbell[*], R. Carter Hill

Department of Finance and Economics, Mississippi State University, Mississippi State, MS 39762, USA

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

Abstract

In this paper, we conduct a set of Monte Carlo sampling experiments to examine the effect of design characteristics on the inequality restricted maximum entropy (RME) estimator. We generate data under varying design characteristics, and estimate the parameters using maximum entropy and least squares estimation, both with and without parameter inequality restrictions. As part of the experimental design we vary the sample size, the distribution of the regressors, the distribution of the errors, the degree of collinearity, the signal-to-noise ratio, and the specification error. We compare the alternative estimators on the basis of mean square error.

JEL Classification: C13; C14; C49

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

*Corresponding author. Tel.: 1-662-325-1516; Fax: 1-662-325-1977.
 E-mail address: rcampbell@cobilan.msstate.edu
Mailing address: Box 9580, 327 McCool Hall, Mississippi State, MS 39762

1. Introduction

This paper examines the effects of design characteristics on the inequality restricted maximum entropy estimator. We conduct a set of Monte Carlo experiments and examine the effects of design characteristics on the empirical risk of least squares and maximum entropy estimators, both with and without linear inequality restrictions. Golan, Judge, and Miller (1996) [hereinafter GJM] discuss the generalized maximum entropy (GME) estimator, apply it to a linear regression model, and carry out a set of experiments comparing the GME estimator to least squares, restricted least squares, and ridge regression. GJM vary the degree of collinearity in their sampling experiment and find that the GME estimator has lower risk than the alternative estimators, especially when there is a high degree of collinearity.

We add to the GJM contribution by studying a broader set of design characteristics in our experiments. On the basis of experimental results, GJM find that GME estimation should be considered when data are highly collinear. By examining additional data characteristics, we can gain some insight into other cases where researchers may benefit from GME estimation. We examine the effects of changes in the sample size, the distribution of the regressors, the distribution of the errors, the condition number, the signal-to-noise ratio, and the specification error on the risk of the alternative estimators. In addition, we estimate the model using a restricted maximum entropy (RME) estimator. Campbell and Hill (2004) discuss how to impose binding inequality restrictions on the GME estimator through the parameter support matrix.

Section 2 discusses GME estimation in the linear regression model. We discuss parameter inequality restrictions and the RME estimator in this section. Section 3 presents the Monte Carlo experimental design and variations in the experimental design. Section 4 gives results from our experiments and examines the impact that each of the design characteristics has on the performance of the alternative estimators. Section 5 gives results from response surface regressions on the MSE of the alternative estimators while Section 6 concludes the paper.

2. Generalized maximum entropy estimation in the general linear model

In our experiments, we estimate a linear regression model of the form

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

where y is a $T \times 1$ vector of sample observations on the dependent variable, X is a $T \times K$ matrix of explanatory variables, e is a $T \times 1$ 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.[1] Therefore, we reparameterize the linear model such that the unknown parameters and errors take the form of probabilities. 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. 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 would specify boundaries of $z_{k1} = 0$ and $z_{k2} = 1$ when estimating the marginal propensity to consume. 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. GJM (1996, p. 138) discuss the width of the parameter support and conclude that the effects of specifying wide support bounds are small when the prior mean is unchanged. We extend this research by examining the effects of design characteristics under binding inequality restrictions, which change the prior mean of the unknown parameters.[2]

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 kth 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

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

---

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

[2] The prior mean of a parameter is equal to the support points times their prior probabilities or $z_k' \tilde{p}_k$, where $\tilde{p}_k$ are the prior probabilities and are uniform in GME.

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 $w_i \in [0,1]$ 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 may be difficult in practice. Chebychev's inequality states that

$\Pr\left[| X - \mu | < c\sigma\right] \geq c^{-2}$, where X is a random variable with mean μ and variance σ^2 and c is a constant.

Following Pukelsheim (1994), GJM suggest setting the error bounds as $v_{i1} = -3\sigma$ and $v_{i2} = 3\sigma$. We obtain

GME estimates using both the 3σ - rule (GME3) and a more conservative 4σ - rule (GME4). The 3σ - rule

guarantees that at least 88.8% of the unknown errors fall within the error bounds while the 4σ - rule guarantees

that at least 93.75% of the unknown errors fall within the error bounds. In practice, when a researcher does not

know σ, the sample standard deviation of y, σ_y, can be used. Since σ_y will generally be larger than the true

σ this will result in slightly wider error bounds. Thus, an empirical 3σ - rule corresponds more closely with

the 4σ - rule based the true value of σ, and may result in even wider error bounds.

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 ith 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

$$e = Vw = \begin{bmatrix} v'_1 & 0 & \cdots & 0 \\ 0 & v'_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v'_T \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_T \end{bmatrix}, \tag{5}$$

where e is a $T \times 1$ vector of random errors, V is a $T \times TJ$ matrix of support points, and w is a $TJ \times 1$ vector of

unknown weights such that $w_{ij} > 0$ and $w'_i i_J = 1$ for all i.

The reparameterized model in matrix form is written 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. GJM refer to the maximum entropy estimator for the reparameterized model as the generalized maximum entropy (GME) estimator.[3] The GME parameter and error estimates are given by

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

and

$$\hat{e}_{GME} = Vw,\tag{8}$$

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

Shannon (1948) 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

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

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

$$(I_T \otimes i'_J)w = i_T,\tag{12}$$

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 T errors.

The solutions to the GME constrained optimization problem are

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

and

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

---

[3] Our GME estimator corresponds to the GME-D estimator given by GJM (p. 86)

where x_k is the $T \times 1$ vector of observations for the kth explanatory variable and λ is a $T \times 1$ vector of Lagrange multipliers for the data constraint. 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, and can be written as

$$\hat{\beta}_{GME} = (X'X)^{-1}X'y - (X'X)^{-1}X'\hat{e}_{GME} = (X'X)^{-1}X'(y - e_{GME}) . \tag{15}$$

Thus, GME minimizes the SSE for a fitted regression line that passes through the mean of $y - \hat{e}_{GME}$ rather than through the mean of y. As $\hat{e}_{GME} \rightarrow 0$ (narrower error bounds), the GME estimator converges to the OLS estimator. As the error bounds are made wider the GME estimator is shrunk towards 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 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 $\hat{\beta}_k = z'_k \hat{p}_k$).

2.1 Restricted maximum entropy estimation in the general linear model

An economic researcher often has sign or other information about the parameters that can be expressed as a linear inequality restriction. We impose parameter sign restrictions on both the least squares and maximum entropy estimators. Judge et al. (1988, pp. 822-825) show that the inequality restricted least squares (IRLS) estimator is biased, but dominates the OLS estimator, under a squared error loss measure, as long as the restrictions are true or nearly true.

Using the parameter support matrix we impose 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. Instead, the parameters are bounded because they must be and the bounds are not based on prior

---

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

information. We impose parameter sign restrictions following Campbell and Hill (2004) and examine how the restricted maximum entropy (RME) estimator performs as we vary the degree of collinearity, the sample size, the distribution of the errors and regressors, the signal-to-noise ratio, and the specification error.

If we have nonsample information that $\beta_k > 0$ we specify the support vector for β_k to take only positive values such as $z'_k = \begin{bmatrix} 0 & 5 & 10 & 15 \end{bmatrix}$, where z_k is the $M \times 1$ parameter support vector for β_k. In this case the RME estimate

$$\hat{\beta}_k = 0\hat{p}_{k1} + 5p_{k2} + 10\hat{p}_{k3} + 15p_{k4} > 0, \tag{16}$$

since $\hat{p}_{km} \geq 0$ for all M support points. Note that the parameter estimate will be shrunk toward the prior mean which is non-zero in this case. Likewise, for a negative sign restriction, $\beta_k < 0$, we specify the parameter support vector to take only negative values. For these parameter sign restrictions, the parameter support matrix is block diagonal. From equation (3), each unknown parameter, β_k, is associated with a support vector, z'_k, and the off-diagonal elements in the support matrix are equal to zero. In this case, the solutions to the RME optimization problem are given by equations (13) and (14).[5] Applications of restricted maximum entropy estimation include Fraser (2000) and Shen and Perloff (2001). Section 3 describes our Monte Carlo experiments and gives the elements of the experimental design.

3. Monte Carlo experimental design

GJM (pp. 133-137) carry out a Monte Carlo experiment comparing the empirical risk of OLS, IRLS, ridge regression, and GME estimators under varying degrees of collinearity. They find that the GME estimator has lower mean squared error (MSE) than the alternative estimators at all levels of collinearity, but the GME estimator performs especially well compared to the alternative estimators when the degree of collinearity is high. This is consistent with other shrinkage estimators. We carry out essentially the same Monte Carlo experiments, but in addition to the degree of collinearity we vary the sample size, the distribution of the errors, the distribution of the regressors, the signal-to-noise ratio, and the specification error. We examine the effects of the

---

[5] Campbell and Hill (2004) also discuss estimation under restrictions such as $\beta_k > \beta_j$, where the parameter support matrix is not block diagonal.

design characteristics on the RME estimator as well as the OLS, IRLS, and GME estimators. We estimate the

linear model

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + e_t, \qquad t = 1, \ldots, T \tag{17}$$

The model is written in matrix form as in equation (1):

$$y = X\beta + e,$$

where X is a $T \times 4$ matrix of random regressors, $\beta = \begin{bmatrix} 2, & 1, & -3, & 2 \end{bmatrix}'$ is a known parameter vector, and e is a

$T \times 1$ vector of random errors.

To vary the degree of collinearity, we obtain the singular value decomposition of $X = QLR$, where Q is a

$T \times K$ matrix, L is a $K \times K$ diagonal matrix with eigenvalues $l_i \geq 0$, and R is a $K \times K$ matrix such

that $R'R = I_K$. We replace the eigenvalues in L with the vector $a = \begin{bmatrix} \sqrt{\dfrac{2}{1+\mu}}, & 1, & 1, & \sqrt{\dfrac{2\mu}{1+\mu}} \end{bmatrix}'$, where μ is

a known constant, to form a new matrix L_a. We use L_a to create a new matrix of regressors $X_a = QL_aR$.

Belsley, Kuh, and Welsch (1980) define the condition number of $X'X$ to be $(\gamma_1 / \gamma_K)^{1/2}$, where γ_1 and γ_k are

the largest and smallest eigenvalues, respectively. The largest and smallest eigenvalues for $X_a'X_a$ are

$$\gamma_1 = \frac{2\mu}{1+\mu} \quad \text{and} \quad \gamma_K = \frac{2}{1+\mu}. \quad \text{Thus, the condition number } \kappa(X_a'X_a) = \mu^{1/2} \text{ is specified } a\ priori.$$

During each of $N = 1{,}000$ Monte Carlo iterations we generate the dependent variable $y = X_a\beta + e$. In

addition, we generate a hold-out sample as $y_0 = X_{a0}\beta + e_0$, which we use to examine out-of-sample prediction.

We vary the estimation sample size using the values $T = 10, 50, 100, 300,$ and 500. We use a hold-out sample of

size $T_0 = 50$ for all experiments.

Following GJM, we vary the degree of collinearity. Belsley, Kuh, and Welsch (1980) conclude that

condition numbers greater than about 30 indicate that collinearity may be a problem. We consider condition

numbers $\kappa(X_a'X_a) = 1, 10, 20, 40, 60, 80, 100, 250,$ and 500. Additionally, we vary the distribution of both

the errors and the regressors. In our experiments we draw both the errors and regressors from a standard normal

distribution as well as from $t_{(3)}$ and a $\chi^2_{(5)}$ distributions, correcting the mean to 0 and the variance to 1. We

chose a $t_{(3)}$ distribution since it is symmetric with a mean of 0 and a variance of 1, but has thicker tails than the standard normal distribution. We chose a $\chi^2_{(5)}$ distribution in order to examine a non-symmetric distribution. Since each column of X has the same expected length, we do not scale the data to unit length as suggested by Belsley et al.

We next consider the signal-to-noise (s/n) ratio. Since we do not include an intercept term the signal-to-noise ratio is equal to

$$s/n = \beta'(\sigma^2 (X'X)^{-1})^{-1} \beta = \frac{\beta'X'X\beta}{\sigma^2} \quad . \tag{18}$$

We vary the signal-to-noise ratio by changing the error variance σ^2. However, because X is randomly drawn we cannot specify the exact signal-to-noise ratio *a priori.* Therefore, we will discuss the effects of the signal-to-noise ratio in terms of σ^2. Holding everything else constant, an increase in σ^2 represents a decrease in the signal-to-noise ratio.

Finally, we examine the effects of specification error. For a linear inequality restriction, $\beta_i \geq r_i$, define the specification error $\delta_i = r_i - \beta_i$. Thus, the specification error will be positive if the restriction is not true. Judge et al. (1988) show that the IRLS estimator dominates the OLS estimator under a squared error loss measure as long as the restrictions are nearly true. In all of our experiments, we compare the OLS and GME estimators to the IRLS and RME estimators with the parameter estimates restricted to take the correct signs. To examine the effects of specification error, we consider the parameter $\beta_2 = 1$ and impose the restrictions $\beta_2 \geq 0, 0.5, 0.8, 0.9,$ 1.0, 1.1, and 1.2, which represent increasing specification error. Table 1 summarizes the dimensions of our experimental design.

4. Results

We present the Monte Carlo results using a separate table for each design characteristic. We start with a "base case" experiment using $T = 100$ in-sample observations, standard normal errors, standard normal regressors, $\kappa(X_a'X_a) = 10$, $\sigma^2 = 1$, and only the parameter sign restrictions for our restricted estimators. From this basic design, we vary each design characteristic holding all other design characteristics constant.

4.1 Sample size

In this section, we examine the effects of the estimation sample size on the alternative estimators. We consider sample sizes ranging from 10 observations to 500 observations. As we increase the sample size we cannot hold the signal-to-noise ratio constant since it depends on the random X matrix whose dimensions vary with the sample size. The signal-to-noise ratio ranges from about 19 to 29 as we draw data using different sample sizes. In the OLS model, the R^2 increases as the signal-to-noise ratio increases. Table 2 gives the MSE of the alternative estimators as the sample size varies. The alternative estimators include OLS, IRLS, GME3 (which uses the 3σ - rule for error bounds), GME4 (which uses the 4σ - rule for error bounds), and two RME estimators. Note that we use the true value of σ, which is known in our experiments, rather than the sample value σ_y. We use the 3σ - rule for both RME estimators, but we change the prior mean. The RME3-I estimator has an "incorrect" prior mean vector of $\begin{bmatrix} 5 & 5 & -5 & 5 \end{bmatrix}'$ while for the RME3-C estimator we specify a "correct" prior mean vector of $\begin{bmatrix} 2 & 1 & -3 & 2 \end{bmatrix}'$, which is the true parameter vector.

The results in Table 2 show that the MSE is roughly constant for OLS as we vary sample size. Judge et al. (1988, p. 866) show that the variance of the OLS estimators are given by

$$\text{var}(b_j) = \sigma^2 \sum_{i=1}^{K} p_{ji}^2 / \gamma_i \tag{19}$$

where p_i is the ith characteristic vector and γ_i is the ith characteristic root of $X'X$. Since we hold constant the error variance and the characteristic roots, the variance of b_j changes only with $\sum_{i=1}^{K} p_{ji}^2$, which does not depend on sample size.

As expected, imposing correct inequality restrictions, in this case the correct signs of the parameters, results in lower risk. Thus, the IRLS estimator has lower MSE than the OLS estimator for all sample sizes. GJM (p. 135, Table 8.6.1) find that the IRLS estimator has just slightly lower (and in some cases higher) empirical risk than the OLS estimator. However, GJM restrict the estimates to the range [-10, 10] whereas we restrict the parameter estimates to take the correct sign.[6]   Judge et al. (1988, pp. 822-824) show that the IRLS estimator

---

[6] Empirically, we find that very few of the OLS parameter estimates lie outside the [-10, 10] range. Thus, the GJM parameter restrictions have very little impact and the risk measures for IRLS and OLS are similar.

dominates the OLS estimator under a squared error loss measure as long as the restrictions are true or nearly true. Our results are consistent with this result.

We find that the GME and RME estimators have lower risk than the OLS estimator in most cases. We obtain three results of note here: 1) the GME and RME estimators have higher risk as the sample size increases, 2) the GME4 estimator has lower risk than the GME3 estimator for all sample sizes and particularly in large samples, and 3) the risk of the RME estimators depends on the prior mean of the estimator. Note that throughout the paper we compare the estimators on the basis on MSE. It is possible that the relative rankings of the alternative estimators could change under a different performance measure.

Mittelhammer, Judge, and Miller (2000) discuss large sample properties of the GME estimator. They show that under certain regularity conditions derived by Mittelhammer and Cardell (1997) the GME-NM estimator (as defined by GJM, p.88), which shrinks the error support boundaries as the sample gets larger, is consistent and asymptotically normal. Among the Mittelhammer and Cardell regularity conditions are that the true error values are contained within the error bounds and the true parameters are contained within the parameter bounds. Our GME and RME estimators are inconsistent since the processes used to generate errors are unbounded. Therefore, some of the generated errors fall outside the bounds we specified. This is particularly true for the t- and chi-square distributions, which have thicker tails. In addition, we use the GME-D estimator, which does not shrink the error support boundaries as the sample size increases.

GJM (p. 88) suggest using the 3σ - rule (Pukelsheim 1994) for setting the error bounds since at most one-ninth of the unknown errors falls outside this range. We also consider a 4σ - rule, which has at most one-sixteenth of the errors fall outside the range. We expect that wider error bounds will lead to greater shrinkage and lower variance of the GME estimator. However, the reduction in variance may be offset by larger bias. In general, our experimental results show that the GME4 estimator has lower MSE than the GME3 estimator.[7] While the variance increases with sample size for both estimators, it does not increase as quickly for the GME4 estimator as it does for the GME3 estimator.

---

[7] This does not imply that a GME5 estimator would have still lower MSE. We did observe lower variance and higher bias for the GME4 estimator compared to the GME3 estimator. The increase in bias would presumably outweigh the reduction in variance at some point.

Finally, we note the increase in MSE for our inequality restricted (RME) estimators. Inequality restrictions have a greater effect on the GME estimator than simply restricting the estimators. For example, our GME parameter support for β_1 is $\begin{bmatrix} -10 & -5 & 0 & 5 & 10 \end{bmatrix}'$, which has a prior mean of 0. When we restrict the parameter estimates to be positive we obviously must specify a positive prior mean. For the RME3-I estimator we specify the parameter support for β_1 as $\begin{bmatrix} 0 & 2.5 & 5 & 7.5 & 10 \end{bmatrix}'$, which has a prior mean of 5. For the RME3-C estimator we specify the parameter support for β_1 as $\begin{bmatrix} 0 & 1 & 2 & 3 & 4 \end{bmatrix}'$, which has a prior mean of 2. When our prior mean is incorrect (RME3-I), imposing binding inequality restrictions on the GME estimator leads to an increase in MSE for small sample sizes. However, this is not true if the researcher has prior information about both the sign and magnitude of a parameter. The RME3-C estimator shrinks $\hat{\beta}_1$ toward 2 (the true parameter value) and results in both lower variance and bias. These are just two of many possible RME formulations that restrict the parameter estimates to take the correct sign.

Tables 3 and 4 give results as we vary sample size with the errors and regressors drawn from t- and chi-square distributions respectively. GJM (p. 142) examine the effects of non-normal errors on risk measures. They find that GME still has lower risk than the OLS estimator when errors are drawn from a t- or chi-square distribution. When errors are drawn from a t- distribution GJM specify wider error bounds to account for the thicker tails. Thus, we expect to see a larger difference between the GME3 and GME4 estimators when the errors follow a t- distribution. When the errors are drawn from a chi-square distribution GJM specify a skewed error support to account for the fact that the chi-square distribution is skewed. We continue to specify the symmetric error supports $\begin{bmatrix} -3\sigma & 0 & 3\sigma \end{bmatrix}'$ and $\begin{bmatrix} -4\sigma & 0 & 4\sigma \end{bmatrix}'$ for our GME3 and GME4 estimators respectively since Chebychev's inequality holds for any distribution and we have adjusted the means to zero for the random errors.

With errors and regressors drawn from a t- distribution the OLS estimator has basically the same risk as in the case of normal errors and regressors. The IRLS estimator has higher MSE than in the normal case, but still has lower MSE than the OLS estimator. However, the GME3, GME4, and RME3-I estimators have much higher MSE than they do when the errors are normal. This is in contrast with the results obtained by GJM who find only a slight increase in the risk of the GME estimator when the errors are non-normal. However, GJM use

a sample size of 10 observations. With only 10 observations we find that the GME and RME estimators have lower MSE than the OLS and IRLS estimators, which is consistent with GJM. But the MSE for the GME3 and GME4 estimators increases more rapidly than in the case of normal errors. We find that the RME3-C estimator still has the lowest MSE for all sample sizes. With errors from a t-distribution the IRLS estimator has smaller MSE than GME3 or GME4 for any sample size greater than 10 and the OLS estimator has smaller MSE than GME3 or GME4 for sample sizes greater than 50 observations.

The results for the OLS and IRLS estimators are similar in the case of chi-square errors and regressors as in the case of t- errors and regressors. The MSE for OLS is roughly the same as in the normal case while the IRLS has higher MSE than in the normal case. The GME3, GME4, and RME3-I estimators again perform relatively poorly as the sample size increases. In Tables 3 and 4 we change the distribution for both the errors and the regressors. In section 4.2 we examine whether each estimator is affected by the error distribution, the distribution of regressors, or both.

4.2 Distribution of errors and regressors

In this section, we change the distributions of the errors and regressors. We draw the random errors and regressors from a standard normal as well as from a t- distribution with 3 degrees of freedom and a chi-square distribution with 5 degrees of freedom, correcting the mean to 0 and the variance to 1. Thus, we draw from three different distributions for both the errors and regressors, resulting in nine possible combinations. Table 5 gives results for all nine combinations.

The results in Table 5 show that the MSE for the OLS estimator does not change much as we change the distribution of the errors and regressors (although there appears to be a slight increase in MSE when the errors are chi-square). The IRLS estimator has lower MSE than the OLS estimator for all combinations, which is expected since we are imposing restrictions that are true. However, the gains are much smaller when the regressors are drawn from a t- or chi-square distribution. Holding the distribution of regressors constant, the MSE for the IRLS estimator changes very little in response to changes in the error distribution. The GME and RME estimators perform best when the errors are normal regardless of the distribution of the regressors. Thus, while the IRLS estimator is affected by the distribution of regressors the GME estimators are affected primarily by the distribution of the error term.

Table 6 gives the MSE of the alternative estimators with normal regressors and varying errors, but with error variance $\sigma^2 = 5$. Table 7 gives the MSE of the alternative estimators with normal errors and varying regressors with condition number $\kappa(X_a' X_a) = 100$. Since the data are relatively "noisy" in these experiments we expect the maximum entropy estimators to perform well relative to least squares. As the variance of the error term increases and the signal-to-noise ratio decreases, the variance and the MSE of the OLS estimator increase. Comparing the results in Table 6 to the first three rows in Table 5, we observe that the MSE of the OLS estimator is exactly 5 times larger in each case, which is the same factor by which we increase the variance of the errors. The MSE of the IRLS and GME estimators also increase, but by less than 5 times. There do not appear to be any significant gains for GME and RME relative to IRLS as the signal-to-noise ratio decreases.

GJM find that the GME estimator has lower MSE than the OLS, IRLS, and ridge regression estimators when the data are collinear. Comparing the results from Table 7 to rows 1, 4, and 7 in Table 5, we observe that as we increase the condition number from 10 to 100, the MSE of the OLS estimator increases by roughly 100 times. The MSE of the IRLS estimator with parameter sign restrictions also increases, but by a much smaller amount. The MSE of the GME3 and GME4 estimators is lower with the higher degree of collinearity. As the collinearity increase, there is less information in the data and the parameter estimates are shrunk toward the prior means resulting in a lower variance for the GME and RME estimators.

For GME we observe a slight increase in bias and a larger decrease in variance as we increase the condition number. The RME3-I estimator has higher MSE for t- and chi-square errors since the increase in bias is not offset by the decrease in variance. Because we specify the RME3-I estimators with a prior mean of 5, the increase in bias is larger than the decrease in variance. The RME3-C estimator has roughly constant MSE as we increase the condition number. All of the GME estimators have lower MSE than the OLS and IRLS estimators when the condition number is equal to 100. Section 4.3 examines the effects of collinearity on the alternative estimators in more detail.

4.3 Degree of collinearity

GJM show that GME has lower squared error loss than OLS and IRLS in their sampling experiments, and that the difference increases with the degree of collinearity. We find the same results here, but we also examine the RME estimator. In addition, we examine the effects of collinearity combined with a higher error variance.

Table 8 gives MSE for the alternative estimators as the degree of collinearity varies with $\sigma^2 = 1$ while Table 9 does the same when $\sigma^2 = 5$. Table 10 examines the same changes in condition numbers, but with a smaller sample size (T = 50).

Table 8 shows that as the degree of collinearity increases, the MSE of the OLS estimator increases at a much faster rate than does the MSE of the IRLS, GME, and RME estimators. This is consistent with the results obtained by GJM. GJM find that the MSE of the IRLS estimator with the parameter estimates restricted to the range [-10, 10] increases at nearly the same rate as OLS. This is not surprising since the OLS parameter estimates do not fall outside this range very often. However, the MSE of the IRLS estimator with parameter sign restrictions does not increase nearly as much as it does for the OLS estimator when the condition number increases. While the MSE of the GME and RME estimators is still lower than the MSE of the IRLS estimator, the difference is not nearly as large as in the experiments by GJM. This can be attributed to the fact that we are using parameter sign restrictions rather than the restriction range that GJM chose. We again find that the GME4 estimator has the lowest MSE and that the RME3-I estimator with incorrect prior mean has higher MSE than the "unrestricted" GME estimators while the RME3-C estimator with correct prior mean has lower MSE than the "unrestricted" GME estimators.

Table 9 shows the MSE for the alternative estimators when the variance of the errors is equal to 5, thus reducing the signal-to-noise ratio. The MSE for the OLS estimator is exactly five times the MSE obtained with an error variance of 1. The MSE for the IRLS estimator is roughly two times larger than it is when the error variance is equal to 1, and the MSE's of the GME3, GME4, and RME3-I estimators are roughly three times larger than when the error variance is equal to 1. Thus, we find that the empirical risk of the IRLS estimator becomes closer to the empirical risk of the GME estimators as we increase the error variance.

Table 10 shows that the MSE of the GME3, GME4, and RME3-C estimators is lower when the sample size is smaller. This is consistent with Tables 2-4. The reason for this is that when the sample size and the signal-to-noise ratio are small the GME estimators place less weight on the data and shrink the parameter estimates toward their prior means. Thus, while the bias of the GME and RME estimators is greater when the data is less informative the variance of the estimators is smaller, resulting in a lower MSE. The RME3-I estimator has higher MSE when we reduce the sample size because the parameter estimates are shrunk toward incorrect values.

4.4 Signal-to-noise ratio

We have already shown that GME and RME have much lower MSE than the OLS and IRLS estimators when the signal-to-noise ratio is low (the data are "noisy"). We show this more clearly here as we examine the effects of increasing the error variance using the values $\sigma^2 = 1, 2, 3, 4,$ and 5. In Tables 11-14 we show this effect for normal errors and regressors, t- errors and regressors, chi-square errors and regressors, and for a condition number of 100, respectively.

In Table 11, the MSE of the OLS estimator increases in the same proportion as the error variance. The MSE for the IRLS, GME, and RME estimators also increases as the error variance increase, but at a much slower rate. As we increase the error variance, the GME estimators are shrunk more just as they are when we have a high degree of collinearity. However, we now observe that the increase in bias is not offset by the reduction in variance. The GME3 and RME3-I estimators perform comparably to IRLS. The GME4 estimator has the lower MSE than IRLS for all values of σ^2. The RME3-C estimator, which shrinks the parameter estimates toward the correct values has much lower MSE than the alternative estimators.

With errors and regressors drawn from a t- distribution, as in Table 12, the MSE of the OLS estimator again increases in the same proportion as the error variance. The GME and RME estimators do not generally perform well relative to IRLS with t- errors and regressors (see Table 3). However, the advantage for IRLS disappears as we increase the error variance. The RME3-C estimator has lower MSE than IRLS for all levels of σ^2 and the RME3-I estimator, which shrinks toward incorrect values, even has lower MSE than IRLS when σ^2 is equal to 5. The RME3-I estimator is heavily biased in this case since the parameter estimates are shrunk toward incorrect values. For the IRLS estimator, both the variance and bias increase with σ^2.

Drawing the errors and regressors from a chi-square distribution, as in Table 13, produces similar results to drawing from a t- distribution. However, the GME and RME estimators perform better relative to IRLS and OLS under a chi-square distribution. The GME4 and RME3-C estimators have lower MSE than the IRLS estimator for all levels of the error variance, and the RME3-I estimator has lower MSE than the IRLS estimator when $\sigma^2 > 2$. We again find that both the bias and variance of the IRLS estimator increase fairly rapidly with σ^2. The bias and variance also increase for the GME and RME estimators, but at a much slower rate than they do for the IRLS estimator.

Finally, comparing Table 14 to Table 11, we see again that as the degree of collinearity increases the MSE of the OLS estimator increases substantially. The MSE of the IRLS estimator increases slightly while the MSE of the GME and RME estimators decreases. With respect to the error variance, we find the greatest gains in the MSE of the GME and RME estimators relative to IRLS when the error variance is low.

4.5 Specification error

In this section, we impose stronger restrictions on β_2 than the parameter sign restrictions. In our experiments the true value of β_2 is 1. Table 15 shows MSE for the alternative parameters as we vary the specification error. Tables 16 and 17 show the same, but with a greater degree of collinearity. Table 18 gives results with $\sigma^2 = 5$, representing a smaller signal-to-noise ratio.

As expected, imposing correct inequality restrictions on the parameter β_2 leads to the restricted estimators (IRLS, RME3-I, and RME3-C) having a lower MSE than the OLS estimator. As shown by Judge et al. (1988) the IRLS estimator has lower MSE than OLS even when the restrictions are nearly true, as when we restrict $\beta_2 \geq 1$. We observe that the MSE of the IRLS begins to increase as the specification error $\delta > 0$ while the MSE of the RME estimators increases even as we impose correct information. This occurs because as we impose stronger restrictions we change the prior mean of the parameter estimates. For example, the RME3-C estimator has a parameter support for β_2 of $\begin{bmatrix} 0 & 0.5 & 1 & 1.5 & 2 \end{bmatrix}'$ with a prior mean of 1 when we restrict $\beta_2 \geq 0$. But when we restrict $\beta_2 \geq 0.5$ we specify a parameter support of $\begin{bmatrix} 0.5 & 0.9 & 1.3 & 1.7 & 2 \end{bmatrix}'$ with a prior mean of 1.28. Thus, the bias increases as we impose stronger inequality restrictions. The restrictions reduce the variance slightly for the RME estimators, but the larger increase in bias results in higher MSE. We also note that the RME estimators violate the Mittelhammer and Cardell (1997) regularity conditions and are inconsistent when the parameter support does not contain the true parameters.

In Tables 16 and 17, we show that increasing the degree of collinearity does not change the effects of specification error on the MSE of the alternative estimators. The MSE of the IRLS estimator decreases until the specification error is equal zero and then it increases. The MSE of the RME estimators increases even as true restrictions are imposed. However, as we showed earlier the GME and RME estimators have lower MSE relative to the IRLS and OLS estimators as we increase the degree of collinearity.

In Table 18, as we increase the error variance, we observe the same effects of specification error on the IRLS and RME estimators. All of our GME programs were written using the GAUSS constrained optimization module. GAUSS code for sample programs demonstrating our Monte Carlo experiments are available at:

http://www.bus.lsu.edu/academics/economics/faculty/chill/personal/irme.htm.

5. Response surfaces for Monte Carlo experiments

In this section, we estimate response surfaces for our Monte Carlo experiments. Hendry (1984) and Davidson and MacKinnon (1993) discuss response surfaces as a means of summarizing the results from a set of Monte Carlo experiments. We estimate the following response surface regression

$$\frac{MSE(\hat{\beta})}{MSE(OLS)} = \alpha_1 + \alpha_2 T_i + \alpha_3 TE_i + \alpha_4 CHE_i + \alpha_5 TR_i + \alpha_6 CHR_i + \alpha_7 CN_i + \alpha_8 \sigma_i^2 + \alpha_9 \delta_i + \alpha_{10}\delta_i^2 + \mu_i \qquad (20)$$

$i = 1,\ldots,N_{MC}$, where T_i is the sample size of the estimation sample, TE_i is equal to one for t- errors and zero otherwise, CHE_i is equal to one for chi-square errors and zero otherwise, TR_i is equal to one for t- regressors and zero otherwise, CHR_i is equal to one for chi-square regressors and zero otherwise, CN_i is the condition number, σ_i^2 is the error variance, δ_i is the constraint specification error, and $MSE(\hat{\beta})$ is the MSE of the estimator of interest. We estimate response surfaces for the IRLS, GME3, GME4, RME3-I and RME3-C estimators using $N_{MC} = 88$ unique observations from our Monte Carlo experiments.

We include a squared term for specification error in the IRLS equation since the MSE of the IRLS estimator is minimized when $\delta_i = 0$ and we do not expect a linear relationship between MSE and δ_i for IRLS. Since we do not impose inequality restrictions on the GME3 and GME4 estimators, we set $\alpha_9 = \alpha_{10} = 0$ for these response surfaces. Table 19 gives results from our response surface regressions, which summarize the results of our experiments. Note that the R^2 value is higher for the unrestricted GME estimators. Since the restrictions themselves explain some of the variation in MSE the amount explained by the design characteristics is lower for the restricted estimators.

While the effect of sample size is not significant for the IRLS estimator, sample size has a positive impact on the MSE of the maximum entropy estimators. The effect of sample size is significant at the 5% level of

significance for the GME3, GME4, RME3-I, and RME3-C estimators. This is consistent with the hypothesis that the GME and RME estimators should perform better relative to OLS when the sample size is smaller and there is less information in the data.

The coefficients for drawing errors from a t- or chi-square distribution is also positive for all of the alternative estimators, but are not significant for the IRLS estimator. The error distribution significantly impacts the MSE of the GME3, GME4, RME3-I, and RME3-C estimators. Since the t- and chi-square distributions have thicker tails, there is a greater chance that the error bounds do not contain the true error values. The effects are larger for the GME3 estimator than for the GME4 estimator since the wider bounds contain more of the unknown errors.

The coefficient for drawing regressors from a t- distribution is negative and insignificant for all of the maximum entropy estimators. The coefficient for t- regressors is positive, but insignificant for the IRLS estimator. The coefficient for drawing regressors from a chi-square distribution is positive and insignificant for the GME4 and RME3-I estimators, and is negative and insignificant for the GME3 and RME3-C estimators. However, the coefficient for chi-square regressors is positive and significant at the 10% level for the IRLS estimator. Our results indicate that the error distribution affects the GME and RME estimators since these estimators require us to place bounds on the unknown errors. However, the distribution of regressors affects the IRLS estimator but not the GME and RME estimators. This makes sense for GME and RME since we are required to place bounds on the unknown parameters, but not on the regressors.

As expected, the coefficient on condition number is negative for all of the alternative estimators. This coefficient is significant at the 5% level for the IRLS, RME3-I, and RME3-C estimators, but not for the GME3 and GME4 estimators. This is somewhat surprising since the GME3 and GME4 estimators have relatively low MSE as the condition number grows large. However, since our dependent variable is $\frac{MSE(\hat{\beta})}{MSE(OLS)}$, the regressions may have negative coefficients due to the tremendous increase in the $MSE(OLS)$ rather than anything in the MSE of our alternative estimators. Looking at Tables 8, 9, and 10, we see that the MSE of the OLS estimator increases very rapidly while the MSE of the other estimators remains roughly constant. The ratio of MSE's quickly approaches zero for all of the alternative estimators.

The coefficient for the error variance is negative for all of the alternative estimators and is significant at the 5% level for the GME3, GME4, RME3-I, and RME3-C estimators. This is consistent with our expectations that the GME and RME estimators should perform well relative to OLS as we increase the signal-to-noise ratio. Finally, we observe that the coefficients for both δ and δ^2 are positive, but not significant for the IRLS, RME3-I, and RME3-C estimators. The insignificance of these variables is likely due to the fact that we did not vary the specification error over a very wide range. Looking at Tables 15-18, we observe that the ratio of the MSE of IRLS, RME3-I, and RME3-C to the MSE of OLS is not changing much.

6. Summary and Conclusions

We carry out a Monte Carlo study to examine the effects of design characteristics on the inequality restricted maximum entropy estimator. This research extends the original study by GJM in two important ways. First, in addition to the effects of collinearity, we examine the effects of the sample size, the distribution and variance of the unknown errors, the distribution of the regressors, the signal-to-noise ratio, and the specification error on the alternative estimators. Examining these additional characteristics is important in examining when a researcher may want to use GME estimation. In addition, we examine the effects of these design characteristics on an inequality restricted maximum entropy (RME) estimator. We summarize our experimental results below.

As expected, we find that the GME and RME estimators perform well in terms of MSE when the sample size is small. When the sample size is small there is less information in the data, the GME and RME estimators are shrunk toward their prior means, and the variance of the GME and RME estimators is small relative to the variance of the OLS estimator. As we increase the sample size, the variance of the GME and RME estimators increases slightly and the bias also increases. Our GME and RME estimators do not satisfy the Mittelhammer and Cardell (1997) regularity conditions since the error bounds do not contain the true error values for all observations. Thus, our GME and RME estimators are inconsistent. As we increase the sample size there are likely more observations that fall outside of these bounds. This is particularly true in the case of t- or chi-square errors when the tails of the error distribution are heavier.

Comparing the MSE of the alternative estimators, we find that the GME4 estimator has lower MSE than the GME3 estimator. The wider bounds increase the degree of shrinkage resulting in a smaller variance and a larger

bias. As we increase the sample size the variance increases at a much slower rate for the GME4 estimator than for the GME3 estimator. In addition, we find that the IRLS estimator with parameter sign restrictions performs nearly as well, and in some cases better, compared to the GME and RME estimators. The RME3-C estimator, which shrinks the parameter estimates toward the true parameter values has the lowest MSE in all cases.

As shown in Table 5, we find that the MSE of the GME and RME estimators is affected by the distribution of the errors. Drawing errors from a t- or chi-square distribution leads to more errors falling outside the bounds since there is more mass in the tails of these distributions than in the normal distribution. We find that the IRLS estimator is affected by the distribution of the regressors rather than the distribution of the errors.

We find that the GME and RME estimators perform better relative to the OLS estimator as the condition number increases. This is consistent with the findings of GJM and with theory since the degree of shrinkage is larger the higher the condition number. The variance of the GME and RME estimators decreases as we increase the condition number, which results in a lower MSE in some cases as the condition number increases. However, in contrast to GJM we find that the MSE of the IRLS estimator is much lower than the MSE of the OLS estimator and nearly as low as the MSE of the GME3, GME4, and RME3-I estimators. GJM use the GME parameter bounds as restrictions for the IRLS estimator while we restrict each variable to take the correct sign.

The GME and RME estimators perform well relative to OLS as the error variance increases. The MSE of the OLS estimator increases by the same factor as the error variance increases. The MSE of the IRLS, GME, and RME estimators also increase with σ^2, but at a slower rate. Again this is consistent with the hypothesis that shrinkage estimators should perform well when there is less information in the data.

We observe that the MSE of the IRLS estimator decreases when we impose restrictions that are true or nearly true. Consistent with theory, we find that the MSE of the IRLS is minimized when the specification error is zero. That is, we impose restrictions that are exactly true. However, we find that for our RME estimators additional restrictions increase the prior means and lead to an increase in bias and MSE.

While imposing parameter sign information greatly improves MSE in the case of the IRLS estimator this is not necessarily the case for the RME estimator. For the RME estimator the prior mean, or the value the parameters are shrunk toward, is very important. We specify two RME estimators, one that shrinks the parameter estimates toward incorrect values (RME3-I) and one that shrinks the parameter estimates toward the

true parameter values (RME3-C). In most cases we find that the GME3 and GME4 estimators with no sign restrictions have lower MSE than our RME3-I estimator. This is in contrast to least squares estimation where imposing correct sign information always reduces MSE. However, the RME3-C estimator always has lower MSE than the GME estimators. We conclude that a researcher using GME estimation should impose parameter sign information only if they also have prior information on the magnitude of the unknown parameters.

Our Monte Carlo experiments show that the GME estimator has lower MSE than the OLS and IRLS estimators when the information in the data is limited, such as for a small sample size, high condition number, or large error variance. In these cases the GME estimator is slightly biased, but has much lower variance than the OLS estimator. We also examine the RME estimator, which can greatly improve estimation if the researcher has an idea of the magnitude of the parameters, but which leads to an increase in MSE if we shrink the parameters toward the wrong values. Although we do not report the results in our tables, we calculate the prediction MSE for a hold-out sample of observations in each of our experiments. The results show that estimators with a lower MSE also predicted better out-of-sample. Thus, the GME and RME estimators generally predict much better for a hold-out sample than does the OLS estimator.

Acknowledgements

References

Belsley, D.A., Kuh, E., and Welsch, R.E., (1980), *Regression Diagnostics*, New York: John Wiley and Sons.

Campbell, R.C., and Hill, R.C., (2004), "Imposing Parameter Inequality Restrictions Using the Principle of Maximum Entropy", working paper.

Davidson, R., and MacKinnon, J.G., (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Fraser, I., (2000), "An Application of Maximum Entropy Estimation: The Demand for Meat in the United Kingdom", *Applied Economics*, 32: 45-59.

Golan, A., Judge, G., and Miller, D., (1996), *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, New York: John Wiley and Sons.

Hendry, D.F., (1984), "Monte Carlo Experimentation in Econometrics", In Griliches, Z., and Intriligator, M.D. (Eds.), *Handbook of Econometrics, Vol. II*, Amsterdam, North-Holland: Elsevier Science Publishers: 937-976.

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., and 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., and Lee, T., (1988), *Introduction to the Theory and Practice of Econometrics*, 2nd ed., New York: John Wiley and Sons.

Mittelhammer, R., and Cardell, S., (1997), "On the Consistency and Asymptotic Normality of the Data Constrained GME Estimator in the GLM", mimeo, Washington State University, Pullman, WA.

Mittelhammer, R.C., Judge, G.G., and Miller, D.J., (2000), *Econometric Foundations*, New York: Cambridge University Press.

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

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

Shen, E.Z., and Perloff, J.M., (2001), "Maximum Entropy and Bayesian Approaches to the Ratio Problem", *Journal of Econometrics*, 104: 289-313.

Tables

Table 1. Dimensions of the Experimental Design

Variable Tested	Location	Sample Size	Error Distribution	Regressor Distribution	Condition Number	Error Variance	Specif. Error(β_2)
Sample size	Table 2	10 to 1000	Normal	Normal	10	1	-1
Sample size	Table 3	10 to 1000	t-	t-	10	1	-1
Sample size	Table 4	10 to 1000	Chi-square	Chi-square	10	1	-1
Error/Regressor Distribution	Table 5	100	N, t, χ^2	N, t, χ^2	10	1	-1
Error Dist.	Table 6	100	N, t, χ^2	Normal	10	5	-1
Regressor Dist.	Table 7	100	Normal	N, t, χ^2	100	1	-1
Condition no.	Table 8	100	Normal	Normal	1 to 500	1	-1
Condition no.	Table 9	100	Normal	Normal	1 to 500	5	-1
Condition no.	Table 10	50	Normal	Normal	1 to 500	1	-1
Signal-to-noise	Table 11	100	Normal	Normal	10	1 to 5	-1
Signal-to-noise	Table 12	100	t-	t-	10	1 to 5	-1
Signal-to-noise	Table 13	100	Chi-square	Chi-square	10	1 to 5	-1
Signal-to-noise	Table 14	100	Normal	Normal	100	1 to 5	-1
Specif. Error	Table 15	100	Normal	Normal	10	1	-1 to 0.2
Specif. Error	Table 16	100	Normal	Normal	40	1	-1 to 0.2
Specif. Error	Table 17	100	Normal	Normal	100	1	-1 to 0.2
Specif. Error	Table 18	100	Normal	Normal	10	5	-1 to 0.2

Table 2. MSE as Sample Size Varies
{**Normal errors, Normal regressors, $\kappa(X_a{}'X_a)=10$, $\sigma^2=1$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$** }

Sample Size	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
$T=10$	19.89	54.937	13.089	4.159	3.504	7.268	0.296
$T=50$	19.09	51.414	9.323	4.955	2.899	7.017	0.751
$T=100$	28.67	50.184	8.743	7.453	2.937	6.578	1.406
$T=300$	27.42	57.605	17.961	30.913	3.308	20.197	4.274
$T=500$	28.20	53.108	8.838	64.210	5.789	27.123	6.045

Table 3. MSE as Sample Size Varies
{*t*- **errors**, *t*- **regressors, $\kappa(X_a{}'X_a)=10$, $\sigma^2=1$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$** }

Sample Size	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
$T=10$	24.53	47.615	21.240	12.600	10.692	13.362	1.226
$T=50$	20.71	62.536	30.959	57.723	36.628	29.520	4.543
$T=100$	17.25	56.902	25.159	113.352	75.353	57.971	7.418
$T=300$	29.38	46.143	13.943	140.067	134.686	55.284	6.848
$T=500$	20.33	53.785	20.172	191.079	133.549	41.626	7.786

Table 4. MSE as Sample Size Varies

{ χ^2 errors, χ^2 regressors, $\kappa(X_a'X_a) = 10$, $\sigma^2 = 1$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Sample Size	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
$T = 10$	18.12	48.853	14.701	8.616	7.438	18.798	0.779
$T = 50$	21.81	56.525	14.093	33.057	10.874	26.132	4.220
$T = 100$	26.68	58.499	30.037	73.885	22.422	46.861	7.598
$T = 300$	24.19	54.364	28.667	159.461	80.264	58.690	8.868
$T = 500$	26.69	48.915	10.362	148.005	122.170	45.645	5.789

Table 5. MSE as Distribution of Regressors and Errors Varies

{T = 100, $\kappa(X_a'X_a) = 10$, $\sigma^2 = 1$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Regressors/ Errors	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
N / N	28.67	50.184	8.743	7.453	2.937	6.578	1.406
N / t	28.67	50.472	8.788	108.789	66.845	43.160	8.064
N / χ^2	28.67	55.808	9.227	75.500	18.087	33.780	7.443
t / N	17.25	54.386	21.930	15.090	9.592	21.082	1.579
t / t	17.25	56.902	25.159	113.352	75.353	57.971	7.418
t / χ^2	17.25	54.470	21.298	87.336	28.690	45.041	7.040
χ^2 / N	26.68	50.264	25.857	9.234	5.494	22.574	1.404
χ^2 / t	26.68	51.228	23.783	113.859	68.548	59.978	8.402
χ^2 / χ^2	26.68	58.499	30.037	73.885	22.422	46.861	7.598

Table 6. MSE as Distribution of Errors Varies

{T = 100, Normal regressors, $\kappa(X_a'X_a) = 10$, $\sigma^2 = 5$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Regressors/ Errors	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
N / N	5.73	250.922	19.844	16.670	7.345	19.178	2.094
N / t	5.73	252.358	19.240	159.310	104.763	63.016	7.245
N / χ^2	5.73	279.041	20.106	121.995	40.102	54.086	7.135

Table 7. MSE as Distribution of Regressors Varies

{T = 100, Normal errors, $\kappa(X_a'X_a) = 100$, $\sigma^2 = 1$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Regressors/ Errors	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
N / N	28.89	4708.97	10.901	4.817	2.624	6.545	1.399
t / N	17.24	5132.05	56.832	14.522	10.538	22.482	1.548
χ^2 / N	26.85	4744.44	141.346	8.402	5.823	23.957	1.413

Table 8. MSE as Degree of Collinearity Varies

{T = 100, Normal errors, Normal regressors, $\sigma^2 = 1$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

$\kappa(X_a'X_a)$	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
1	18.00	4.11	5.92	5.20	3.40	6.95	1.46
10	28.67	50.18	8.74	7.45	2.94	6.58	1.41
20	28.83	191.36	10.04	6.61	2.71	6.54	1.39
40	28.87	756.06	10.58	5.71	2.64	6.56	1.39
60	28.88	1697.23	10.75	5.38	2.63	6.53	1.40
80	28.88	3014.86	10.83	5.17	2.63	6.53	1.40
100	28.89	4708.97	10.90	4.82	2.62	6.55	1.40
250	28.89	29414.68	10.93	4.71	2.62	6.54	1.40
500	28.89	117649.34	10.93	4.69	2.62	6.54	1.40

Table 9. MSE as Degree of Collinearity Varies

{T = 100, Normal errors, Normal regressors, $\sigma^2 = 5$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

$\kappa(X_a'X_a)$	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
1	3.60	20.54	14.58	18.12	9.86	24.00	2.05
10	5.73	250.92	19.84	16.67	7.34	19.18	2.09
20	5.77	956.78	20.35	16.23	7.27	19.14	2.09
40	5.77	3780.28	20.46	15.67	7.26	19.14	2.08
60	5.78	8486.13	20.45	15.70	7.25	19.12	2.12
80	5.78	15074.32	20.39	15.71	7.25	19.11	2.07
100	5.78	23544.85	20.36	15.71	7.25	19.11	2.07
250	5.78	147073.37	20.24	15.52	7.25	19.10	2.07
500	5.78	588246.69	20.20	15.54	7.25	19.10	2.07

Table 10. MSE as Degree of Collinearity Varies

{T = 50, Normal errors, Normal regressors, $\sigma^2 = 1$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

$\kappa(X_a'X_a)$	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
1	18.00	4.01	3.67	3.82	3.24	6.02	0.75
10	19.09	51.41	9.32	4.96	2.90	7.02	0.75
20	19.10	196.45	10.41	4.25	2.65	7.16	0.74
40	19.11	776.63	11.06	3.82	2.58	7.11	0.74
60	19.11	1743.58	11.19	3.67	2.56	7.05	0.74
80	19.11	3097.32	11.23	3.62	2.56	7.05	0.74
100	19.11	4837.84	11.23	3.61	2.56	7.06	0.74
250	19.11	30220.44	11.29	3.42	2.55	7.04	0.74
500	19.11	120872.59	11.32	3.40	2.55	7.03	0.74

Table 11. MSE as Signal-to-Noise Ratio Varies
{T = 100, Normal errors, Normal regressors, $\kappa(X_a'X_a) = 10$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Error Variance	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
$\sigma^2 = 1$	28.67	50.184	8.743	7.453	2.937	6.578	1.406
$\sigma^2 = 2$	14.34	100.369	12.139	9.932	4.396	10.610	1.688
$\sigma^2 = 3$	9.56	150.553	14.973	12.032	5.573	13.846	1.887
$\sigma^2 = 4$	7.17	200.738	17.496	14.394	6.536	16.656	2.015
$\sigma^2 = 5$	5.73	250.922	19.844	16.670	7.345	19.178	2.094

Table 12. MSE as Signal-to-Noise Ratio Varies
{T = 100, *t*- errors, *t*- regressors, $\kappa(X_a'X_a) = 10$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Error Variance	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
$\sigma^2 = 1$	17.25	56.902	25.159	113.352	75.353	57.971	7.418
$\sigma^2 = 2$	8.63	113.805	40.200	131.597	89.445	65.574	7.081
$\sigma^2 = 3$	5.75	170.707	52.954	144.537	98.186	69.618	6.749
$\sigma^2 = 4$	4.31	227.610	64.807	152.950	100.932	70.639	6.520
$\sigma^2 = 5$	3.45	284.512	76.113	159.697	101.889	71.219	6.204

Table 13. MSE as Signal-to-Noise Ratio Varies
{T = 100, χ^2 errors, χ^2 regressors, $\kappa(X_a'X_a) = 10$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Error Variance	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
$\sigma^2 = 1$	26.68	58.499	30.037	73.885	22.422	46.861	7.598
$\sigma^2 = 2$	13.34	116.997	51.805	93.454	30.960	54.639	7.944
$\sigma^2 = 3$	8.89	175.496	70.649	108.682	37.223	58.785	7.880
$\sigma^2 = 4$	6.67	233.995	88.350	119.405	41.962	61.786	7.797
$\sigma^2 = 5$	5.34	292.494	105.413	127.814	45.594	63.669	7.584

Table 14. MSE as Signal-to-Noise Ratio Varies
{T = 100, Normal errors, Normal regressors, $\kappa(X_a{'}X_a) = 100$, restrictions: $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Error Variance	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
$\sigma^2 = 1$	28.89	4708.97	10.901	4.817	2.624	6.545	1.399
$\sigma^2 = 2$	14.44	9417.94	13.508	8.131	4.215	10.584	1.705
$\sigma^2 = 3$	9.63	14126.91	15.948	11.016	5.440	13.886	1.900
$\sigma^2 = 4$	7.22	18835.88	18.231	13.472	6.428	16.591	2.019
$\sigma^2 = 5$	5.78	23544.85	20.356	15.714	7.251	19.110	2.067

Table 15. MSE as Specification Error Varies
{T = 100, Normal errors, Normal regressors, $\kappa(X_a{'}X_a) = 10$, $\sigma^2 = 1$, restrictions: $\beta_1 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Restriction	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
$\beta_2 \geq 0$	28.67	50.184	8.743	7.453	2.937	6.578	1.406
$\beta_2 \geq 0.5$	28.67	50.184	8.152	7.453	2.937	7.404	1.542
$\beta_2 \geq 0.8$	28.67	50.184	7.962	7.453	2.937	7.934	1.601
$\beta_2 \geq 0.9$	28.67	50.184	7.932	7.453	2.937	8.123	1.621
$\beta_2 \geq 1$	28.67	50.184	7.919	7.453	2.937	8.318	1.690
$\beta_2 \geq 1.1$	28.67	50.184	7.924	7.453	2.937	8.528	1.712
$\beta_2 \geq 1.2$	28.67	50.184	7.946	7.453	2.937	8.740	1.765

Table 16. MSE as Specification Error Varies
T = 100, Normal errors, Normal regressors, $\kappa(X_a{'}X_a) = 40$, $\sigma^2 = 1$, restrictions: $\beta_1 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Restriction	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
$\beta_2 \geq 0$	28.87	756.057	10.578	5.714	2.643	6.559	1.394
$\beta_2 \geq 0.5$	28.87	756.057	9.894	5.714	2.643	7.415	1.538
$\beta_2 \geq 0.8$	28.87	756.057	9.688	5.714	2.643	8.009	1.597
$\beta_2 \geq 0.9$	28.87	756.057	9.656	5.714	2.643	8.174	1.609
$\beta_2 \geq 1$	28.87	756.057	9.642	5.714	2.643	8.372	1.677
$\beta_2 \geq 1.1$	28.87	756.057	9.648	5.714	2.643	8.451	1.705
$\beta_2 \geq 1.2$	28.87	756.057	9.673	5.714	2.643	8.805	1.759

Table 17. MSE as Specification Error Varies
T = 100, Normal errors, Normal regressors, $\kappa(X_a'X_a) = 100$, $\sigma^2 = 1$, restrictions: $\beta_1 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Restriction	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
$\beta_2 \geq 0$	28.89	4708.970	10.901	4.817	2.624	6.545	1.399
$\beta_2 \geq 0.5$	28.89	4708.970	10.211	4.817	2.624	7.389	1.540
$\beta_2 \geq 0.8$	28.89	4708.970	10.006	4.817	2.624	7.952	1.594
$\beta_2 \geq 0.9$	28.89	4708.970	9.975	4.817	2.624	8.149	1.614
$\beta_2 \geq 1$	28.89	4708.970	9.962	4.817	2.624	8.358	1.682
$\beta_2 \geq 1.1$	28.89	4708.970	9.968	4.817	2.624	8.561	1.710
$\beta_2 \geq 1.2$	28.89	4708.970	9.992	4.817	2.624	8.789	1.763

Table 18. MSE as Specification Error Varies
T = 100, Normal errors, Normal regressors, $\kappa(X_a'X_a) = 10$, $\sigma^2 = 5$, restrictions: $\beta_1 \geq 0$, $\beta_3 \leq 0$, $\beta_4 \geq 0$ }

Restriction	Signal-to-Noise	OLS	IRLS	GME3	GME4	RME3-I	RME3-C
$\beta_2 \geq 0$	5.73	250.922	19.844	16.670	7.345	19.178	2.094
$\beta_2 \geq 0.5$	5.73	250.922	19.255	16.670	7.345	20.011	2.201
$\beta_2 \geq 0.8$	5.73	250.922	19.052	16.670	7.345	20.567	2.247
$\beta_2 \geq 0.9$	5.73	250.922	19.012	16.670	7.345	20.756	2.268
$\beta_2 \geq 1$	5.73	250.922	18.988	16.670	7.345	20.949	2.329
$\beta_2 \geq 1.1$	5.73	250.922	18.980	16.670	7.345	21.144	2.354
$\beta_2 \geq 1.2$	5.73	250.922	18.987	16.670	7.345	21.312	2.396

Table 19. Results from Response Surface Regressions (t-values in parentheses)

Variable	IRLS	GME3	GME4	RME3-I	RME3-C
Constant	0.12444	-0.17813	-0.17592	0.08217	0.01442
	(2.10)	(-1.88)	(-2.18)	(0.94)	(1.02)
Sample Size	-0.00008	0.00475	0.00325	0.00092	0.00019
	(-0.30)	(10.00)	(8.02)	(2.39)	(3.01)
t- Errors	0.04093	1.28927	0.89463	0.38590	0.06536
	(0.42)	(7.37)	(6.00)	(2.71)	(2.83)
Chi-Sq Errors	0.01318	0.86167	0.29855	0.23067	0.05234
	(0.14)	(4.92)	(2.00)	(1.62)	(2.27)
t- Regressors	0.12382	-0.13686	-0.02077	-0.05414	-0.03147
	(1.29)	(-0.79)	(-0.14)	(-0.38)	(-1.37)
Chi-Sq Regressors	0.16842	-0.02939	0.08980	0.05455	-0.02000
	(1.76)	(-0.17)	(0.61)	(0.38)	(-0.87)
Condition #	-0.00071	-0.00034	-0.00010	-0.00086	-0.00013
	(-3.02)	(-0.81)	(-0.27)	(-2.44)	(-2.26)
σ^2	-0.01539	-0.07309	-0.04339	-0.04107	-0.00894
	(-1.17)	(-3.04)	(-2.12)	(-2.11)	(-2.84)
δ	0.03300			0.06941	0.01097
	(0.13)			(0.19)	(0.18)
δ^2	0.14454			0.23856	0.03687
	(0.57)			(0.64)	(0.61)
R^2	0.334	0.795	0.704	0.442	0.417