CEFAGE-UE Working Paper
2010/02

GME versus OLS
Which is the best to estimate utility functions?

Cesaltina Pires, Andreia Dionísio, Luís Coelho

Departamento de Gestão, Universidade de Évora and CEFAGE-UE

GME *versus* OLS - which is the best to estimate utility functions?

Cesaltina Pires, Andreia Dionísio, Luís Coelho

Universidade de Évora, Cefage-UE

Largo dos Colegiais, 2, 7000 Évora, Portugal

March 25, 2010

Abstract

This paper estimates von Neumann and Morgenstern utility functions comparing the generalized maximum entropy (GME) with OLS, using data obtained by utility elicitation methods. Thus, it provides a comparison of the performance of the two estimators in a real data small sample setup. The results confirm the ones obtained for small samples through Monte Carlo simulations. The difference between the two estimators is small and it decreases as the width of the parameter support vector increases. Moreover the GME estimator is more precise than the OLS one. Overall the results suggest that GME is an interesting alternative to OLS in the estimation of utility functions when data is generated by utility elicitation methods.

Keywords: Generalized maximum entropy; Maximum entropy principle; von Neumann and Morgenstern utility; Utility elicitation

JEL classification: C13; C14; C49; D81

1 Introduction

In this paper we explore the potentialities of the generalized maximum entropy estimation of von Neumann and Morgenstern utility functions using only partial information about the agent's preferences and his risk tolerance. The main goal of this study is to compare the results of generalized maximum entropy estimation with traditional estimation methods in order to investigate the possible advantages of the generalized maximum entropy approach.

One of the core assumptions of decision theory is that a decision maker observed behaviour can be rationalized in terms of the underlying preference ordering. One implication of this assumption is that the preference ordering can be inferred from the decision maker observed behaviour. Although economists favour revealed preference data in the estimation of utility functions there are many circumstances where such data is not available. Thus, in order to estimate the decision maker utility function, one often needs to use utility elicitation methods based on surveys or experiments. However the use of utility elicitation methods presents several difficulties with important implications for the decision analyst. First, different methods of utility elicitation yield different results. Thus one needs to be careful in designing the elicitation method in such a way that does not generate biased results and should be cautious in interpreting the results as measurement errors are likely to be present. Second, the methods of elicitation and validation of results are sometimes difficult to implement. Third, the use of elicitation methods is likely to lead to relatively few observations since the decision maker might not be willing to answer to many questions. In this paper we will not explore the problems related with the elicitation process itself. However, in our estimation approach, we take into account the fact that data was generated by utility elicitation methods. In particular, we will explicitly

acknowledge the fact that the number of observations is very small and that there may exist measurement errors.

Since the elicitation of preferences is not a goal of this research, we use data from Abdellaoui, Barrios and Wakker (2007). These authors elicit the preferences of 47 decision makers using a choice-based method (trade-off method) and a choiceless method (strength-of-preference method) and estimate the parameters of three utility functions (power, exponential, and expo-power). In this paper, we apply generalized maximum entropy estimation (GME) to these data and compare it with least squares estimation (OLS). In addition, we use bootstrap to obtain confidence intervals and compare the precision of the GME and the OLS estimators.

Generalized maximum entropy has the ability to estimate the parameters of a regression model without imposing any constrains on the probability distribution of errors and it is robust even when we have ill-posed problems, namely with very small samples (Golan *et al.*, 1996; Campbell and Hill, 2005, 2006).

The ability of GME to estimate economic relationships has already been explored by several authors (see, for instance, Golan and Judje, 1992; Golan and Perloff, 2002; Kitamura and Stuzer, 2002). Most applications of GME relate to ill-posed problems or ill-conditioned problems (non-stationarity, number of unknown parameters exceeds the number of observations, collinearity). Golan *et al.* (1994) use GME/GCE to derive a Social Accounting Matrix without having to impose the standard identification procedures. Paris and Howitt (1998) employ GME in a mathematical programming framework to estimate cost parameters for individual farm level data. Fraser (2000) uses GME to estimate the demand for meat in the United Kingdom with data that are subject to collinearity. Golan, Judge, and Perlof (1997) and Golan, Perlof and Shen (2001) use GME to estimate

a censored regression model. More recently, Campbell and Hill (2005, 2006) demonstrate how to impose inequality constraints in GME through the parameter support matrix.

To the best of our knowledge our work is the first one to apply GME in the estimation of utility functions and to provide evidence on the comparison of GME with least squares estimation with real data small samples. This paper is organized as follows: Section 2 presents the concepts of entropy, maximum entropy principle and generalized maximum entropy estimation. In Section 3 we present the estimation process and the data we used in this paper. The main results are presented in Section 4, where we compare the results of GME and OLS estimation. The last section of the paper contains the main conclusions of this research.

2 Entropy, maximum entropy principle and GME

2.1 Background theory

In order to understand the concepts underlying generalized maximum entropy estimation (GME) it is worth explaining the meaning of entropy and the maximum entropy principle, since these two concepts are important building blocks of GME. The concept of entropy was introduced by Shannon (1948) in the context of information theory. Entropy is a measure of uncertainty or a measure of the information generated by observing the outcome of a random event. Consider a future event X with n possible outcomes and let $(p_1, p_2, \cdots, p_n)$ be the associated probability distribution. If a particular outcome has a high probability we will not be very surprised if it occurs or, in other words, observing that outcome is not very informative. This suggests that the information contained in one observation is inversely related with its probability, a feature which is well captured if one

uses the $-\log(\cdot)$ function to measure the information content. The concept of entropy proposed by Shannon applies to events that have not occurred yet, thus the relevant measure is the expected information: the expected value of the $-\log(\cdot)$ function. The entropy of the probability distribution $\mathbf{p} = (p_1, p_2, \cdots, p_n)$ is the measure $H(\mathbf{p})$ given by:

$$H(\mathbf{p}) \equiv \sum_{i=1}^{n} p_i(-\ln p_i) = -\sum_{i=1}^{n} p_i \ln p_i$$

and $0 \cdot \ln(0) = 0$. When one outcome has $p_k = 1$ (the remaining have $p_i = 0$, for $i \neq k$) the entropy is $H(\mathbf{p}) = 0$. On the other hand when all outcomes are equally likely, $p_i = \frac{1}{n}$ for all i, $H(\mathbf{p})$ reaches a maximum.

According to Laplace's principle of insufficient reason, if we have no information about a particular probability distribution, then one should assume that all outcomes are equally likely. Consequently the principle of insufficient reason corresponds to the maximization of the entropy function. The maximum entropy (ME) principle proposed by Jaynes (1957) generalizes the previous idea to a context where there exists some information on the probability distribution. For instance, if particular moments of the probability distribution generating the data are know then, among the probability distributions that satisfy these moment conditions, one should choose the one that maximizes entropy. In other words, we choose $\mathbf{p}$ so as to maximize $H(\mathbf{p})$ subject to all the existent restrictions (moment conditions, cumulative probability constraints, adding constraints) on the random variable X. Considering m moment or cumulative probability constraints, the ME problem can be written as follows:

$$\begin{aligned} &\mathbf{p}^* = \arg\max -\sum_i p \ln p_i, \quad \text{s.t.} \\ &\textstyle\sum_{i=1}^{n} p_i = 1 \\ &\textstyle\sum_{i=1}^{n} h_j(X_i)\, p_i = b_j \qquad j = 1, ..., m \\ &p_i \geq 0 \qquad\qquad\qquad i = 1, ..., n. \end{aligned} \tag{1}$$

When restriction j is a cumulative probability constraint, $h_j(X_i)$ is an indicator function over an interval for cumulative probability and b_j is the value of the cumulative probability. When restriction j is a moment constraint $h_j(X_i)$ is equal to X_i, $X_i^2, \cdots$ (depending on the moment being considered) and b_j is the corresponding moment of the distribution. It should be highlighted that the solution of the ME procedure is consistent with the known information but it entails maximal uncertainty regarding all the other things. According to several authors (see for example Soofi, 2000, and Golan, 2002) this principle uses only relevant information and eliminates all irrelevant details from the calculations by averaging over them. The ME principle can also be applied to the density function of continuous random variables.

In spite of the potentialities of the ME approach, the Generalized Maximum Entropy (GME) proposed and employed by Judge and Golan (1992) increases substantially the number of possible extensions and applications in economics, particularly when we have ill-posed problems in linear models. In order to explain GME conveniently, let us consider the following standard regression model in matrix form:

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

where $\mathbf{y}$ is a $(T \times 1)$ vector, $\mathbf{X}$ is a $(T \times K)$ matrix, β is a $(K \times 1)$ vector and $\mathbf{e}$ is a $(T \times 1)$ vector of errors. According to Golan *et al.* (1996) the problem is ill-posed if there is not enough information contained in $\mathbf{X}$ and the noisy data $\mathbf{y}$ to permit the recovery of the K-dimensional β parameter vector by traditional estimation methods. A ill-posed problem may arise from several sources: (i) non-stationarity or other model specification issues may cause the number of unknown parameters to exceed the number of data points; (ii) the data is mutually inconsistent; (iii) the experiment may be badly designed, which may cause collinearity in $\mathbf{X}$. The use of OLS (or other traditional estimation methods) in

those situations may cause: (a) arbitrary parameters; (b) the solution may be undefined; and (c) the estimates can be highly unstable given the high variance or the low precision of the estimated parameters.

The GME generalizes the maximum entropy approach and reparameterizes the linear model such that the unknown parameters and the unknown errors are in the form of probabilities. In other words, the GME estimates the probability distribution of each of the K parameters and estimates the probability distribution of each of the T errors. Judge and Golan (1992) assume that each β_k can be viewed as a discrete random variable with M potential outcomes (with $2 \leq M \leq \infty$). Let $\mathbf{z}_k$ be the $(M \times 1)$ support vector for parameter β_k and assume that the elements in $\mathbf{z}_k$ are ordered from the lowest, z_{k1}, to the largest, z_{kM}. Then, for each parameter β_k, there exists a probability vector $\mathbf{p}_k$ such that

$$\beta_k = \begin{bmatrix} z_{k1} & \cdots & z_{kM} \end{bmatrix} \begin{bmatrix} p_{k1} \\ \vdots \\ p_{kM} \end{bmatrix}. \tag{3}$$

The parameter support is based on prior information of economic theory or in previous experiences (see for example, Fraser, 2000; and Campbell and Hill, 2005, 2006).

Generalizing for all k (with $k = 1, \cdots, K$), the vector β can be defined by:

$$\beta = \mathbf{Z}\mathbf{p} = \begin{bmatrix} \mathbf{z}_1' & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_2' & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{z}_K' \end{bmatrix} \cdot \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_K \end{bmatrix} \tag{4}$$

where $\mathbf{Z}$ is a $(K \times KM)$ matrix of support points and $\mathbf{p}$ is a $(KM \times 1)$ vector of unknown weights such that $p_{km} > 0$ and $\mathbf{p}_k' \mathbf{i}_M = 1$ for all k (where $\mathbf{i}_M$ is a $M \times 1$ vector of ones). In a similar way, to estimate the distribution of the errors, admit that we have $\mathbf{J} \geq 2$

support points and let $\mathbf{V}$ be the $(T \times TJ)$ matrix of support points, and $\mathbf{w}$ be a $(TJ \times 1)$ vector of weights on these support points. The unknown errors are defined as

$$\mathbf{e} = \mathbf{V}\mathbf{w} = \begin{bmatrix} \mathbf{v}_1' & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_2' & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{v}_T' \end{bmatrix} \cdot \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_T \end{bmatrix}. \tag{5}$$

Once again, the weights satisfy $w_{tj} > 0$ and $\mathbf{w}'_t \mathbf{i}_J = 1$ for all t.

The model can be reparameterized in matrix form as follows

$$\mathbf{y} = \mathbf{X}\mathbf{Z}\mathbf{p} + \mathbf{V}\mathbf{w}, \tag{6}$$

where $\mathbf{y}$, $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{V}$ are known and we estimate the unknown $\mathbf{p}$ and $\mathbf{w}$ vectors using the maximum entropy approach. The regression model can be expressed as follows

$$\max H(\mathbf{p}, \mathbf{w}) = -\mathbf{p}' \ln \mathbf{p} - \mathbf{w}' \ln \mathbf{w} \tag{7}$$

$$s.t.$$

$$\mathbf{y} = \mathbf{X}\mathbf{Z}\mathbf{p} + \mathbf{V}\mathbf{w} \tag{8}$$

$$(\mathbf{I}_K \otimes \mathbf{i}'_M)\,\mathbf{p} = \mathbf{i}_K \tag{9}$$

$$(\mathbf{I}_T \otimes \mathbf{i}'_J)\,\mathbf{w} = \mathbf{i}_T, \tag{10}$$

where $\otimes$ is the Kronecker product. Equation (8) is a data constraint and equations (9) and (10) are additivity constraints, which require that the probabilities sum to one for each of the K parameters and T errors.

The solution of the GME model (Golan *et al.*, 1996) is:

$$\hat{p}_{km} = \frac{\exp\left(z_{km}\mathbf{x}'_k\hat{\lambda}\right)}{\sum_{m=1}^{M}\exp\left(z_{km}\mathbf{x}'_k\hat{\lambda}\right)} \qquad m = 1, \cdots, M \quad \text{and} \quad k = 1, \cdots, K \tag{11}$$

and

$$\hat{w}_{tj} = \frac{\exp\left(v_{tj}\hat{\lambda}_t\right)}{\sum_{j=1}^{J}\exp\left(v_{tj}\hat{\lambda}_t\right)}, \qquad j = 1, \cdots, J \quad \text{and} \quad t = 1, \cdots, T \tag{12}$$

where $\hat{\lambda}$ is the vector of Lagrange multipliers for the data constraints.

The endpoints of the probability supports can be either symmetric or asymmetric, depending on the problem under consideration. For the error term is usual to use a symmetric representation centred on zero. Pukelsheim (1994) suggests setting error bounds as $v'_{t1} = -3\sigma$ and $v'_{tJ} = 3\sigma$, where σ is the standard-deviation of e. Since σ is unknown, it can be replaced by an estimate. Campbell and Hill (2006) consider two possibilities: (1) $\hat{\sigma}$ from the OLS regression, and (2) the sample standard deviation of y. These authors found better results using the second alternative.

Fraser (2000) alerts to the problem of choosing the number of elements in the supports (M and J). Generally, M and J are determined, more by computational time rather than accuracy of estimation. Increasing the number of points in a support matrix and maintaining the width, the variance of the uniform distribution will decrease. Of course it will need more computational time, but using Gauss 8.0 the calculus are quite fast. Mittelhammer *et al.* (2002) conclude that the quality of the GME estimates depends on the quality of the supports chosen, and the Monte Carlo experiments suggest that the GME with wide supports will often perform better than OLS, especially for small samples ($n < 25$).

According to Mittelhammer *et al.* (2002) the GME estimator is not constrained by any extraneous assumptions. The information used is the observed information contained in the data, the information contained in the constraints and the information of the structure of the model, including the choices of the supports for the $\beta'_k s$.

2.2 Applications of entropy, ME and GME in decision analysis

The applications of entropy, ME and GME to economics are not new and cover a wide set of fields, including finance, game theory and agricultural economics (see for example Buchen and Kelly, 1996; Gulko, 1998; Samperi, 1999; Stuzer, 1996, 2000). In this section we briefly review recent applications related with decision analysis with particular emphasis on the estimation of utility functions.

Yang and Qiu (2005) use the notion of entropy and propose an expected utility-entropy measure of risk aversion in portfolio management. The authors conclude that using this approach it is possible to solve a class of decision problems which cannot be dealt with the expected utility or mean-variance criterion. The ME principle has also been applied in decision analysis. For example, Fritelli (2000) derives the relative entropy minimizing martingale measure under incomplete markets and demonstrates its connection with the maximization of exponential utility. In a different approach Abbas (2004) presents an optimal question-algorithm to elicit von Neumann and Morgenstern utility values using the ME principle and Abbas (2006a) uses the discrete form of ME principle to obtain a joint probability distribution using lower order assessments. Sandow *et al.* (2006) use the minimization of cross-entropy (or relative entropy) to estimate the conditional probability distribution of the default rate as a function of a weighted average bond rating, concluding that the modeling approach is asymptotically optimal for an expected utility maximizing investor. Friedman *et al.* (2007) explore an utility-based approach to some information measures, namely the Kullback-Leibler relative entropy and entropy using the example of horse races. On the other hand, Darooneh (2006) uses the ME principle to find the utility function and the risk aversion of agents in a exchange market.

As described in the previous section, the ME principle applies to estimation of prob-

ability distributions. However the principle has also been applied in the estimation of utility functions describing ordinal preferences by Herfert and La Mura (2004) and in the estimation of von Neumann and Morgenstern utility functions by Abbas (2006b). Herfert and La Mura (2004) scale utility so as to behave as a probability measure whereas Abbas (2006b) scales utility so that marginal utility behaves as a probability measure. Since we are interested in estimating von Neumann and Morgenstern utility functions let us analyze in greater detail Abbas's approach.

The main insight in Abbas (2006b) is to draw an analogy between utility and probability and to use this analogy to apply ME to the estimation of the utility function. The starting point is to normalize the utility function such that it ranges between zero and one. Then, defining the utility density function as the derivative of the normalized utility function one can show that the utility density function is non-negative and integrates to unit, precisely the same properties that define a probability density function. Consequently, the ME principle can be applied to the estimation of the utility density function when there exists partial information on the agent's preferences. For instance, if one only knows the preference ordering then the ME principle would imply that the marginal utilities are constant and hence the utility function is a linear function. On the other hand, if some utility values are known the ME utility approach requires that the utility function satisfies these values' constraints. This implies that the maximum entropy utility, as Abbas calls it, is simply the function composed by the linear segments joining the known utility values.

The maximum entropy utility approach proposed by Abbas is conceptually appealing and it has the advantage of making no assumptions on the functional form of the utility function. However the maximum utility entropy also has some drawbacks. The first one

is that it assumes that there does not exist any measurement error in the utility values. However this assumption is likely to be violated when data is obtained through utility elicitation methods. The second problem is that the ME utility is non-differentiable, which contradicts the assumption used in the definition of the utility density function. In addition, non-differentiability may lead to difficulties if one is interested in using the estimated utility to solve decision problems, such as portfolio optimization. For these reasons we have decided not to pursue Abbas approach. Instead we use GME to estimate a particular functional form of the utility function. Note that our approach may be considered semi-parametric, since we assume a particular functional form for the utility function but do not impose restrictions on the probability distribution of the error terms.

3 Estimation and Results

In this section we describe the data used in this paper, as well as the methodology applied to estimate Betas through OLS and GME, and the inference process developed in order to evaluate the accuracy of both methods and the statistical significance of the estimated Betas.

3.1 Estimation of Betas

Our dataset is taken from Abdellaoui, Barrios and Wakker (2007). These authors estimate the parameters of three utility functions (power, exponential, and expo-power) with only one parameter. Abdellaoui, Barrios and Wakker (2007) elicit the preferences of decision makers using a choice-based method (trade-off method) and a choiceless method (strength-of-preference method). The experience was made with 47 students from the

Ecole Normale Supérieure of Cachan. The authors used a software programme for observing indifferences while avoiding bias. The trade-off method was carried out before the strength-of-preference method because its answers were used as inputs in the second method.

We use a parametric family of utility function, more specifically a power function, that is defined by

$$\begin{cases} y = x^r \text{ if } r > 0 \\ y = \ln x \text{ if } r = 0 \\ y = -x^r \text{ if } r < 0 \end{cases},$$

where r is a parameter related with the coefficient of risk aversion (the relative risk aversion is equal to $r - 1$), x is the income and y are the utility values. Note that the power utility function has constant relative risk aversion (CRRA).

In order to estimate the parameter r, we logarithmize both side of the equation and use the following econometric model (for example, assuming that $r > 0$)

$$\ln y = r \ln x + e$$

$$Y = \beta X + e \tag{13}$$

where $Y = \ln y$, $X = \ln x$ and $\beta = r$. The model (13) will be estimated using OLS (Ordinary Least Squares) and GME.[1]

In our database, we have 6 observations for each participant, which is clearly a very small sample to estimate the participant's utility function. This fact may be problematic

[1]We used normalized data for comparative effects. Utility values are between 0 (worst outcome) and 1 (best outcome). Income values are also normalized between 0 (lowest income) and 1 (highest income). This explains why the regression does not include an intercept term. Note that the observation corresponding to the lowest income was not considered in the estimation as ln0 cannot be computed. Since utility values are continuous it is not correct to use logit models.

when we apply the most traditional method of regression estimation: OLS. With small samples, traditional statistical inference under OLS is based on the assumption that residuals are normally distributed (with large samples no such assumption is needed). However, if data is not normally distributed the traditional OLS inference results will be incorrect under small samples. Given this context, it is interesting to investigate whether GME is a better estimator than OLS or not. In this paper, the main reason to use GME is precisely the fact that our samples are very small. To use GME we do not need to assume any theoretical probability distribution for the residuals even to make statistical inference.

In a first phase, we estimate the utility function parameter using the Least Squares approach. We found that all the Betas are positive, being the highest 2.6429 and the smallest 0.5925. For all the regression models estimated we tested the Gauss-Markov assumptions, and we found that the hypothesis of normality, no autocorrelation and homocedasticity are not rejected for any of the 47 estimated regressions. It should be noted that, with such small number of observations, the power of the normality tests is small (meaning that the null may be false and yet we fail to reject it). Thus it seems sensible to use GME and compare it with OLS.

Given the results of the estimated Betas through OLS we specify the support vector with different M for β. We used three different support vectors in order to estimate Betas: parameter support A: $M = 9$; parameter support B: $M = 17$; parameter support C: $M = 37$ (see Table 1).

The prior mean, considering that all support points are equally likely, is 1 for all the support vectors. This prior mean is equal to the mean of the 47 Betas estimated through Least Squares and it is also the one more justifieable in theoretical terms sonce

Parameter Support
$z_k' = [-1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3]$
$z_k' = [-3, -2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4, 5, 5]$
$z_k' = \begin{bmatrix} -8, -7.5, -7, -6.5, -6, -5.5, -5, -4.5, -4, -3.5, -3, -2.5, -2, 1.5, \\ -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4, 5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10 \end{bmatrix}$

Table 1: Parameter Support for GME

$\beta = 1$ corresponds to risk-neutrality.[2] Note that as we go from support vector A, to support vector B, to support vector C the width of the support vector becomes larger (the distance between two contiguous support points is always the same and equal to 0.5). We also defined the error support for our GME estimates, following the considerations of Campbell and Hill (2006). Using the mean of the standard-deviation from the OLS regressions, which equal to 0.3118, the $\pm 3\sigma$ rule of Golan *et al.* (1996) and Campbell and Hill (2006) results on the error support of $\mathbf{v}_i' = [-1, -0.5, 0, 0.5, 1]$, with $J = 5$.

Figure 1 compares GME and OLS estimators for each of the 47 regressions, for the three support vectors (A, B and C). The results show that the GME estimates do not differ very much from the OLS estimates in terms of signs and magnitudes for all the support vectors. It is important to mention that as the width of the support vector increases, the difference between the GME and the OLS estimates became smaller. Increasing the width of the support vector means that we are allowing the parameter to take values in a larger interval, which implies that we are giving less information to the problem. A wider support vector implies that the GME estimator will be more determined by data

[2]Note that we use the same support vector for all the individuals. Alternatively we could use specific support vector for each individual (based on the individual OLS estimator).

constraints than by the prior information, which explains why the GME estimator becomes closer to the OLS estimator. On the other hand, as we will see in the next section, assuming a less informative prior decreases the precision of the GME estimator.

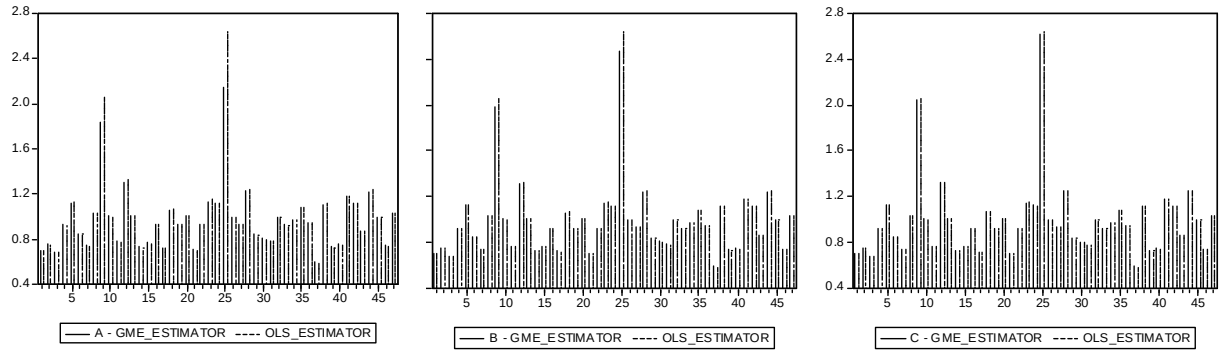


Figure 1: OLS estimator versus GME estimator for all the 47 participants. A -Support vector with $M = 9$, B - Support vector with $M = 17$, C - Support vector with $M = 37$.

All the participants show a positive r (or β) and the majority show a difference between GME estimates and OLS estimates around 0.0005. We can verify that the majority of the participants exhibit a β very close to one, which means that these participants evidence risk neutrality. Only 8 participants evidence risk aversion ($\beta > 1$) and about 6 reveal to be risk lovers ($\beta < 1$).

Figure 2 reveals that exist a strong and linear relationship between the GME and OLS, especially for the GME estimates using a support vector with $M = 37$. The significant similarities between the two estimates evidence the potentialities of the GME estimator, since it is a more general method, without assumptions about the theoretical distributions of variables and residuals. These similarities also point to the good choice of the support values. It is worth mentioning that the support vectors $\mathbf{z}$ were chosen using the results of the OLS estimates as reference.

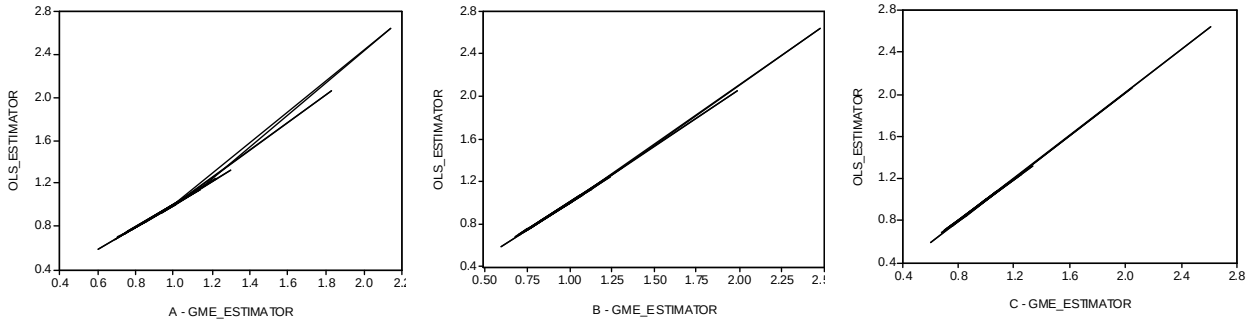


Figure 2: Regression line between OLS estimator and GME estimator. A - Support vector with $M = 9$, B - Support vector with $M = 17$, C - Support vector with $M = 37$.

According to the results obtained, we can say that GME estimates are very close to OLS estimates, and also that GME estimates do not change very much as we change the parameter support (note that for most of the individual regressions the changes are infinitesimal). This means that an uninformative support vector would produce results consistent with OLS and GME estimates obtained with a more informative support vector. This gives confidence and robustness when we choose the support vector and respective size. Next subsection examines the precision of the GME estimator through the use of a bootstrap.

3.2 Confidence Intervals

In order to compare the precision of the GME and OLS estimators we construct confidence intervals. We use a bootstrap to obtain interval estimates for the GME and the OLS estimator, estimating standard errors by resampling the original data. There are several studies that use bootstrap methods to estimate standard errors. For example Horowitz (1997) presents a bootstrap method for computing confidence intervals where

the t-statistics is obtained from the resampled data and the interval estimates are computed as $\hat{\beta} \pm t_c S_{\hat{\beta}}$, where t_c is the bootstrap t-statistics and $S_{\hat{\beta}}$ is the asymptotic standard error of the estimator. Since we do not know the asymptotic distribution of the GME estimator we follow the approach used by Mooney and Duval (1993) and by Campbell and Hill (2006), who used the percentile method to obtain the confidence intervals. It should be noted that we also used the percentile approach to estimate the confidence intervals for the OLS estimators. There are two reasons behind our choice: *(i)* since we are not sure that errors are normally distributed, using the t-statistics to determine OLS confidence intervals with our small sample would not be correct; *(ii)* for comparison purposes it seems more adequate to use the same procedure for the two estimators, OLS and GME.

The confidence intervals for GME and OLS were estimated by resampling from the original data and estimating the model $T = 2000$ times. We then order the resulting estimates and find the 0.05^{th}, 2.5^{th}, 5^{th}, 95^{th}, 97.5^{th} and 99.5^{th} percentiles and construct the confidence intervals with 90%, 95% and 99% confidence level.

As we can see from Figure 3 the GME interval estimates have almost always smaller width than the OLS interval estimates, since most of the differences are negative for all the support vectors we used. This may be a sign of higher precision of the GME estimator. We may note too that as we increase the width of the support vector, the precision of the GME estimates becomes smaller, which implies that the difference in the precision of GME and OLS estimators decreases. Another interesting thing, is the fact that, all the confidence intervals (for GME and OLS estimators) include only positive numbers, excluding the zero in all circumstances (even for a confidence level of 99%). This means that all the parameters estimated (through GME and OLS) are statistically significant at 1% level of significance.

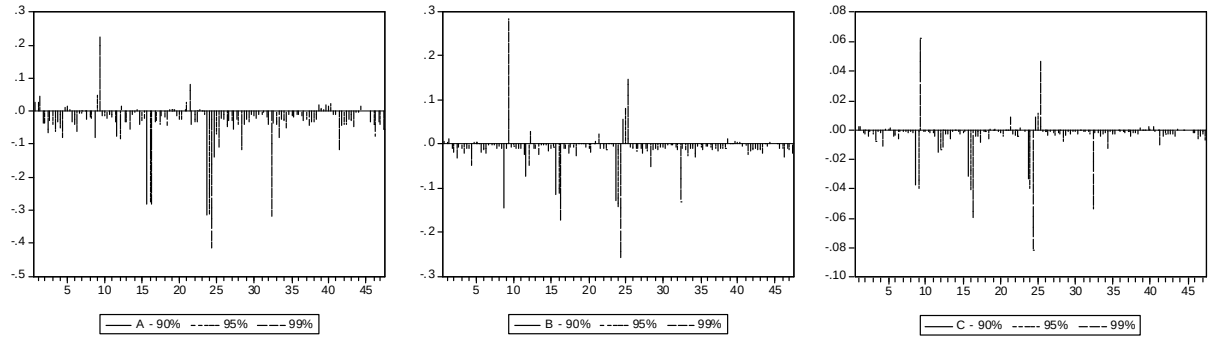


Figure 3: Difference of width between the confidence intervals of the GME and the OLS estimates (calculated by: $Width_{GME} - Width_{OLS}$). A - Support vector with $M = 9$, B - Support vector with $M = 17$, C - Support vector with $M = 37$.

These results are corroborated by the statistical tests for the significance of each β estimated using OLS and the t-Student probability distribution. However, the statistical tests based on t-Student probability distribution are not robust for our estimates, since we have very small samples for each model.

As we mentioned above, the standard error for each Beta was calculated using bootstrap. As we can see from Figure 4, the magnitude of the standard errors for the GME estimator are most of the times, smaller than the standard error of the OLS estimator. This result corroborates the result found about the higher precision of the GME estimator face to the OLS estimator, especially when we deal with small samples.

4 Conclusions

When utility elicitation methods are used to obtain data on the decision maker's preferences it is very likely that the number of observations generated is small. However, with

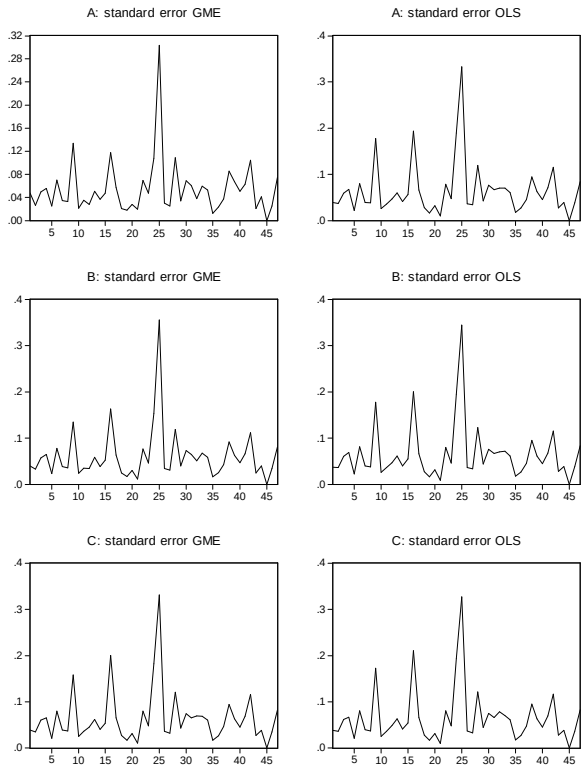


Figure 4: Standard errors for GME and OLS estimates, obtained using bootstrap. A - Support vector with $M = 9$, B - Support vector with $M = 17$, C - Support vector with $M = 37$.

such small samples the use of traditional estimation and inference techniques may not be adequate. Moreover, using Monte Carlo simulations, previous studies have shown that with small samples the generalized maximum entropy estimator outperforms OLS (see, for example, Mittelhammer *et al.*, 2002). This suggests that GME might be particularly well suited to estimate utility function when data is obtained through utility elicitation methods.

In this paper we compare the performance of the GME with OLS in estimating von Neumann and Morgenstern utility functions using data obtained through utility elicita-

tion methods. We used data from Abdellaoui, Barrios and Wakker (2007) who elicited preferences for 47 individuals. An interesting contribution of this paper is that it provides evidence on the comparative performance of GME and OLS under small samples using real data (as opposed to the generated data used in Monte Carlo simulations). Thus our paper complements previous results on the comparison of the GME with OLS.

We reached several interesting conclusions. First, the two estimators do not differ much and their difference becomes smaller as we increase the width of the parameter support vector used in the GME estimation. It is well known that GME combines prior information with observed data. A wider support vector corresponds to less informative priors, so the GME estimator gives more weight to the data and hence it is closer to the OLS. Second, the GME is more precise than OLS leading to less wide confidence intervals, but the difference between the width of the two confidence intervals decreases as we use less informative support vectors. Since GME is a biased estimator, these results show an interesting trade-off between bias and precision of the GME estimator. Using a less informative support vector decreases the bias but it also decreases the precision of the GME estimator. Thus whenever we have prior information about the interval where the parameter varies we should use it, since it enables us to reduce the width of the support vector and consequently increase the precision of the GME estimator. On the other hand, if we have no prior information, one should use a wide support vector. In this case GME still outperforms OLS in terms of precision, but the difference in their relative performance is small.

Acknowledgement

Financial support from Fundacao da Ciencia e Tecnologia, Lisbon, is gratefully acknowledged by the first and third authors, under the contract PTDC/GES/70529/2006

and the second author, under the contract PTDC/GES/73418/2006.

The authors are also grateful to Mohammed Abdellaoui, Carolina Barrios and Peter Wakker for allowing us to use their dataset.

References

[1] Abbas, A., (2006*a*). Entropy methods for joint distributions in decision analysis. *IEEE Transactions on Engineering Management*, **53**, 146-159.

[2] Abbas, A., (2006*b*). Maximum entropy utility. *Operations Research*, **54**.

[3] Abbas, A., (2004). Entropy methods for adaptive utility elicitation. *IEEE Transactions on Systems, Man and Cybernetics*, **34**, 169-178.

[4] Abdellaoui, M., Barrions C. and Wakker, P., 2007. Reconciling introspective utility with revealed preference: Experimental arguments based on prospect theory. *Journal of Econometrics*, **138**, 356-378.

[5] Campbell, R. and Carter Hill, R., (2005). A Monte Carlo study of the effect of design characteristics on the inequality restricted maximum entropy estimator. *Review of Applied Economics*, **1**, 53-84.

[6] Campbell, R. and Carter Hill, R., (2006). Imposing parameter inequality restrictions using the principle of maximum entropy. *Journal of Statistical Computation and Simulation*, **76**, 985-1000.

[7] Darooneh, A., (2006). Utility Function from Maximum Entropy Principle. *Entropy*, **8**, 18-24.

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

[9] Friedman, C.; Huang, J. and Sandow, S., (2007). A utility-based approach to some information measures. *Entropy*, **9**, 1-26.

[10] Fritelli, M., (2000). Entropy and utility maximization in Finance. *Mathematical Finance*, **10**, 39-52.

[11] Golan, A., (2002). Information and entropy econometrics–Editor's View. *Journal of Econometrics*,**107**, 1-15.

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

[13] Golan. A. Judge, G. and Perlof, J., (1997). Estimation and inference with censored and ordered multinomial response data. *Journal of Econometrics*, **79**, 25-51.

[14] Golan, A., Perloff, J. and Shen, E., (2001). Estimating a demand system with nonnegativity constraints: mexican meat demand. *The Review of Economics and Statistics*, **83**(3), 541-550.

[15] Gulko, B., (1998). *The Entropy Pricing Theory – Market Beliefs, Valuation, Incomplete Markets*, Asset Pricing, PhD Dissertation, UMI – Dissertations Services.

[16] Herfert, M. and La Mura, P., (2004). Estimation of consumers preferences via ordinal decision-theoretic entropy, Working-paper 64, Leipzig Graduate School of Management.

[17] Judge, G. G. and Golan, A., (1992) Recovering information in the case of ill-posed inverse problems with noise. Mimeo Department of Agricultural and Natural Resources, University of California, Berkeley, CA.

[18] La Mura, P., (2003). *Decision-Theoretic Entropy. Proceedings of the 9th Conference on Theoretical Aspects of Rationality and Knowledg*e, Bloomington, Indiana, USA.

[19] Mittelhammer, R. Cardell, N. and Marsh, T. (2002). The data-constrained GME estimator of the GLM: asymptotic theory and inference, Mimeo, University of Washington.

[20] Mooney, C. and Duval, R., (1993). Bootstrapping: a nonparametric approach to statistical inference. Gage University Paper Series on Quantitative Applications in the Social Sciences, 07-095, Newbury Park, CA, Sage.

[21] Paris, Q. and Howitt, R., (1998). An analysis of ill-posed production problems using maximum entropy. *American Journal of Agricultural Economics*, **80**, 128-138.

[22] Pukelsheim, F., (1994). The three sigma rule. *American Statistician*, **48**, 88-91.

[23] Samperi, D., (1999). Entropy and Statistical Model Selection for Asset Pricing and Risk Management, preprint in http://papers.ssrn.com

[24] Sandow, S.; Friedman, C.; Gold, M. and Chang P., (2006). Economy-wide bond default rates: a maximum expected utility approach. *Journal of Banking and Finance*, **30**, 679-693.

[25] Shannon, C., (1948). A Mathematical Theory of Communication. *Bell Systems Tech.*, **27**, 379-423, 623-656.

[26] Soofi, E., (2000). Principal information theoretic approaches. *Journal of the American Statistical Association*, **95**, 1349-1353.

[27] Stuzer, M., (1996). A simple nonparametric approach to derivative security valuation. *Journal of Finance*, **51**, 5, 1633-1652.

[28] Stuzer, M., (2000). A simple entropic derivation of a generalized Black-Scholes option pricing model. *Entropy*, **2**, 70-77.

[29] Yang, J. and Qiu, W., (2005). A measure of risk and a decision-making model based on expected utility and entropy. *European Journal of Operational Research*, **164**, 792-799.