An application of general maximum entropy to utility

Paulo Ferreira*

Center for Advanced Studies in Management and Economics,
University of Évora (CEFAGE-UE),
Largo dos Colegiais, 2, 7000 Évora, Portugal
and
Agrarian Superior School of Elvas,
Edifício do Trem Alto,
Avenida 14 de Janeiro,
7350-903 Elvas, Portugal
E-mail: pjsf@uevora.pt
*Corresponding author

Andreia Dionísio

Center for Advanced Studies in Management and Economics,
University of Évora (CEFAGE-UE),
Largo dos Colegiais, 2, 7000 Évora, Portugal
E-mail: andreia@uevora.pt

Abstract: Methodologies related to information theory have been increasingly used in studies in economics and management. In this paper, we use generalised maximum entropy as an alternative to ordinary least squares in the estimation of utility functions. Generalised maximum entropy has some advantages: it does not need such restrictive assumptions and could be used with both well and ill-posed problems, for example, when we have small samples, which is the case when estimating utility functions. Using linear, logarithmic and power utility functions, we estimate those functions and confidence intervals and perform hypothesis tests. Results point to the greater accuracy of generalised maximum entropy, showing its efficiency in estimation.

Keywords: generalised maximum entropy; GME; linear function; utility; power function; logarithmic function.

Biographical notes: Paulo Ferreira is an Invited Assistant in Agrarian Superior School of Elvas. He received his PhD in Management in University of Évora, Portugal. He has published several books and papers in different refereed journals, related to different subjects: microeconomics, macroeconomics, statistics, econometrics or quality management.

Andreia Dionísio is an Assistant Professor in Universidade de Évora where she teaches courses of data analysis, decision models, research methods and marketing research. She is Vice-Director of the CEFAGE-UE and conducts

research in areas related to quantitative methods, including statistics, econometrics and econophysics. She has published several books and papers in different refereed journals.

1 Introduction

Entropy is a concept from thermodynamics and has recently been used in other research areas, particularly in economics [see, for example, Fraser (2000) or Golan et al. (1996a, 1996b, 1996c)]. In this research area, classical models often consider perfect information which in many cases requires the existence of assumptions to make estimation of functions. This is the case in traditional methods of regression estimation, such as ordinary least squares (OLS), which requires the error to be independent and normally distributed. Although methods are very useful, even for their simplicity, they can fail in circumstances where all assumptions are not verified, including the presence of perfect information (see, for example, Golan et al., 1996a).

Some new approaches to the treatment of problems in economics use methods that exploit the use of entropy. Use of the concept of entropy is growing over time, giving rise to other concepts and methodologies, such as maximum entropy (ME), cross-entropy (CE) or generalised maximum entropy (GME). The application of these methods began in the analysis of financial markets, and was later extended to other areas such as industrial economics (see, for example, Golan et al., 1998, 2000, 2001), agricultural economics (Golan et al., 1996b) or microeconomics and utility analysis – studies such Abbas (2004, 2006a, 2006b) use techniques developed from entropy, in order to solve their research problems related to utility.

This paper explores the use of GME, applying it to decision theory, by estimating utility functions. This methodology has advantages because it does not need restrictive assumptions for the variables' behaviour and it could be used both in well and ill-posed problems. For example, this methodology fits well when we have very small samples ($n < 10$), which is common in utility estimation problems. GME is also more efficient than other classic methodologies used for estimation and inference, since estimators are more precise (having lower variances) (see, for example, Golan at al., 1996a).

This paper is organised as follows. Section 2 describes the theoretical approach of the GME used in this work, identifying its potential advantages over OLS, as well as some aspects to consider in its use, notably the choice of priors. Section 3 presents several applications of this methodology in economics, and Section 4 makes the application to a specific case. Section 5 concludes the study.

2 Generalised maximum entropy

GME is a methodology that derives from the principle of ME. Used in several areas of research, such as medicine, physics, chemistry, biology, politics and others, it has had growing application in economics, primarily in the finance area.

The classic problem of ME is estimation of a probability distribution without knowing all the moments of this distribution. To solve this problem of lack of data, Jaynes (1957) proposed the principle of ME, in order to maximise Shannon's entropy,

subject to a set of restrictions that can be m moment conditions, aggregates or other restrictions. According to Jaynes (1957), in any inference problem, probabilities must be collected through this principle, using all available and relevant information.

GME is a modelling proposed by Golan et al. (1996a). The objective remains to make a prediction for the population or build an image with the available information which is often incomplete when solving problems in economics. GME has the advantage of increasing the number of possible applications in economics, especially in the presence of linear ill-posed problems (or with their linearisation).

According to Golan et al. (1996a), ill-posed problems can arise because of:

1   problems in model specification, which results in the impossibility to estimate all unknown parameters

2   mutually inconsistent data or lack of a sufficient number of points for estimation

3   collinearity problems between variables.

The problem lies in the fact that use of traditional methodologies, such as OLS, can cause problems in both the bias of the estimators and their efficiency. This can be caused by unstable solutions with high variance and lack of precision, resulting in arbitrary parameters or even open-ended solutions.

Briefly, GME generalises the concept of ME through reparametrisation of a linear model where unknown elements (parameters and errors) are in the form of probabilities. After reparametrisation, GME estimates the distribution probability of the relevant parameters and errors.

Considering a model given by $y = X\beta + e$, where y is a vector of dimension ($T \times 1$), X a matrix ($K \times T$), β a vector ($K \times 1$) and e an error vector of dimension ($T \times 1$). Only having information about the signal of parameters and about error components (β and e), such as knowing expected signals and values for the magnitude of the effects of variables, it is taken, according to Golan et al. (1996a) that each parameter β_k can be understood as a discrete random variable with a number of results M in which $2 \leq M \leq \infty$. Considering these results as a possible vector z_k of dimension ($M \times 1$), with z_{k1} and z_{kM} being respectively the minimum and maximum values, it is possible to write β_k as a combination of z_k and p_k so that:

$$\beta_k = [z_{k1} \ldots z_{km}] \begin{bmatrix} p_{k1} \\ \cdots \\ p_{kM} \end{bmatrix} \tag{1}$$

Vector z_k should be built based on sampling or experimental information or according to economic theory [see the examples of Fraser (2000) and Campbell and Hill (2001, 2005)]. Generalising for all parameters ($k = 1,\ldots,K$), we can define the vector β as follows:

$$\beta = Zp = \begin{bmatrix} z_1' & 0 & \cdots & 0 \\ 0 & z_2' & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & z_K' \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \cdots \\ p_K \end{bmatrix} \tag{2}$$

In this case Z is a matrix of support values for parameters with dimension ($K \times KM$) and $\boldsymbol{p}$ a vector of dimension ($KM \times 1$) for unknown probabilities and $p_{km} > 0$ and $p'_{kiM} = 1$ for all k.

In addition to reparametrisation of parameters, it is possible to do something similar to the errors. Assuming we have $\boldsymbol{J} \geq 2$ support points, where V is an array of support points of dimension ($T \times TJ$) and $\boldsymbol{w}$ is a vector of dimension ($TJ \times 1$) of probabilities for these points of support, we can define the following matrix:

$$e = Vw = \begin{bmatrix} v'_1 & 0 & \ldots & 0 \\ 0 & v'_2 & \ldots & 0 \\ \ldots & \ldots & \ddots & \ldots \\ 0 & 0 & \ldots & v'_T \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_T \end{bmatrix} \tag{3}$$

For the same reason, probabilities must be $w_{tj} > 0$ and $w'_{tiJ} = 1$ for all values of t.

Using the above matrix, it is possible to write the reparametrised model given by

$$\mathbf{y} = \mathbf{XZp} + \mathbf{Vw}. \tag{4}$$

In this case the values of $\mathbf{y}$, $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{V}$ are known. The objective is to estimate the unknown parameters of vectors $\mathbf{p}$ and $\mathbf{w}$ through the ME principle. The model to estimate is the following:

$$\max H(p, w) = -p' \ln p - w' \ln w$$
$$s.t.$$
$$y = ZXp + Vw \tag{5}$$
$$(i_K \otimes i'_M) \, p = i_K$$
$$(i_T \otimes i'_J) \, w = i_T$$

where $\otimes$ is Kronecker's product. The second equation is a restriction and the following equations are additivity restrictions, which require the sum of probabilities to be equal to unity, both for K parameters and for T errors.

One of the issues that can arise when implementing GME is related to the priors for parameters and error. Choice of the size of priors for parameters is related to the type of information used in the model: if less information is given to the problem, the prior's range is greater, so the result tends to approach other traditional methods of estimation, namely OLS. Choice of priors is also related to the information available for the variables, and in these circumstances the prior should be centred on the value indicated by theory. Other prior values should be equidistant from each other.

For the prior of error, most authors use the 3σ rule (Pukelsheim, 1994) which states that points must be between $v'_{t1} = -3\sigma$ and $v'_{tJ} = 3\sigma$ where σ is the standard deviation of the error. Since the value of σ is not known, it should be replaced by an estimate. Campbell and Hill (2001) indicate that this estimate may be the standard error of OLS regression or the standard deviation of the dependent variable.

The fit of estimates depends on the quality of the chosen support elements for priors, as mentioned by Mittelhammer et al. (2002). With small samples (less than 25 observations), these authors even conclude that GME results are better than other

traditional methods in terms of efficiency. The same authors also add that GME estimators are not affected by external assumptions.

3   ME and GME applications in utility

There are several applications of ME and GME in economics. One area that has aroused much interest in the recent past is the estimation of utility of economic agents through ME and GME. Use of these methods is also linked to the similarity between the theoretical problems of entropy and utility, as indicated by Candeal et al. (2001).

One possibility for analysing the utility of agents is through revealed preference techniques, watching consumers' behaviour and inferring about their preferences. But estimation of utilities continues to be a problem. Estimating the utility can be a typical problem where we do not know all the information about distribution probability for the economic agent. The identification of the utility is subjective, which can skew the results. In addition, it is a problem that can be time-consuming for analysts and agents themselves, possibly generating few observations.

When there is no uncertainty, identification of the order of hypotheses is sufficient to get utility. However, when uncertainty exists, this information is not sufficient and one possibility is to analyse the expected utilities according to the von Neumann and Morgenstern (1944) method.

Within this context, Abbas (2006b) presents an approach to obtaining the utility function when there is only partial information about preferences, through the analogy between the utility function and the concept of entropy. Abbas (2004) also uses an algorithm to identify utilities when there is information about the order of choices, that information being obtained from a questionnaire with binary response, to obtain consumer preferences more easily.

To analyse the decision of an agent under conditions of uncertainty, Abbas (2006a) proposes a method based on ME for the joint distribution probability, assuming the possibility of dependence between variables, which sometimes is not considered in solving problems using uncertainty. The same author (Abbas, 2006a) uses ME to identify utility values for a distribution probability.

The use of ME to examine utility has some advantages. According to Abbas (2006a), these advantages are: it incorporates all the information available at the time of decision making; no assumptions are introduced in a particular distribution probability; it applies both to quantitative and qualitative variables; and it is not limited to using only moments and correlation coefficients.

However, this approach still has some limitations. On the one hand, it is assumed that there are no measurement errors in the values of utility. However, this assumption may be violated if data is obtained through elicitation methods (see Pires et al., 2010).

Another problem relates to the impossibility of differentiating the utility problem obtained by ME. This leads to an additional problem: without differentiability, optimisation problems turn out to be limited in their analysis, using this method. Therefore, GME is used to estimate utility functions. Regarding the problem of measurement errors, they may continue to exist, because this paper continues to use data from utility elicitations. To investigate whether there is a way to minimise these measurement errors through application of GME is a possible area for future research. For example, Pires et al. (2010) use GME to estimate utility functions using data obtained

by utility elicitation methods and compare the performance of this methodology with OLS, confirming that the GME estimator is more precise than the OLS one.

Besides utility, GME is also used in other research areas. One of the first studies applying GME was in industrial economics. Golan et al. (1996c) use GME to estimate the market shares of 20 companies in different industries, using only information from four different concentration indicators to estimate all market shares. As an ill-posed problem, ME turns out to be the solution to the problem. Thus, entropy is maximised subject to these restrictions and to the condition that the sum of quotas is equal to unity. Golan et al. (1998) use it in game theory, comparing GME with maximum likelihood estimators and concluding that GME is more efficient. Golan et al. (2000) use GME to estimate pricing strategies and advertising in Coca-Cola and Pepsi, providing consistent and efficient estimators. Meanwhile, Fraser (2000) also uses GME to estimate the demand for meat in the UK due to the presence of an ill-posed problem from the existence of multi-collinearity problems. Golan et al. (2001) estimate the demand for five types of meat in Mexico, using GME with non-negativity constraints, with several advantages when compared with templates or two steps with maximum likelihood: firstly, this is a robust method, even when the error is not normal; in addition, this method is more efficient than the previous one (displaying a smaller mean square error), it does not use any assumptions about the structure and function of the error in both problems and behaves well in ill-posed problems.

GME has also been used in agricultural economics. Golan et al. (1996b) used this methodology to estimate a dynamic problem that may be incomplete or inadequate and have not determined a problem. The goal was to estimate the state variable and analyse its dynamics, and they used the example of the amount of fish in a given aquatic environment to show the utility of its use. A similar analysis is performed by Golan et al. (1994) and Golan and Vogel (2000). In the first case, it is intended to obtain estimates for multi-sector matrices from incomplete data. In the second case, using a matrix with complete data for a period, the aim is to estimate the dynamic matrix of flows between two distinct periods.

Directed to the best knowledge of income distribution, since the data are traditionally presented at intervals, Wu (2003) uses GME to analyse the distribution of income in the USA, finding empirical evidence for the effectiveness of this method.

4 An application in utility

This paper aims to evaluate the use of GME to estimate utility functions, from elicitation data used by Bleichrodt et al. (2001)[1]. In this paper, the authors propose a quantitative modification procedure for the elicitation of traditional utilities (such as the certainty equivalent) used in the context of risk and uncertainty. According to the authors, since traditional methods assume expected utility in the elicitation of utilities, this may lead to the same bias. This bias can be noted in some instances where, under the assumption of expected utility, there are differences in those elicitations. This new method is explored in an experiment that considers the hypothetical choices between optimal medical treatments, with the participation of 51 university students in two interviews separated by two weeks. The database consists of 48 individuals, as in three cases data was not obtainable.

In the presence of small samples (seven observations for each individual), traditional methodologies, including OLS may show estimators' properties to be different from those desired due to the fact that we can be in the presence of an ill-posed problem. To solve such problems, Golan et al. (1996a) proposed the use of GME. Therefore, we use OLS and GME with the objective of estimating various utility functions. After estimating these functions, we compare the results of both methods to see which one is the best estimation method, in terms of efficiency and accuracy of estimates, evaluating the potential advantage of using GME compared with OLS (namely due to the small number of observations, where utility is a possible example of this kind of problem). Utility functions are estimated with OLS and GME based on the elicitation data of Bleichrodt et al. (2001).

4.1   OLS and GME estimations

The choice of utility functions to estimate was made using the functions that are most studied in the literature (see, e.g., Stott, 2006) and also those that could be used because they are linear or can be linearised. So linear, logarithmic and power functions were estimated, as described in Table 1[2].

Table 1   Utility functions used

Utility function	Functional form	Linearised form
Linear	$U = \beta x$	$U = \beta x$
Power	$U = x^{\beta}$	$\ln U = \beta \ln x$
Logarithmic	$U = \beta \ln x$	$U = \beta \ln x$

The estimation was performed using two different methods: OLS and GME method. The aim is to compare the estimates of these two methods since, being in the presence of a small number of observations, GME may be a more efficient method of estimation (see, e.g., Golan et al., 1996a).

The estimation of utility functions through GME requires the definition of priors to be used in the estimation method, both for the estimated parameters and the error. Regarding the parameters (in this study, the value of β in each of the utility functions), when there is no information on the parameters, the definition of priors is traditionally made centring the vector on the 0 value, then choosing large enough limits that could not influence results.

Linear function is a risk neutrality function, whatever the value of β. In this case, economic theory does not indicate a specific value for the parameter. It only needs to be positive to make function's description possible. In this case, we proceeded with the estimation of function with OLS and then calculated the average of the estimated parameters, therefore being made centring on the value of 1. In the case of power function, the value of β indicates the degree of risk neutrality. It takes the unit value for risk neutral agents; if the value is greater than 1, the agent is risk loving; if the parameter is lower than one, the agent is risk averse. As such, the value considered as a reference for the identification of priors is $\beta = 1$ (in this case, the average parameters estimated by OLS was also equal to unity). Finally, regarding the logarithmic function, by definition this function corresponds to a profile of a risk adverse agent, so the utility function exists for any non-zero value of β. Once again we use priors centred on 1, a value which

comes from the mean of OLS estimation of β. Power and logarithmic functions have characteristics that make them widely used in the financial literature, because in the presence of risk aversion, utility grows at a decreasing rate with increasing wealth.

Since the choice of different priors may influence the results, experiments were made with seven different priors for the parameter of the utility functions, and these can be found in Table 2.

Table 2   Priors used for parameters

Priors used
$z_k^1 = \{-8, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
$z_k^2 = \{-10, -9, -8, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
$z_k^3 = \{-12, -11, -10, -9, -8, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}$
$z_k^4 = \{-8, -7.5, -7, -6.5, -6, -5.5, \dots 7.5, 8, 8.5, 9, 9.5, 10\}$
$z_k^5 = \{-6, -5.5, -5, -4.5, -4, -3.5, \dots, 5.5, 6, 6.5, 7, 7.5, 8\}$
$z_k^6 = \{-4.5, -4, -3.5, -3, -2.5, -2, -1.5, -1, 0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6\}$
$z_k^7 = \{-3, -2, -1, 0, 1, 2, 3, 4, 5\}$

For the error, priors used followed the 3σ rule (Pukelsheim, 1994), and for the value of σ both estimates of the standard deviation of the dependent variable and standard deviation from OLS were used. Values can be found in Table 3. The software used was Gauss 8.0.

Table 3   Priors used for error

Priors σ_Y	*Priors σ_{OLS}*
$v_{power} = \{-0.75, -0.375, 0, 0.375, 0.75\}$	$v_{power} = \{-0.18, -0.09, 0, 0.09, 0.18\}$
$v_{logarithmic} = \{-1.00, -0.05, 0, 0.05, 1.00\}$	$v_{logarithmic} = \{-0.15, -0.075, 0, 0.075, 0.15\}$
$v_{linear} = \{-1.00, -0.05, 0, 0.05, 1.00\}$	$v_{linear} = \{-0.25, -0.125, 0, 0.125, 0.25\}$

Firstly, we estimated those functions with OLS and GME. Estimates are not significantly different in most cases. Results of comparison of OLS and GME parameters, using both values for σ are not presented in this paper, but will be provided by the authors if requested.

To compare the accuracy of estimates, confidence intervals were constructed for parameter values, also using both approaches. We used the same priors that were previously identified. In the intervals using the value of the standard deviation of the dependent variable as an estimate for the value of σ confidence intervals with 2,000 and 4,000 replicas were calculated. For intervals that use the value of standard deviation of the OLS intervals, 2,000 and 3,000 replicas were calculated (it was not possible in this second case, to calculate confidence intervals with 4,000 replicas, due to lack of computing power for this purpose; since there were no significant differences in results when using different numbers of replicas, it is considered that the results are not significantly different).

In relation to confidence intervals, they general point to a higher amplitude for intervals calculated using OLS, indicating possibly greater accuracy of GME in obtaining estimates. In some cases, as in the calculation of confidence intervals for the power function and using the estimate of standard error of the dependent variable to the value of σ, the confidence intervals of GME always have smaller amplitude, for any of the 48 individuals in the sample and for any of the priors used.

Figure 1   Difference in 99% confidence intervals amplitude (*amplitude*$_{GME}$ – *amplitude*$_{OLS}$) for the linear function parameter, using standard error of the dependent variable in GME estimation
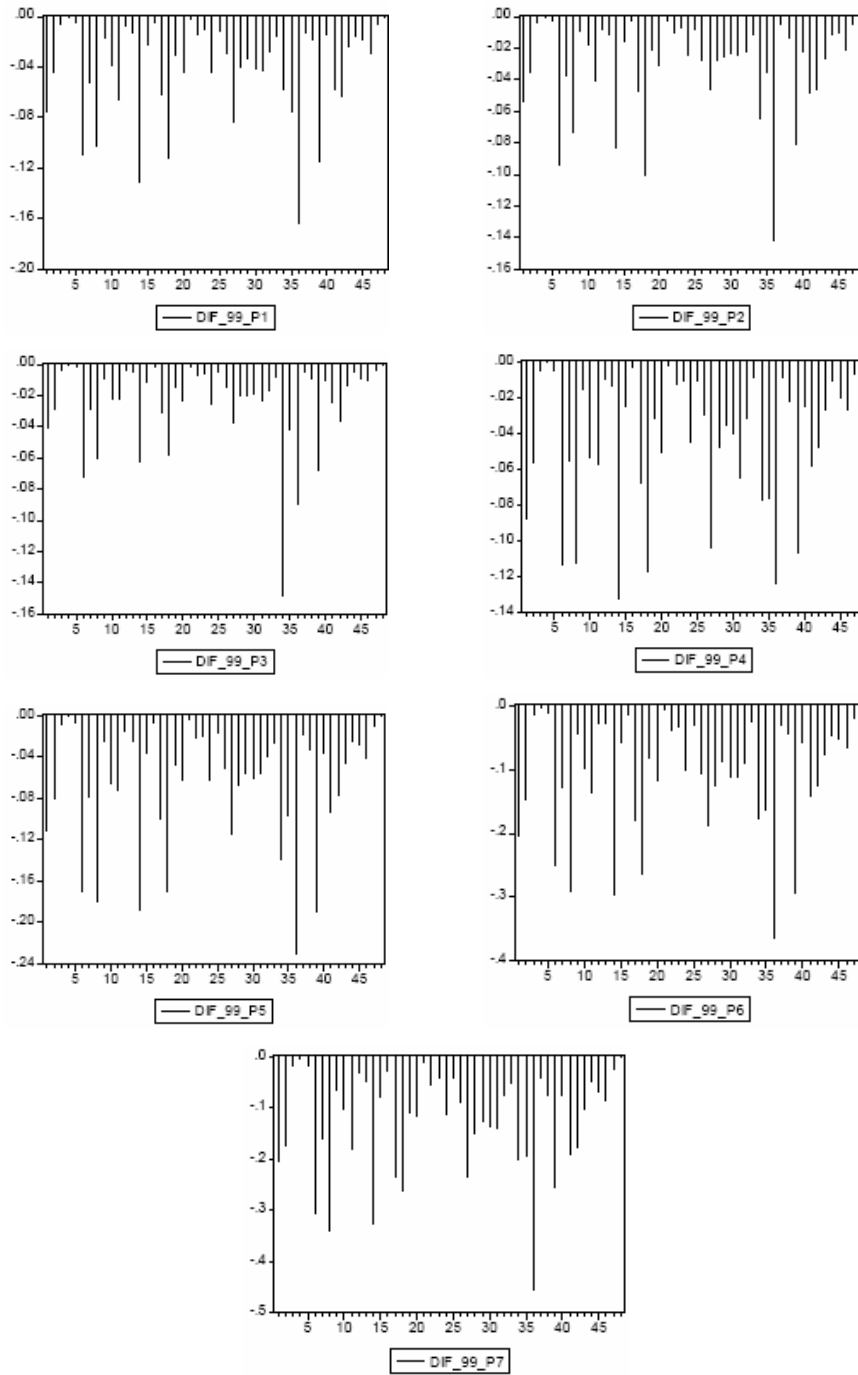
Figure 2 Difference in 99% confidence intervals amplitude (*amplitude$_{GME}$* − *amplitude$_{OLS}$*) for the power function parameter, using standard error of the dependent variable in GME estimation
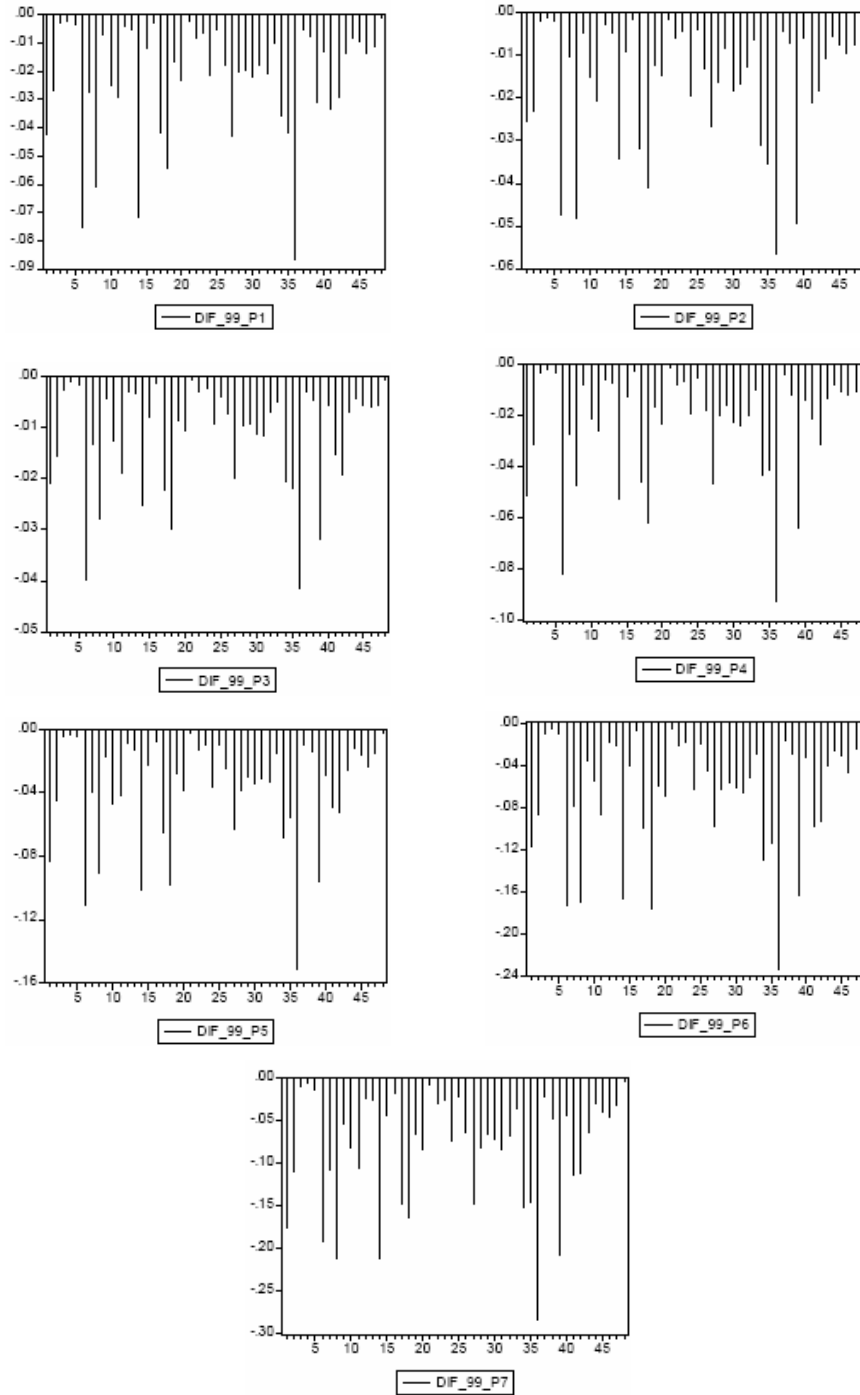
Figure 3   Difference in 99% confidence intervals amplitude ($amplitude_{GME} - amplitude_{OLS}$) for the logarithmic function parameter, using standard error of the dependent variable in GME estimation
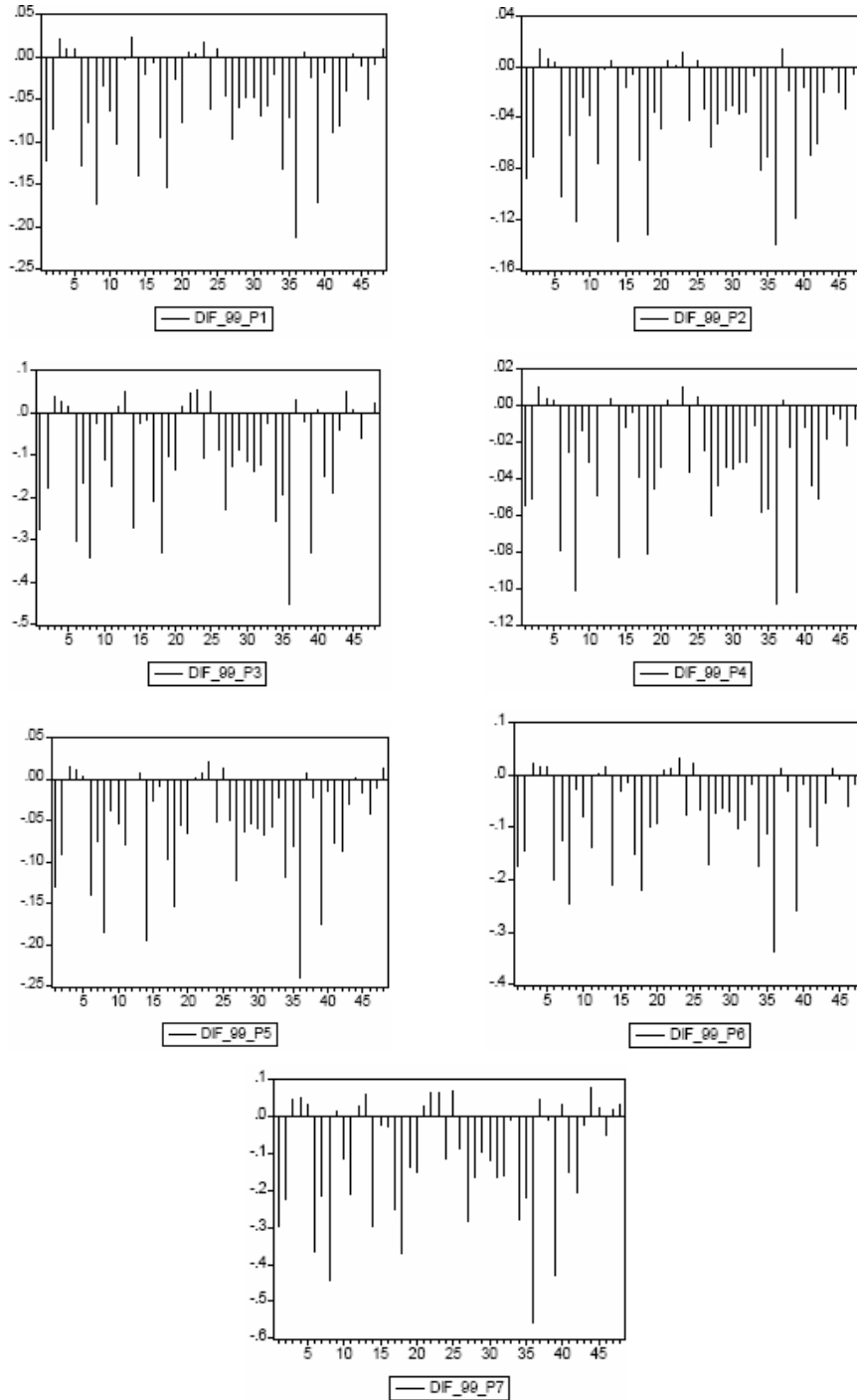
Figure 1 shows the differences in amplitudes between 99% confidence intervals in the case of linear function.

Figure 2 shows the same information for confidence intervals of the estimated parameter for the power function, while Figure 3 presents the results of the logarithmic function. In all cases, GME estimation is done using the standard deviation of the dependent variable. The results of the confidence intervals for levels of 99% using the standard deviation of OLS, as well as the confidence intervals at 95% and 90% are not shown because there were no significant differences of interpretation but again they are available on request.

The fact that GME intervals have smaller amplitude indicates that this methodology is a priori more accurate in its results and more efficient. This can also be seen by examining the standard errors resulting from the construction of confidence intervals. Because these figures have similar interpretation to those of confidence intervals, they are not used in this work but are also released on request.

We also proceeded to calculation of critical values by GME. These critical values were the basis for conducting hypothesis tests in order to evaluate the statistical significance of parameters estimated by GME. Creating samples from 3 to 9 observations, we calculated percentiles corresponding to 0.5%, 2.5%, 5%, 95%, 97.5% and 99.5%, with 50,000 replications of samples. We were unable to obtain critical values for larger samples due to lack of computing power. All samples were obtained by random values from the t-student distribution with $n - 1$ degrees of freedom. The choice of t-student instead of normal distribution is related to two aspects: first, the fact that the samples used are small, and secondly because it addresses the possible leptocurtosis distribution, allowing stronger tails and values more concentrated around the mean. For the parameter, we used the prior $\{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$ and the error prior was built according to the 3σ rule. Critical values of GME can be found in Table 4.

Table 4   Critical values for GME to test parameters individually

df	0.005	0.025	0.05	0.95	0.975	0.995
2	–4.9918	–3.7344	–2.3334	2.3697	3.8506	4.9927
3	–4.6363	–2.716	–1.6462	1.634	2.7152	4.6861
4	–4.1934	–2.1221	–1.3308	1.3037	2.1124	4.18
5	–3.8324	–1.7699	–1.1037	1.0744	1.7046	3.7116
6	–3.3737	–1.477	–0.953	0.9834	1.5213	3.5339
7	–3.1023	–1.2972	–0.8684	0.878	1.3316	3.185
8	–2.9474	–1.2009	–0.8003	0.7993	1.187	2.9689

Obtaining critical values aims to evaluate the statistical significance of estimated parameters. In the case of linear and logarithmic utility functions, we tested the significance of each estimated parameter, with hypotheses H_0: $\beta_j = 0$ and H_1: $\beta_j \neq 0$. In the case of the power function, it is intended to verify the degree of risk aversion, so hypotheses are H_0: $\beta_j = 1$ and H_1: $\beta_j \neq 1$.

The statistical test is a traditional t-statistic given by $t = \dfrac{\hat{\beta} - \beta_{H0}}{\sqrt{\text{var} \, \hat{\beta}}}$, with $\hat{\beta}$ the parameter being estimated, β_{H0} the value of the parameter under test and $\sqrt{\text{var} \, \hat{\beta}}$ the standard deviation of the estimated parameter.

In order to calculate the above equation, it is necessary to obtain the variances of the parameters. To do this, we follow the procedure of Mittelhammer and Cardell (1997) and Fraser (2000) which allows us to calculate statistical tests. We reject the null hypothesis of the test when the observed value of the statistical test exceeds the critical value corresponding to the desired level of significance.

For both linear and logarithmic functions, results are similar: we reject all null hypotheses at a significance level of 1%, so all parameters are statistically significant (the same conclusion can be observed through confidence intervals). The conclusions are the same, regardless of the prior used and the value used to estimate σ.

For the power function, where the test is around the unit value to check the agents' type of risk aversion, we also always reject the null hypothesis. However, the significance level for rejection is not always the same. In the case of using the standard deviation of the OLS to estimate the values of GME, rejection with a significance level of 10% is made on six occasions and in all other tests the hypothesis is rejected with a significance level of 5%. When using the standard deviation of the dependent variable, the number of rejections with a significance level of 10% increases to 10. What follows is that parameters are significantly different from unity, so agents are not risk neutral. We can conclude that only one agent has a risk aversion profile, all others being risk lovers.

The distribution values calculated in this research are smaller than the values of t-student distribution due to the fact that of being generated for a variable β where $E(\beta) = 0$ which is standardised. In addition, there can be some asymmetry in the distribution of β. In any case, even taking the test for the t-student distribution, we also reject the null hypothesis for all parameters of linear and logarithmic functions, so that they are statistically significant.

4.2   *Selecting the most appropriate utility function*

After calculating estimates for the different utility functions, and finding that GME estimates are more accurate, we intend to identify which of these functions is the one that best fits data. In order to compare the estimates, we used two different indicators. First we used the linear correlation coefficient in order to identify the relationship between the estimated and true observed value for each individual. The correlation coefficient shows the closeness between the estimated and observed values, so the best function will be the one that shows the highest correlation value. However, as there may be non-linear elements and because we are in the presence of small samples (which could lead to the option of estimating functions through GME instead of OLS), we decided also to use an additional criterion: the CE. CE is a divergence measure that evaluates the distance between two different distributions, which takes zero value if distributions are equal. Thus, the utility function that shows the lowest value of CE is the best.

The correlation coefficient values between the observed and estimated values are all very high. The lowest correlation value is for the logarithmic function (0.9778), with the highest value being the linear function (0.98558). Results can be seen in Table 5 (in some cases, the correlation coefficient is always the same), with no significant differences depending on the estimate of σ which is used. For each function, values did not vary significantly, depending on the priors used.

Table 5   Correlation coefficient between estimated and true value of utility

	σ_y				σ_{OLS}		
	Power	*Logarithmic*	*Linear*		*Power*	*Logarithmic*	*Linear*
OLS	0.98403	0.97779	0.98558	OLS	0.98403	0.97779	0.98558
Prior 1	0.98405	0.97795	0.98558	Prior 1	0.98404	0.97779	0.98558
Prior 2	0.98404	0.97790	0.98558	Prior 2	0.98404	0.97779	0.98558
Prior 3	0.98404	0.97787	0.98558	Prior 3	0.98404	0.97779	0.98558
Prior 4	0.98405	0.97796	0.98558	Prior 4	0.98404	0.97779	0.98558
Prior 5	0.98406	0.97806	0.98558	Prior 5	0.98404	0.97779	0.98558
Prior 6	0.98404	0.97829	0.98558	Prior 6	0.98404	0.97780	0.98558
Prior 7	0.98409	0.97846	0.98558	Prior 7	0.98404	0.97780	0.98558

For CE, the lowest values are those of the linear function, independently of the value of σ used. When we use σ from standard error of the dependent variable, CE ranges from 0.07162 (for the OLS estimates) to 0.07258 (there is a slight difference between the values, depending on the prior used in GME). When using the value of σ from the standard error of OLS, values are always about 0.072. The second best function is the power function (always with values around 0.078) and, finally, the logarithmic function (always with values higher than 0.1). Results can be found in Table 6. The linear function is the one that presents the greatest proximity between the estimated and observed values for the utility, regardless of the prior used, as this is the utility function where CE is lowest.

Table 6   Cross entropy between estimated and observed value of utility

	σ_y				σ_{OLS}		
	Power	*Logarithmic*	*Linear*		*Power*	*Logarithmic*	*Linear*
OLS	0.07812	0.10523	0.07162	OLS	0.07812	0.10523	0.07162
Prior 1	0.07835	0.10795	0.07193	Prior 1	0.07840	0.10565	0.07209
Prior 2	0.07835	0.10709	0.07184	Prior 2	0.07840	0.10565	0.07209
Prior 3	0.07829	0.10658	0.07179	Prior 3	0.07839	0.10564	0.07209
Prior 4	0.07846	0.10809	0.07195	Prior 4	0.07840	0.10566	0.07210
Prior 5	0.07866	0.10991	0.07214	Prior 5	0.07841	0.10569	0.07210
Prior 6	0.07844	0.11424	0.07258	Prior 6	0.07844	0.10576	0.07212
Prior 7	0.08257	0.12355	0.07349	Prior 7	0.07851	0.10592	0.07216

As can be seen, the values obtained by CE and correlation coefficient are in agreement in the ordering of different utility functions, and the function that best fits is the linear one, while the worst fitting function to data is the logarithmic function.

In terms of microeconomic theory applied to finance, linear utility function is the least used function due to the limitations that it presents. In addition, the most common risk profile for agents is risk-adverse. The fact that in this work, the best function is the linear one can be related to the data used.

The paper from which data is collected uses information about health states. In health problems it is also common to use another kind of utility function, namely the

quality-adjusted life years (QALY), which presents an intertemporal approach (see, e.g., Guerrero and Herrero, 2005). The fact that the utility functions associated with health can be multi-attribute, and not only related to one variable (see, e.g., Feeny et al., 2002), can also lead to these results.

5   Conclusions

In the field of economics and management use of methodologies related to information theory has increased in recent years, with the study of utility being one of the preferred areas of research. Methodologies for estimating the relationship between variables, such as GME, have been used because they have potential advantages when compared to some traditional methods, namely OLS. This advantage has to do with several factors inherent in some of the phenomena studied, including for example the complexity of financial markets, the fact of working with chronological sequences that are stationary only in asymptotic terms, the existence of multi-collinearity between variables and the fact that in some studies we have only small samples (such as the study of the utility functions of economic agents). In addition, methodologies used for estimation are models which require several assumptions, which does not happen in GME.

The objective of this paper is to compare OLS and GME as part of a study that involves the estimation of utility functions. The fact that we have small samples (seven observations per individual) may lead to results different from desirable using OLS, particularly in terms of parameter efficiency, which may ultimately have the result of less accuracy in the estimates.

Parameters are estimated for three different utility functions: power, logarithmic and linear using both methods. In the case of GME, it is necessary to define priors which are then used in the estimation of functions. Because there may be differences in the estimation of parameters depending on the existence of different priors seven different priors were used for each utility function, all centred on 1. For linear and logarithmic functions, the choice was made from OLS estimations while in the power function the choice was made considering risk neutrality as the benchmark.

For the errors, it is also necessary to identify priors. In this case we followed the 3σ Pukelsheim (1994) rule, this being the estimated value of σ obtained by two different methods: from the values of the OLS estimation and from the value of standard deviation of the dependent variable (the results are qualitatively similar, whatever the estimate used).

Confidence intervals were calculated to compare GME and OLS, with the conclusion that GME estimations are more accurate, since the majority of the confidence intervals obtained have a smaller amplitude when compared to those obtained with OLS. These findings reinforce the idea of previous studies that GME may be more accurate and efficient in the presence of small samples.

We also obtained standard deviations of the estimated parameters in order to test their statistical significance. We simulated critical values to be able to perform appropriate hypothesis tests for GME. Tests allowed rejection of the null hypothesis of no statistical significance in linear and logarithmic functions, so that all parameters are significant.

Since GME estimations have better results, we continue the analysis comparing the three utility functions studied, to identify the one that best fits to data, with the correlation coefficient (to measure the relationship between estimated and observed values) and CE

criterion (to analyse the proximity between observed and calculated values). The results of these two criteria are qualitatively similar, indicating that the best function is the linear one, followed by the power function and the logarithmic function, which can be explained by the type of data used.

Overall, it may be noted that in special situations, including samples of a small size, GME has its advantages as a method of estimation over the traditional OLS. These advantages result primarily in greater estimate precision and flexibility of the estimation method, which does not include strict restrictions on error behaviour, behaviour and distribution of variables and sample sizes.

However, it should be noted that GME may also have some limitations. One is the choice of priors to be used. Although, a priori, they do not influence the results too much, the truth is that there is some subjectivity in the selection, which can bias the results. Another limitation of GME relates to its difficulty in terms of calculation. Of course, not being a method of exceptional difficulty, it requires greater computing effort compared to OLS.

Acknowledgements

References

Abbas, A. (2004) 'Entropy methods for adaptative utility elicitation', *IEEE Transactions on Systems Man and Cybernetics – Part A: Systems and Humans*, Vol. 34, No. 2, pp.169–178.

Abbas, A. (2006a) 'Entropy methods for joint distributions in decisions analysis', *IEEE Transactions on Engineering Management*, Vol. 53, No. 1, pp.146–159.

Abbas, A. (2006b) 'Maximum entropy utility', *Operations Research*, Vol. 54, No. 2, pp.277–290.

Bleichrodt, H., Pinto, J. and Wakker, P. (2001) 'Making descriptive use of prospect theory to improve prescriptive applications of expected utility', *Management Science*, Vol. 47, No. 11, pp.1498–1514.

Campbell, R. and Hill, C. (2001) *Maximum Entropy Estimation in Economic Models with Linear Inequality Restrictions*, Departmental Working Papers No. 2001-11, Department of Economics, Louisiana State University.

Campbell, R. and Hill, C. (2005) 'A Monte Carlo study of the effect of design characteristics on the inequality restricted maximum entropy estimator', *Review of Applied Economics*, Vol. 1, No. 1, pp.53–84.

Candeal, J., De Miguel, J., Induráin, E. and Mehta, G. (2001) 'Utility and entropy', *Economic Theory*, Vol. 17, No. 1, pp.233–238.

Feeny, D., Furlong, W., Torrance, G., Goldsmith, C., Zhu, Z., Depaw, S., Denton, M. and Boyle, M. (2002) 'Multiattribute and single-attribute utility functions for the health utilities index mark 3 system', *Medical Care*, Vol. 40, No. 2, pp.113–128.

Fraser, I. (2000) 'An application of maximum entropy estimation: the demand of meat in the United Kingdom', *Applied Economics*, Vol. 32, No. 1, pp.45–59.

Golan, A. and Vogel, S. (2000) 'Estimation of non-stationary social accounting matrix coefficients with supply-side information', *Economic Systems Research*, Vol. 12, No. 4, pp.447–471.

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

Golan, A., Karp, L. and Perloff, J. (1996b) 'A maximum entropy approach to estimation and inference in dynamic models or counting fish in the sea using maximum entropy', *Journal of Economic Dynamics and Control*, Vol. 20, No. 4, pp.559–582.

Golan, A., Judge, G. and Perloff, J. (1996c) 'Estimating the size distribution of firms using government summary statistics', *The Journal of Industrial Economics*, Vol. 44, No. 1, pp.69–80.

Golan, A., Judge, G. and Robinson, S. (1994) 'Recovering information from incomplete or partial multisectoral economic data', *Review of Economics & Statistics*, Vol. 76, No. 3, pp.541–549.

Golan, A., Karp, L. and Perloff, J. (1998) *Estimating a Mixed Strategy: United and American Airlines*, Institute for Research on Labor and Employment, Working Paper Series, No. 1018.

Golan, A., Karp, L. and Perloff, J. (2000) 'Estimating Coke and Pepsi's price and advertising strategies', *Journal of Business & Economic Statistics*, Vol. 18, No. 4, pp.398–409.

Golan, A., Perloff, J. and Shen, E. (2001) 'Estimating a demand system with nonnegativity constraints Mexican meat demand', *Review of Economics and Statistics*, Vol. 83, No. 3, pp.541–550.

Guerrero, A. and Herrero, C. (2005) 'A semi-separable utility function for health profiles', *Journal of Health Economics*, Vol. 24, No. 1, pp.33–54.

Jaynes, E. (1957) 'Information theory and statistical mechanics', *The Physical Review*, Vol. 106, No. 4, pp.620–630.

Mittelhammer, R. and Cardell, N. (1997) *On the Consistency and Asymptotic Normality of Data-constrained GME Estimator in the General Linear Model*, mimeo, University of Washington, Washington.

Mittelhammer, R.N., Cardell, N. and Marsh, T. (2002) *The Data-constrained GME Estimator of the GLM: Asymptotic Theory and Inference*, mimeo, University of Washington, Washington.

Pires, C., Dionísio, A. and Coelho, L. (2010) *GME versus OLS – Which is the Best to Estimate Utility Functions?*, CEFAGE-UE Working Paper 2010/02.

Pukelsheim, F. (1994) 'The three sigma rule', *American Statistician*, Vol. 48, No. 2, pp.88–91.

Stott, H. (2006) 'Cumulative prospect theory's functional menagerie', *Journal of Risk and Uncertainty*, Vol. 32, No. 2, pp.101–130.

von Neumann, J. and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton University Press, Princeton.

Wu, X. (2003) 'Calculation of maximum entropy densities with application to income distribution', *Journal of Econometrics*, Vol. 115, No. 2, pp.347–354.

Notes

1   Data is available at http://people.few.eur.nl/wakker/data/data2001.01corrsg.htm.

2   In addition to the functions used in this paper, others are used with some frequency. This is the case, for example, of exponential function, given by $U = -e^{-\beta X}$. Failure to use this function in this paper is due to the fact that it could not be directly linearised, so it is not possible to apply these methodologies here.