Estimating utility functions using generalized maximum entropy

Cesaltina Pires [a] , Andreia Dionísio [a] & Luís Coelho [a]

[a] Cefage-UE, Universidade de Evora , Largo dos Colegiais 2, 7000 ,
Évora , Portugal
Published online: 09 Nov 2012.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

Estimating utility functions using generalized maximum entropy

Cesaltina Pires, Andreia Dionísio* and Luís Coelho

Cefage-UE, Universidade de Evora, Largo dos Colegiais 2, 7000 Évora, Portugal

This paper estimates von Neumann and Morgenstern utility functions using the generalized maximum entropy (GME), applied to data obtained by utility elicitation methods. Given the statistical advantages of this approach, we provide a comparison of the performance of the GME estimator with ordinary least square (OLS) 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 are generated by utility elicitation methods.

Keywords: generalized maximum entropy; maximum entropy principle; von Neumann and Morgenstern utility; utility elicitation

JEL Classifications: C13; C14; C49; D81

1. Introduction

One important step in decision analysis is the identification of the decision-maker's preferences. Thus, the estimation of the decision-maker utility function is a prerequisite for the analysis of the given decision problem. 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 these methods is likely to lead to relatively few observations since the decision-maker might not be willing to answer to many questions.[1] The fact that the number of observations generated by utility elicitation methods is small, raises interesting issues regarding the estimation of the decision-maker utility function.

With very small samples, the use of traditional estimation techniques may not be adequate. With small samples, statistical inference under ordinary least squares (OLS) is based on the assumption that the errors are normally distributed. However, if this assumption does not hold,

---

*Corresponding author. Email: andreia@uevora.pt

traditional OLS inference results will be incorrect. On the other hand, generalized maximum entropy (GME) has the ability to estimate the parameters of a regressions without imposing any constraints on the probability distribution of the errors and it is robust even when we have ill-posed problems, namely with very small samples [4,5,10]. Moreover, using Monte Carlo simulations, previous studies have shown that with small samples the generalized maximum entropy estimator outperforms OLS [20]. Considering that GME is well suited to estimate regressions with a small number of observations without imposing constraints on the errors probability distribution, a natural question arises: is GME a good alternative to estimate utility functions when data are obtained through utility elicitation methods?

In this paper, we explore the potentialities of the GME estimation of von Neumann and Morgenstern utility functions using only partial information about the agent's preferences. Since the elicitation of preferences is not a goal of this research, we use data from Abdellaoui *et al.* [2]. 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 GME estimation to these data and compare it with OLS estimation. In addition, we use bootstrap to obtain confidence intervals and compare the efficiency of the GME and the OLS estimators.

This paper can be related with two areas of Economics research. On the one hand, there exists some literature applying the maximum entropy (ME) principle to the estimation of utility functions [1,15]. This literature does not impose a functional form for the utility function, but has the drawback of generating non-differentiable utility functions which is problematic if one is interested in using the estimated utility to solve decision problems, such as portfolio optimization. Our approach is different as we apply GME to estimate a particular functional form of the utility function. This approach may be considered semi-parametric as we assume a particular functional form for the utility function but do not impose restrictions of the probability distribution of the error terms. Thus, one important contribution of our work is that it is the first one to apply GME in the estimation of utility functions.

The paper can also be related to the literature on GME and, in particular, to its applications in Economics. The ability of GME to estimate economic relationships has already been explored by several authors (see, for instance, [12,18,19]). GME applications in Economics cover a wide set of fields, including finance, game theory and agricultural economics (see, for example, [3,8,14,24, 27,28]). 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.* [11] use GME and generalized cross-entropy to derive a Social Accounting Matrix without having to impose the standard identification procedures. Paris and Howitt [22] employ GME in a mathematical programming framework to estimate cost parameters for individual farm-level data. Fraser [8] uses GME to estimate the demand for meat in the UK with data that are subject to collinearity. Golan *et al.* [11,13] use GME to estimate a censored regression model. More recently, Campbell and Carter Hill [4,5] demonstrate how to impose inequality constraints in GME through the parameter support matrix. Although there exist many applications of GME in Economics, our application has one distinctive feature: we use GME to estimate utility functions using a very small number of observations. Thus, an interesting contribution of our study is that it provides evidence on the comparison of GME with least-squares estimation using real data small samples.

This paper is organized as follows: Section 2 presents the concepts of entropy, ME principle and GME estimation and reviews recent applications of entropy and ME to decision theory. The main results are presented in Section 3, where we compare the results of GME and OLS estimation. The last section of the paper contains the main conclusions of this research and provides directions for future research.

2. Entropy, ME principle and GME

2.1 *Background theory*

In order to understand the concepts underlying GME estimation, it is worth explaining the meaning of entropy and the ME principle, since these two concepts are important building blocks of GME. The concept of entropy was introduced by Shannon [25] 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 $(p_1, p_2, \ldots, p_n)$ be the associated probability distribution. The entropy of the probability distribution $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ 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 all outcomes are equally likely, $p_i = 1/n$ 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 ME principle proposed by Jaynes [17] 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 known 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 and adding constraints) on the random variable X. According to several authors [9,26], the ME 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 GME proposed and employed by Judge and Golan [18] increases substantially the number of possible extensions and applications in economics, particularly when we have ill-posed problems in linear models. Let us consider the following standard regression model in matrix form:

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

where $\mathbf{y}$ is a $(T \times 1)$ vector, $\mathbf{X}$ is a $(T \times K)$ matrix, $\boldsymbol{\beta}$ is a $(K \times 1)$ vector and $\mathbf{e}$ is a $(T \times 1)$ vector of errors. According to Golan *et al.* [10] 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 $\boldsymbol{\beta}$ parameter vector by traditional estimation methods. An 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 are mutually inconsistent and (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 ME approach by estimating the probability distribution of each of the K parameters and estimating the probability distribution of each of the T errors. Judge and Golan [18] 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

$$\beta_k = [z_{k1} \cdots z_{kM}] \begin{bmatrix} p_{k1} \\ \vdots \\ p_{kM} \end{bmatrix}. \tag{2}$$

The parameter support is based on prior information of economic theory or in previous experiments [4,5,8].

Generalizing for all k (with $k = 1, \ldots, K$), the vector $\boldsymbol{\beta}$ can be defined by

$$\boldsymbol{\beta} = \mathbf{Z}\mathbf{p} = \begin{bmatrix} \mathbf{z}_1' & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_2' & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{z}_K' \end{bmatrix} \cdot \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_K \end{bmatrix}, \tag{3}$$

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

$$\mathbf{e} = \mathbf{V}\mathbf{w} = \begin{bmatrix} \mathbf{v}_1' & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_2' & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{v}_T' \end{bmatrix} \cdot \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_T \end{bmatrix}. \tag{4}$$

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:

$$\mathbf{y} = \mathbf{XZp} + \mathbf{Vw}, \tag{5}$$

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 ME approach. The regression model can be expressed as follows:

$$\max \quad H(\mathbf{p}, \mathbf{w}) = -\mathbf{p}' \ln \mathbf{p} - \mathbf{w}' \ln \mathbf{w}, \tag{6}$$

$$\text{s.t.} \quad \mathbf{y} = \mathbf{XZp} + \mathbf{Vw}, \tag{7}$$

$$(\mathbf{I}_K \otimes \mathbf{i}_M')\mathbf{p} = \mathbf{i}_K, \tag{8}$$

$$(\mathbf{I}_T \otimes \mathbf{i}_J')\mathbf{w} = \mathbf{i}_T, \tag{9}$$

where $\otimes$ is the Kronecker product. Equation (7) is a data constraint and Equations (8) and (9) 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 [10] is

$$\hat{p}_{km} = \frac{\exp(z_{km}\mathbf{x}_k'\hat{\boldsymbol{\lambda}})}{\sum_{m=1}^{M} \exp(z_{km}\mathbf{x}_k'\hat{\boldsymbol{\lambda}})} \quad m = 1, \ldots, M \text{ and } k = 1, \ldots, K \tag{10}$$

and

$$\hat{w}_{tj} = \frac{\exp(v_{tj}\hat{\boldsymbol{\lambda}}_t)}{\sum_{j=1}^{J} \exp(v_{tj}\hat{\boldsymbol{\lambda}}_t)}, \quad j = 1, \ldots, J \text{ and } t = 1, \ldots, T, \tag{11}$$

where $\hat{\boldsymbol{\lambda}}$ 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, it is usual to use a symmetric representation centred on zero. Pukelsheim [23] 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 Carter Hill [5] 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 [8] alerts to the problem of choosing the number of elements in the supports (M and J). Mittelhammer *et al.* [20] 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.* [20], 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 and ME in decision analysis*

In this section we briefly review recent applications of the entropy and ME principle in decision analysis with particular emphasis on the estimation of utility functions [1,6,7]. In addition, we explain how our work differs from the papers that use ME to estimate utility functions.

As described in the previous subsection, the ME principle applies to estimation of probability distributions. However, the principle has also been applied in the estimation of utility functions describing ordinal preferences by Herfert and La Mura [15] and in the estimation of von Neumann and Morgenstern utility functions by Abbas [1]. Herfert and La Mura [15] scale utility so as to behave as a probability measure, whereas Abbas [1] 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 [1] 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 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 ME utility, as Abbas calls it, is simply the function composed of the linear segments joining the known utility values.

The ME 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 are 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 the 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. To the best of our knowledge our work is the first to apply GME to the estimation of utility functions.

3. Estimation and results

In this section, we describe the data used in this paper, as well as the methodology applied to estimate the parameters of interest 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 coefficients of interest.

3.1 *Estimation of the parameters*

Our data set is taken from Abdellaoui *et al.* [2]. These authors estimate the parameters of three utility functions (power, exponential and expo-power) with only one parameter. Abdellaoui *et al.* [2] elicit the preferences of decision-makers using a choice-based method (trade-off method) and a choiceless method (the 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

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

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.

In order to estimate the parameter r, we logarithmize both sides of the equation and use the following econometric model (e.g. assuming that $r > 0$):

$$\ln y = r \ln x + e,$$
$$Y = rX + e,$$

(12)

where $Y = \ln y$ and $X = \ln x$. The model (12) will be estimated using OLS and GME.[2]

In our database, we have six observations for each participant, which is clearly a very small sample to estimate the participant's utility function. This fact may be problematic 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 errors are normally distributed (with large samples no such assumption is needed). However, if data are 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 good alternative to OLS. In this paper, the main reason to use GME is precisely the fact that our samples are very small.[3] To use GME, we do not need to assume any theoretical probability distribution for the errors even to make statistical inference.

In a first phase, we estimate the utility function parameter using the least-squares approach. We found that the parameter r for all the participants is positive, 2.6429 being the highest and 0.5925 being the smallest. For all the regression models estimated, we tested the

Table 1. Support vectors for GME used in the different experiments performed.

	Center	Width		M	
Experience A	1	$[-1, \ldots, 3]$	9	17	33
Experience B	1	$[-3, \ldots, 5]$	9	17	33
Experience C	1	$[-7, \ldots, 9]$	9	17	33
Experience D	1.5	$[-0.5, \ldots, 3.5]$	9	17	33
Experience E	1.5	$[-2.5, \ldots, 5.5]$	9	17	33
Experience F	1.5	$[-6.5, \ldots, 9.5]$	9	17	33
Experience G	0.5	$[-1.5, \ldots, 2.5]$	9	17	33
Experience H	0.5	$[-3.5, \ldots, 4.5]$	9	17	33
Experience I	0.5	$[-7.5, \ldots, 8.5]$	9	17	33

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.[4] This suggests that, in this particular case, OLS is adequate and thus it seems sensible to use GME and compare it with OLS.

Considering the power utility function, an agent is risk neutral when $r = 1$, risk averse when $r > 1$ and risk lover when $r < 1$. Since we do not know *a priori* the agent's attitude toward risk, the value $r = 1$ is a natural candidate to be the center of the support vector of the GME estimator. Therefore, our first estimations use a support vector centered in 1. However, it is interesting to analyze whether the GME results are sensitive to changes in the support center. Thus, we also compute the GME estimators with support vectors centered in 1.5 and centered in 0.5. In other words, we consider three alternative types of support vectors: (i) centered in 1; (ii) centered in 1.5 and (iii) centered in 0.5. The first scenario is based on economic theory, the last two alternatives are used to test the robustness of the GME estimator. In addition, we are also interested in testing the sensitivity of our results to changes in the width of the support vector (we run experiments with width equal to 4, 8 and 16) and to changes in the number of support points (we run experiments with $M = 9, M = 17$ and $M = 33$). Thus, for each support vector center, we consider 9 different experiments, which leads to a total of 27 experiments.[5] Table 1 summarizes the information about the experiments performed in this paper.

To define the error support vector, we followed Campbell and Carter Hill [5]. We used the standard deviation of the dependent variable (Y), which equal to 0.345, the $\pm 3\sigma$ rule of Golan *et al.* [10] and Campbell and Carter Hill [5] results on the error support of $\mathbf{v}'_i = [-1, -0.5, 0, 0.5, 1]$, with $J = 5$. In order to undertake inference procedures, we follow Mittelhammer *et al.* [20] considerations about the derivation of the standard errors for all the parameters estimated.

All the participants show a positive and statistically significant r in both estimation methods.[6] Figure 1 compares GME and OLS estimators for each of the 47 regressions, for the support vectors centered in 1. 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 becomes smaller, independent of the number of points in each support vector. 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 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.

As mentioned above we computed the GME estimator using support vectors centered in 1, 1.5 and 0.5. The main goal of these experiments is to evaluate the impact on the GME estimators of a
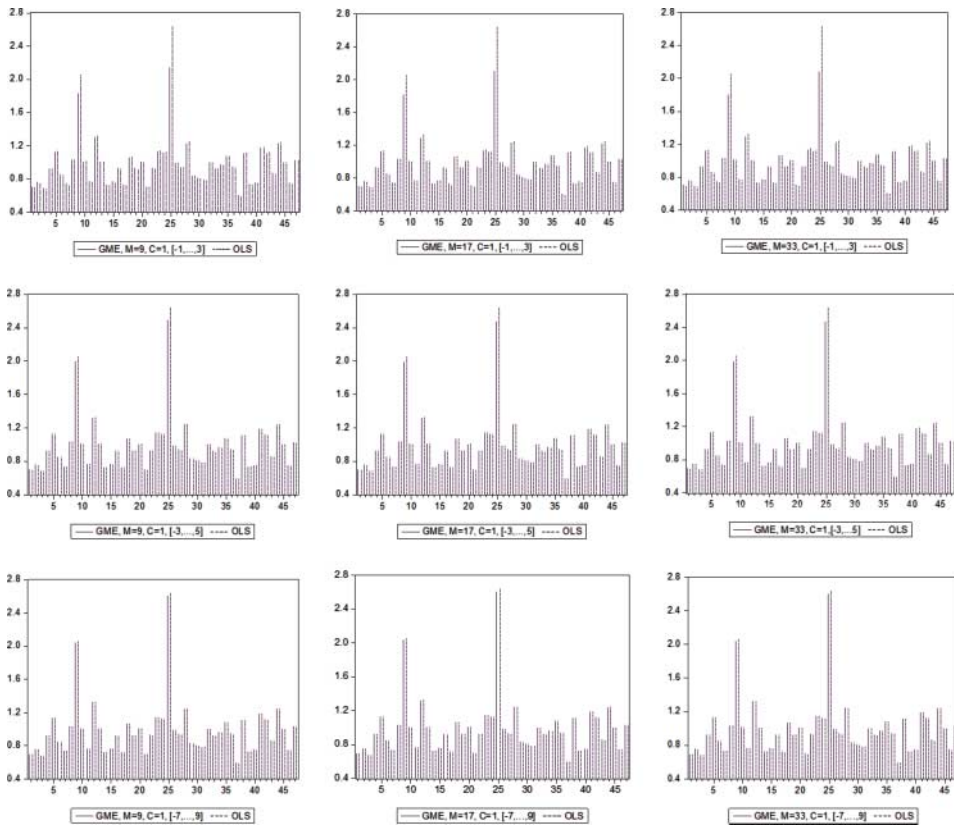
Figure 1. OLS estimator versus GME estimator for all the 47 participants for the support vectors centered in 1.

possibly biased support vector. In fact, these experiments include a 50% positive (and negative) bias with respect to the support vector based on theory and centered in 1. If the results of the GME estimator change a lot when the support vector center varies, then the GME estimator would not be robust.

The comparison of the results for the different experiments shows that the GME estimator did not change in a significant way when we move the center of the support vector from 1 to 1.5 or to 0.5. We observe a slight increase in the values of the parameter when the center is 1.5, especially for the smallest width of the support vector. However, this increase is almost nil for the support vectors with larger width. Similarly, there is a slight decrease in the parameter estimates when the support vector is centered in 0.5, but this decrease almost vanishes when we consider wider support vectors. In fact, for wider support vectors (which means a less informative prior) the results are almost insensitive to the center of the support vector and are very close to the OLS results (Figure 2). Once more, the number of points in the support vector does not influence in a significant way the results.

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 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.

The strong 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 the errors.
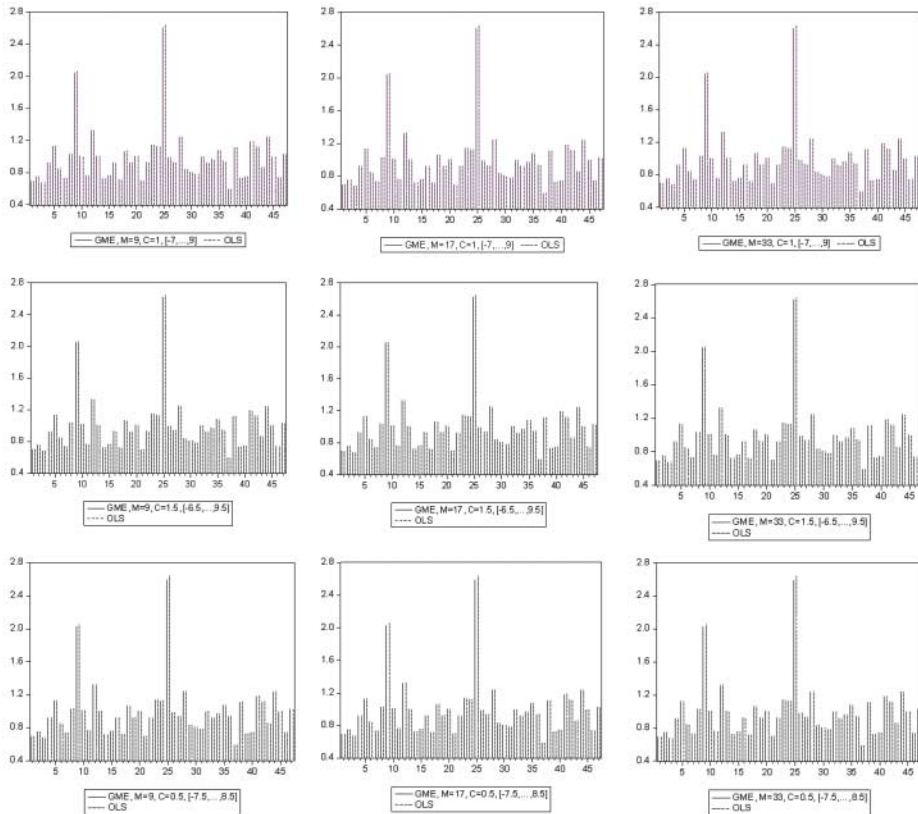
Figure 2. OLS estimator versus GME estimator for all the 47 participants for the support vectors with width = 16 centered in 1, 1.5 and 0.5.

The standard errors of the GME estimator were estimated using the approach of Mittelhammer *et al*. [20]. Figure 3 shows the obtained values for the standard errors for the $C = 1$ and $M = 9$ experiments and OLS. There are not significant differences in the experiments with different number of element in the support vector. The results point to a slight increase in the standard errors as we increase the width of the support vector. This pattern also holds in the experiments centered in 1.5 and 0.5, which do not differ in terms of interpretation.

In order to compare the standard errors of the GME estimator and the standard errors of the OLS estimator, we tested the hypothesis of equal mean of the standard errors using the *t*-Students test and the Wilconxon test. The null of equality of the means and medians of the standard errors ($H_0 : \mu_{\sigma_{\hat{\beta} \text{GME}}} = \mu_{\sigma_{\hat{\beta} \text{OLS}}}$) is not rejected in any experiment, which reveals the small difference between the efficiency of the estimators under consideration.

The next section 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 [16] presents a bootstrap
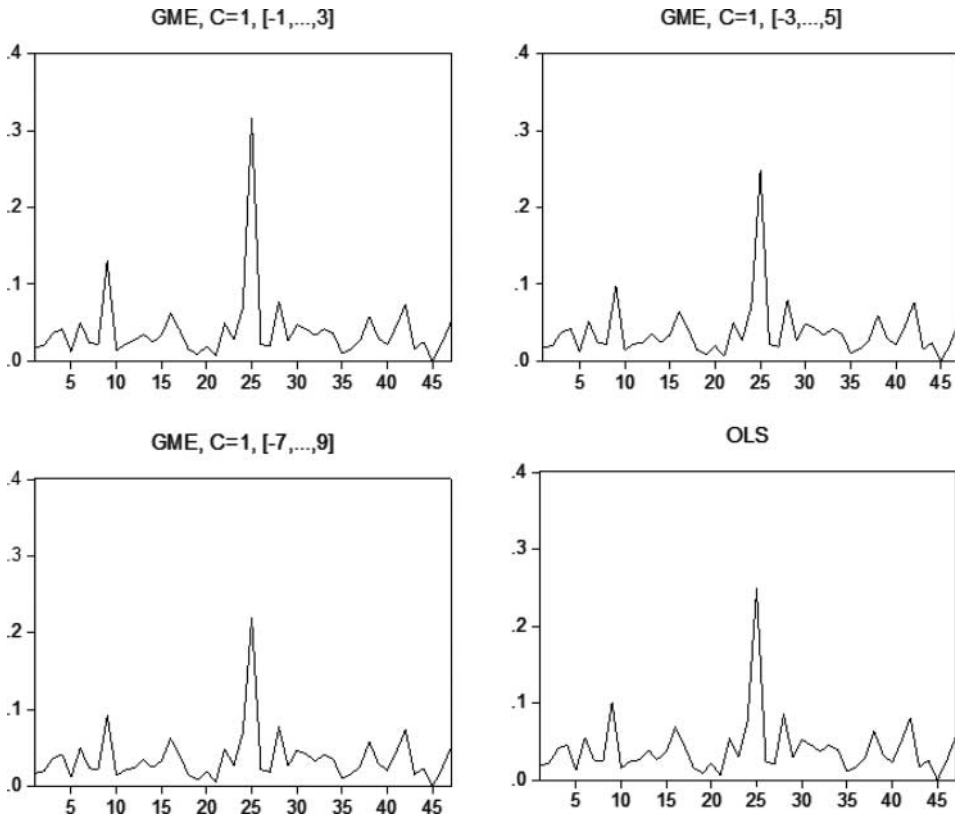
Figure 3. Standard errors of the GME parameter of interest (r) for the experiments centered in 1 and OLS.

method for computing confidence intervals where the t-statistics is obtained from the resampled data and the interval estimates are computed as $\hat{r} \pm t_c S_{\hat{r}}$, where t_c is the bootstrap t-statistics and $S_{\hat{r}}$ 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 [21] and by Campbell and Carter Hill [5], 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) the t-statistic would not be correct if the errors are not normally distributed and (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 0.05th, 2.5th, 5th, 95th, 97.5th and 99.5th percentiles and construct the confidence intervals with 90%, 95% and 99% confidence level.

In order to simplify the exposition, and since there are no substantial differences between the several experiments performed, we will only exhibit the results for the experiments A, B and C (all of them centered in 1) using $M = 9$ points.

As we can see from Figure 4, 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 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
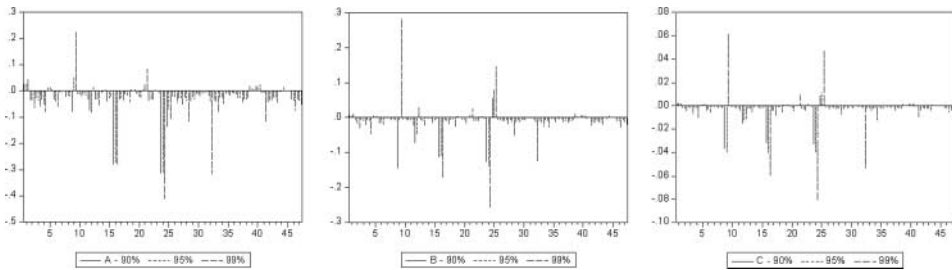
Figure 4. Difference of width between the confidence intervals of the GME and the OLS estimates (calculated by: Width$_{GME}$ − Width$_{OLS}$). Support vectors centered in 1, A – Support vector $[-1, \ldots, 3]$, B – Support vector $[-3, \ldots, 5]$ and C – Support vector $[-7, \ldots, 9]$.

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 the 1% level of significance, confirming the results obtained in the previous subsection.

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 such small samples the use of traditional estimation and inference techniques may not be adequate. Moreover, previous studies have shown that with small samples the GME estimator outperforms OLS. This suggests that GME might be well suited to estimate utility function when data are 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. We used data from Abdellaoui *et al*. [2], 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. Thus, our paper complements previous results, that used Monte Carlo simulations, 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 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.

Our results suggest that GME is an interesting alternative to OLS in the estimation of utility functions using data obtained through utility elicitation methods. Mittelhammer *et al*. [20] show that GME outperforms OLS in small samples in terms of the mean square error. Our results show

that the GME estimators have lower variance, which is consistent with GME having a lower mean square error. It should be mentioned that the advantage of using GME is higher if one has good prior information. In the case of the utility functions estimation, one can use utility theory to choose the center of the support vector of the risk aversion parameter, giving additional support to the use of GME. Another advantage of GME is that it does not require assumptions about the theoretical distributions of the errors.

One important contribution of our work is that it is the first one to apply GME in the estimation of utility functions. Previous literature applied the ME principle to the estimation of utility functions, without imposing a functional form for the utility function. In this study we applied GME to estimate a particular functional form of the utility function. This approach has the advantage of estimating a differentiable utility function, which may be an important feature in solving decision problems.

One possible direction for future work would be to take into account the fact that the utilities are normalized between zero and one. In other words, it would be interesting to use GME to estimate a fractional regression model. Another possible extension would be to explore the choice of the utility functional form using GME.

Acknowledgements

Notes

1. The use of utility elicitation methods presents other difficulties. First, different methods of utility elicitation yield different results. Thus, one needs to be careful in designing the elicitation method. Second, the methods of elicitation and validation of results are sometimes difficult to implement.
2. 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.
3. It is important to mention that the existence of small samples is not the only circumstance when GME can be used with econometric and statistical advantages. See, for instance, Fraser [8] that has large samples but where the existence of multicollinearity justifies the use of GME. On the other hand, Mittelhammer *et al*. [20] show that the quality of the GME estimates often perform better than OLS, in several Monte Carlo experiments for small and large samples.
4. It should be noted that, with such small number of observations, the power of the normality tests is small.
5. Note that we use the same support vector for all the individuals.
6. Considering the large number of experiments and regressions we present most of our results graphically. To provide an example, the appendix presents a table with the results of 3 of our 27 experiments. The remaining results are not included in the paper; however, they are available from the authors upon request.

References

[1] A. Abbas, *Maximum entropy utility*, Oper. Res. 54 (2006), pp. 277–290.
[2] M. Abdellaoui, C. Barrions, and P. Wakker, *Reconciling introspective utility with revealed preference: Experimental arguments based on prospect theory*, J. Econometrics 138 (2007), pp. 356–378.
[3] P.W. Buchen and M. Kelly, *The maximum entropy distribution of an asset inferred from option prices*, J. Financ. Quant. Anal. 31(1) (1996), pp. 143–159.
[4] R. Campbell and R. Carter Hill, *A Monte Carlo study of the effect of design characteristics on the inequality restricted maximum entropy estimator*, Rev. Appl. Econ. 1 (2005), pp. 53–84.
[5] R. Campbell and R. Carter Hill, *Imposing parameter inequality restrictions using the principle of maximum entropy*, J. Stat. Comput. Simul. 76 (2006), pp. 985–1000.

[6] A. Darooneh, *Utility function from maximum entropy principle*, Entropy 8 (2006), pp. 18–24.

[7] A. Dionisio, A.H. Reis, and L. Coelho, *Utility function estimation: The entropy approach*, Phys. A 387 (2008), pp. 3862–3867.

[8] I. Fraser, *An application of maximum entropy estimation: The demand for meat in the United Kingdom*, Appl. Econ. 32 (2000), pp. 45–59.

[9] A. Golan, *Information and entropy econometrics-editor's view*, J. Econometrics 107 (2002), pp. 1–15.

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

[11] A. Golan, G. Judge, and J. Perlof, *Estimation and inference with censored and ordered multinomial response data*, J. Econometrics 79 (1997), pp. 25–51.

[12] A. Golan and J. Perloff, *Comparison of maximum entropy and higher-order entropy estimators*, J. Econometrics 107(1–2) (2002), pp. 195–211.

[13] A. Golan, J. Perloff, and E. Shen, *Estimating a demand system with nonnegativity constraints: Mexican meat demand*, Rev. Econom. Statist. 83(3) (2001), pp. 541–550.

[14] B. Gulko, *The entropy pricing theory – market beliefs, valuation, incomplete markets*, Asset pricing, Ph.D. dissertation, UMI – Dissertations Services, 1998.

[15] M. Herfert and P. La Mura, *Estimation of consumers preferences via ordinal decision-theoretic entropy*, Working-Paper 64, Leipzig Graduate School of Management, 2004.

[16] J.L. Horowitz, *Bootstrap methods in econometrics: Theory and numerical performance*, in *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress*, Vol. III, D.M. Krpes and K.F. Wallis, eds., University Press, Cambridge, 1997, pp. 188–222.

[17] E.T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev. 106 (1957), pp. 620–630.

[18] G.G. Judge and A. Golan, *Recovering Information in the Case of Ill-Posed Inverse Problems with Noise*, Mimeo Department of Agricultural and Natural Resources, University of California, Berkeley, CA, 1992.

[19] Y. Kitamura and M. Stutzer, *An information-theoretic alternative to generalized method of moment estimation*, Econometrica 65 (1997), pp. 861–874.

[20] R. Mittelhammer, N. Cardell, and T. Marsh, *The Data-Constrained GME Estimator of the GLM: Asymptotic Theory and Inference*, Mimeo, University of Washington, 2002. Available at http://ideas.repec.org/f/pmi493.html.

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

[22] Q. Paris and R. Howitt, *An analysis of ill-posed production problems using maximum entropy*, Amer. J. Agric. Econ 80 (1998), pp. 128–138.

[23] F. Pukelsheim, *The three sigma rule*, Amer. Statist. 48 (1994), pp. 88–91.

[24] D. Samperi, *Entropy and statistical model selection for asset pricing and risk management*, preprint (1999). Available at http://papers.ssrn.com.

[25] C. Shannon, *A mathematical theory of communication*, Bell Syst. Technol. 27 (1948), pp. 379–423, 623–656.

[26] E. Soofi, *Principal information theoretic approaches*, J. Amer. Statist. Assoc. 95 (2000), pp. 1349–1353.

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

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

Appendix

This appendix lists some results obtained in this paper. Since the total amount of results is considerable and thus difficult to present completely in the paper, we provide the results for three representative experiments: A, D and G for $M = 9$. These results refer to the experiments with the smallest widths. All remaining results are available from the authors upon request (Table A1).

Table A1. Results for the r estimator, standard-deviation and t-Student statistics for the experiments A, D and G with $M = 9$.

	OLS			A			D			G		
Ind	r	SD	t	r	SD	t	r	SD	t	r	SD	t
1	0.6956	0.0214	32.5449	0.7041	0.0157	44.7849	0.7196	0.0180	40.0330	0.6902	0.0154	44.8172
2	0.7532	0.0243	31.0318	0.7612	0.0186	40.9075	0.7790	0.0209	37.2011	0.7450	0.0185	40.3422
3	0.6817	0.0321	21.2312	0.6904	0.0313	22.0585	0.7054	0.0326	21.6223	0.6771	0.0309	21.9127
4	0.9238	0.0259	35.6274	0.9275	0.0415	22.3740	0.9525	0.0433	21.9877	0.9033	0.0420	21.4823
5	1.1329	0.0106	106.5777	1.1236	0.0122	91.7946	1.1591	0.0155	74.7001	1.0867	0.0221	49.0668
6	0.8482	0.0260	32.5902	0.8548	0.0490	17.4504	0.8762	0.0505	17.3358	0.8347	0.0490	17.0250
7	0.7359	0.0217	33.8649	0.7441	0.0218	34.1853	0.7611	0.0238	31.9997	0.7286	0.0215	33.8198
8	1.0333	0.0118	87.3696	1.0312	0.0221	46.5903	1.0615	0.0250	42.5441	1.0008	0.0258	38.7312
9	2.0593	0.1445	14.2478	1.8332	0.0831	22.0481	1.9465	0.2100	9.2672	1.6953	0.1668	10.1637
10	1.0122	0.0060	168.2814	1.0115	0.0142	71.0845	1.0407	0.0184	56.6203	0.9823	0.0188	52.2859
11	0.7670	0.0228	33.6164	0.7749	0.0207	37.4536	0.7931	0.0230	34.5561	0.7581	0.0206	36.7454
12	1.3280	0.0403	32.9303	1.2969	0.0232	55.8280	1.3443	0.0203	66.3769	1.2449	0.0372	33.4925
13	1.0077	0.0144	69.7747	1.0073	0.0357	28.2384	1.0362	0.0376	27.5331	0.9784	0.0377	25.9732
14	0.7258	0.0230	31.5212	0.7341	0.0223	32.9609	0.7506	0.0242	31.0205	0.7191	0.0220	32.6748
15	0.7635	0.0244	31.2471	0.7712	0.0312	24.7018	0.7893	0.0330	23.9075	0.7546	0.0309	24.3834
16	0.9255	0.0283	32.6901	0.9298	0.0647	14.3613	0.9549	0.0657	14.5368	0.9051	0.0655	13.8171
17	0.7208	0.0281	25.6239	0.7296	0.0367	19.9048	0.7458	0.0380	19.6165	0.7150	0.0364	19.6667
18	1.0667	0.0098	109.0634	1.0625	0.0147	72.2204	1.0944	0.0184	59.6082	1.0300	0.0210	49.0061
19	0.9275	0.0062	149.4901	0.9310	0.0085	108.9040	0.9560	0.0144	66.2675	0.9065	0.0120	75.7870
20	1.0070	0.0083	121.9908	1.0066	0.0195	51.5344	1.0354	0.0228	45.4289	0.9777	0.0229	42.6224
21	0.7003	0.0216	32.4967	0.7088	0.0059	120.0917	0.7246	0.0102	70.7776	0.6947	0.0054	129.0271
22	0.9258	0.0188	49.3154	0.9299	0.0492	18.9142	0.9546	0.0507	18.8197	0.9060	0.0499	18.1724
23	1.1468	0.0159	72.0948	1.1363	0.0273	41.6154	1.1725	0.0285	41.0822	1.0987	0.0337	32.5786
24	1.1286	0.0388	29.0820	1.1199	0.0684	16.3742	1.1548	0.0678	17.0316	1.0831	0.0721	15.0188
25	2.6429	0.1412	18.7238	2.1432	0.4311	4.9710	2.3131	0.5174	4.4702	1.9427	0.2612	7.4385
26	0.9871	0.0117	84.3682	0.9878	0.0216	45.6278	1.0157	0.0247	41.1944	0.9601	0.0243	39.4868
27	0.9379	0.0123	76.2922	0.9409	0.0187	50.3129	0.9665	0.0221	43.7255	0.9159	0.0206	44.4310
28	1.2475	0.0281	44.4436	1.2276	0.0768	15.9758	1.2693	0.0753	16.8670	1.1832	0.0828	14.2848
29	0.8364	0.0215	38.8127	0.8428	0.0257	32.7768	0.8640	0.0281	30.7845	0.8229	0.0259	31.7601
30	0.8084	0.0223	36.1710	0.8160	0.0459	17.7720	0.8354	0.0474	17.6101	0.7978	0.0458	17.4056
31	0.7833	0.0214	36.6108	0.7911	0.0401	19.7272	0.8096	0.0417	19.4211	0.7740	0.0399	19.3747
32	0.9997	0.0217	45.9733	0.9998	0.0338	29.5662	1.0284	0.0357	28.7989	0.9711	0.0359	27.0721
33	0.9226	0.0212	43.4855	0.9265	0.0414	22.3540	0.9512	0.0433	21.9806	0.9024	0.0421	21.4328
34	0.9706	0.0132	73.2925	0.9723	0.0356	27.2782	0.9993	0.0376	26.5527	0.9457	0.0371	25.4664
35	1.0799	0.0094	115.2010	1.0748	0.0100	107.4566	1.1074	0.0147	75.1750	1.0414	0.0185	56.1578
36	0.9441	0.0076	124.9118	0.9468	0.0158	59.9735	0.9727	0.0197	49.4099	0.9215	0.0182	50.6241
37	0.5925	0.0314	18.8504	0.6011	0.0207	29.0866	0.6130	0.0218	28.1787	0.5909	0.0203	29.1119
38	1.1141	0.0201	55.3950	1.1067	0.0589	18.7754	1.1408	0.0595	19.1730	1.0717	0.0621	17.2490
39	0.7292	0.0228	31.9804	0.7375	0.0279	26.4302	0.7541	0.0296	25.4338	0.7223	0.0276	26.1533
40	0.7502	0.0186	40.3042	0.7582	0.0199	38.0232	0.7757	0.0223	34.8439	0.7421	0.0198	37.5160
41	1.1882	0.0248	47.9533	1.1737	0.0448	26.2213	1.2123	0.0448	27.0456	1.1331	0.0503	22.5123
42	1.1219	0.0309	36.3142	1.1138	0.0758	14.6883	1.1480	0.0760	15.1089	1.0787	0.0788	13.6830
43	0.8632	0.0149	57.9088	0.8690	0.0151	57.3714	0.8912	0.0188	47.4368	0.8478	0.0161	52.6881
44	1.2441	0.0246	50.5941	1.2236	0.0232	52.7388	1.2656	0.0227	55.7517	1.1786	0.0340	34.6530
45	1.0000	0.0000	-	1.0000	0.0000	-	1.0286	0.0117	88.2470	0.9715	0.0117	83.3479
46	0.7453	0.0277	26.8616	0.7534	0.0200	37.5886	0.7709	0.0222	34.7498	0.7374	0.0198	37.2032
47	1.0314	0.0176	58.7165	1.0299	0.0510	20.2082	1.0598	0.0523	20.2732	1.0000	0.0529	18.9139