THE UNIVERSITY of York

Discussion Papers in Economics

No. 11/22

The Explanatory and Predictive Power of Non Two-Stage-Probability Theories of Decision Making Under Ambiguity

By
John D Hey and Noemi Pace

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

The Explanatory and Predictive Power of Non Two-Stage-Probability Theories of Decision Making Under Ambiguity

By JOHN D HEY AND NOEMI PACE
Revised July 2012

Representing ambiguity in the laboratory using a Bingo Blower (which is transparent and not manipulable) and asking the subjects a series of allocation questions (which are more efficient than pairwise choice questions), we obtain data from which we can estimate by maximum likelihood methods (with explicit assumptions about the errors made by the subjects) a significant subset of the empirically relevant models of behaviour under ambiguity, and compare their relative explanatory and predictive abilities. Our results suggest that not all recent models of behaviour represent a major improvement in explanatory and predictive power, particularly the more theoretically sophisticated ones.

The past decade has seen an explosion of theoretical work in the modelling of behaviour under ambiguity. Now it is the turn of the experimentalists to investigate the empirical validity of these theories. That is the primary purpose of this paper. Specifically, we complement a growing experimental literature, and, in particular, add to the work of Abdellaoui *et al* (2011), Halevy (2007), Ahn *et al* (2010) and Hey *et al* (2010), though our detailed objectives, methods and results differ in many respects substantially from theirs.

In essence, all these papers (and others) are aimed at the same fundamental objective: to discover which of the many theories of behaviour under ambiguity are empirically most appealing. However our work differs from these earlier works in terms of: (1) the representation of ambiguity (except for Hey *et al*); (2) in terms of the experimental design (except for Ahn *et al*); (3) in terms of the theories being explored; and (4) the econometric methods (except for part of Hey *et al*).

Ambiguity is represented in different ways in the experiments on which these different papers were based. Ambiguity is understood as a situation in which probabilities do not exist or the decision-maker does not know the actual probabilities. Both Halevy and Abdellaoui *et al* use as one of their representations the traditional 'Ellsberg Urn': subjects are told what objects are in the urn but are not told the quantities of each object, so that the probability of drawing any particular object can not be known by the subject. Abdellaoui *et al*, given that their objective is to examine the impact of different *sources* of ambiguity, also consider other sources (changes in the French Stock Index, the temperature in Paris, and the temperature at some randomly drawn remote country – all on a particular day). Ahn *et al*'s representation is simply not to tell the subjects what the precise probability of two of the three possible outcomes was; this is a sort of continuous Ellsberg Urn and inevitably suffers from the usual problem that the subjects may simply consider it as the 'suspicious urn'. In contrast, Hey *et al* used an open and transparent representation: a Bingo Blower. This is also what we used in the experiment reported on in this paper. The Blower removes any possible suspicion; moreover it enables us to carry out two treatments which unambiguously have different amounts of ambiguity.

The papers by Hey *et al* and Abdellaoui *et al* use the 'traditional' form of experimental question: pairwise choice, while Halevy uses reservation price questions. In contrast, Ahn *et al* use the *allocation* type of question pioneered originally by Loomes (1991), but forgotten for many years until revived by Andreoni and Miller (2002) in a social choice context, and later by Choi *et al* (2007) in a risky choice context. In this paper we use allocation questions, which are more informative than pairwise choice questions and probably more reliable than reservation price questions, and thus more able to detect true preferences. In this respect the comments by Wilcox (2007) as to the informative nature of experimental data in general, and pairwise choice questions in particular, should be noted.

The set of theories of behaviour under ambiguity is now very large. Those theories – such as maximin, which do not incorporate a preference functional – have been largely discredited (partly by Hey *et al*); we do not consider them here. Of the remaining theories, one can make a broad distinction between the set of theories that use second-order probabilities[1] and the set that does not. For example, if there are I possible events $i = 1, 2, ..., I$, but the probabilities of them are not known, those theories that use second-order probabilities assume that the decision-maker works on the basis that there is a set of J possible values for these probabilities, with the j^{th} set taking values $p_{1j}, p_{2j}, ..., p_{Ij}$ with the probability that the j^{th}set being true given by $\pi_j, \pi_j = 1, 2, ..., J$. In contrast, the set of theories that do not use second-order probabilities may assume that the p_i may take a range of values in the decision-maker's mind, but he or she does not attach probabilities to these possible values. We restrict attention here to this second set (non two-stage-probability models). This is for three reasons: the way we represent ambiguity in the laboratory (there is no obvious first stage); the complexity of the resulting models in the two-stage-probability set; and problems with distinguishability of the underlying preference functionals (because of the large number of parameters). In contrast, Halevy uses two-stage-probability models because his experimental design effectively makes such models appropriate. One could also argue that the same applies to the Ahn *et al* experiment: there they have three possible outcomes 1, 2 and 3. Subjects are told p_2 but they are not told anything about p_1 and p_3 (except that they obey the usual probability rules). However, if subjects had read footnote 4 of their paper[2] then a two-stage-probability representation would have been natural.

Ahn *et al* make another distinction amongst the various specifications of behaviour under ambiguity: between those specifications that they call *smooth* and those that they

[1] Sometimes called Mulitple Prior models.

[2] Which read "In practice, the probability of one of the 'ambiguous' states was drawn from the uniform distribution over [0,2/3]. This distribution was not announced to the subjects." If the distribution had been revealed to the subjects, the decision problem would have involved compound risk rather than ambiguity."

call *kinked*. Essentially this distinction consists on whether preference depends upon the *ordering* of the outcomes: in Expected Utility theory this is not the case and hence this is a smooth specification; in contrast, Choquet Expected Utility and (as one can tell from its name) Rank Dependent Expected Utility (if one considers this as a theory of behaviour under ambiguity) are kinked. Ahn *et al* do not estimate particular preference functionals but rather two general specifications – one smooth and one kinked. They note that the smooth specification "can be derived from" Recursive Expected Utility (REU, which is a two-stage-probability model); while the kinked specification "... can be derived as a special case of a variety of utility models: MEU, CEU, Contraction Expected Utility, and α-MEU[3]". We note two things: first that the smooth specification does not come *only* from REU (indeed it comes from several other models, such as the Variational model of Maccheroni *et al* (2006)); and secondly, but perhaps more importantly, Ahn *et al* do not estimate any preference functionals that come *specifically* from the models that they mention. So they do not test *directly* any of the recent theories of behaviour under ambiguity.

Abdellaoui *et al* effectively investigate only one model – essentially Rank Dependent Expected Utility (RDEU) theory. This has two key elements, a utility function, which they take to be CRRA (Constant Relative Risk Averse), and a (probability) weighting function, which they take, in the 'Ellsberg Urn' part of the experiment, to be of the Prelec form: $w(p) = (\exp(-(-\ln(p))^{\alpha}))^{\beta}$. It may seem a bit odd using probabilities in a study of ambiguity, but these probabilities are the true probabilities, which, of course, the experimenters know, if not the subjects. Using these functional specifications, RDEU is a special case[4] of Choquet Expected Utility, which we estimate. In the 'Natural Uncertainties' part of the experiment they do not assume any particular form for the weighting function, so RDEU in this context is precisely Choquet – which we estimate. While Abdellaoui *et al* assume CRRA preferences, we estimate both assuming CRRA and CARA preferences.

There are significant econometric differences between these various papers. First, and rather hidden from view, is the fact that we carried out extensive pre-experimental simulations to ensure that we had a sufficient number and an appropriate set of questions to ask the subjects; too many experiments have too few questions and thus lack power to discriminate amongst the theories. Second, the estimation methods vary. Underlying any particular chosen estimation method, there is an assumption about the stochastic specification of the model. Usually this is tacit; it should be explicit, particular as there is an obvious source for the stochastic component of the data – if one is estimating subject by subject (which is the case in all these papers) this comes either from randomness in preferences or from errors made by the subjects. We see no mention in any of this literature of randomness in preferences, so the noise, the stochastic component, must come from errors made by the subjects. We explicitly include a story of such mistakes. Indeed,

[3](Our note) MEU, CEU and α-MEU are respectively MaxMin Expected Utility, Choquet Expected Utility, and Alpha Expected Utility (all of which we consider specifically later).

[4]If they had estimated the weighting function at all points, rather than estimating the parameters of the particular functional form, it would have been precisely Choquet.

we explore two different stochastic specifications, a normal error story and a beta error story, depending upon the functional form we take for the utility function (as we explain later). We estimate the various preference functionals with the stochastic specifications specifically built in to the estimation; this does not appear to be the case in the other papers we have mentioned in this section (though it is the case in Anderson *et al* (2009) which we will discuss later), so their assumptions about errors is not clear.

We also go one step further than all these other papers. Believing that economics is all about predicting, rather than just explaining, we compare our different models by seeing how good they are at predicting. For the importance of this, see Wilcox (2007 and 2011).

In summary: we represent ambiguity in the laboratory in an open and non-manipulable manner; we ask a set of allocation questions to the subjects (obviously with an appropriate incentive mechanism) chosen after extensive simulations; we use maximum likelihood estimation, with two carefully-chosen stochastic specifications, to estimate a significant sub-set of the empirically relevant theories of behaviour under ambiguity, using two different functional forms for the utility function, and to compare their relative goodness of fit; finally we compare the various theories in terms of their predictive ability. To prepare the reader for what is to come, we should warn that the theorists are going to be disappointed: the more recent theories are in general only marginally empirically superior to SEU in terms of fitting but less so in terms of prediction.

The rest of the paper is organised as follows. In the next section we give a brief overview of the theories that we are going to fit to our data. The following section describes the way that our data was generated in our experiment. We then relate what we do to the literature, a part of which we have discussed in this introduction, and give more detail about what others have done. A section describing the technicalities underlying our analysis then follows, after which we present our results. We then conclude.

I. Theories under investigation

This section discusses the theories of decision-making under ambiguity that we investigate. We confine our attention to those theories in which there is an explicit preference functional, and hence we exclude earlier theories which proceed directly to a decision rule.[5] In all the theories which we consider the decision maker is perceived, in any decision problem, as maximising the value of some preference functional. As noted above, we include: (1) Subjective Expected Utility (SEU) theory in which the decision-maker is envisaged as working with subjective probabilities; (2) the Choquet Expected Utility (CEU) model, usually nowadays accredited to Schmeidler (1989), which allows the agent's beliefs to be represented by unique but nonadditive "capacities"; (3) the Alpha Expected Utility model (AEU) of Ghirardato *et al* (2004), which models the agent's beliefs as being represented by a set of probabilities (but without attaching probabilities to the members of this set); (4) Vector Expected Utility (VEU) of Siniscalchi (2009) in which uncertain prospect is assessed according to a baseline expected utility evaluation

[5] Such as, for example, MaxMin (in which the decision-maker looks at the worst that can happen and makes that as good as possible) and MaxMax (in which the decision-maker looks at the best that can happen and makes that as good as possible). See Hey, Lotito and Maffioletti (2010) for the empirical evidence against such theories.

and an "adjustment that reflects the individual's perception of ambiguity and her attitude toward it" (Siniscalchi 2009, p.1); and (5) the Contraction Model of Gajdos, Hayashi, Tallon and Vergnaud (2008) which combines Maxmin with Expected Utility at a particular point in the probability set. We note that SEU is a 'smooth' specification in the sense used above, while all the rest are kinked specifications. We tried to fit the Variational model of Maccheroni *et al* (2006), which is a smooth specification, but without success (in terms of goodness of fit); this may have been the consequence of the particular context of our experiment. We give an overview of these theories below.[6] We restrict attention in both the overview and the detail to decision problems with at most three events - which was the case in our experiment. Call these events E_1, E_2 and E_3 To each event there will be associated an outcome to the decision-maker which consists of an amount of money. We denote the utility of the decision-maker for these three outcomes u_1, u_2 and u_3. For some of the theories – those with a 'rank-dependent' flavouring – the ordering of the outcomes will be crucial and we will assume in what follows that $u_1 \geq u_2 \geq u_3$ though it should be noted that it is not necessarily the case that the ordering of the outcomes is the same as the original ordering of the events: this depends upon the decisions that the decision-maker makes. Let us denote the event which leads to the highest outcome by $E_{(1)}$, that to the second highest outcome by $E_{(2)}$ and that to the lowest outcome by $E_{(3)}$. We note that the set $\{E_{(1)}, E_{(2)}, E_{(3)}\}$ consists of the numbers 1, 2 and 3, though not necessarily in that order. In summary, each of our models (SEU, CEU, AEU, VEU and COM) implies a particular preference functional, respectively equations (1), (2), (4), (6) and (7). It is clear that these are different, except insofar as (2), (4), (6) and (7) reduce to (1) (and hence CEU, AEU, VEU and COM reduce to SEU) when respectively, the CEU capacities are additive, the set D of probabilities consists of a single element, the parameter δ is zero, and the set D of probabilities consists of a single element. Given that (2), (4), (6) and (7) are different it follows that the models are observationally distinguishable: different models imply different preference functionals and hence different decisions. However this is not to deny that the function $A(.)$ in VEU and the sets D in AEU and COM could be such that they lead to the same preference functionals. But the crucial point is that our specifications of the different models imply different preference functionals and hence are observationally distinguishable.

SUBJECTIVE EXPECTED UTILITY THEORY

The preference functional for SEU is given by

$$SEU = \sum_{i=1}^{3} p_i u_i \tag{1}$$

where p_i is the subjective probability that event $E_{(i)}$ occurs. In this case $p_i = \text{Prob}(E_{(i)})$ for all i, and, of course $p_1 + p_2 + p_3 = 1$.

[6] More technical detail are available on this web-site:
http://www.york.ac.uk/economics/research/research-clusters/experimental-economics/research/ongoing-projects/

CHOQUET EXPECTED UTILITY THEORY

According to Schmeidler (1989), the Choquet Expected Utility of a lottery is given by

$$CEU = \sum_{i=1}^{3} \bar{w}_i u_i \tag{2}$$

where the $\bar{w}$'s are weights that depends on nonadditive *capacities* w that satisfy the normalisation conditions and monotonicity. In the context of our experiment, a CEU subject works with six nonadditive capacities $w_{E_{(1)}}, w_{E_{(2)}}, w_{E_{(3)}}, w_{E_{(2)} \cup E_{(3)}}, w_{E_{(1)} \cup E_{(3)}}$ and $w_{E_{(1)} \cup E_{(2)}}$ referring to the three events and their pairwise unions. Crucially, the weights $\bar{w}$ depend upon the ordering of the outcomes:

$$\begin{aligned} \bar{w}_1 &= w_{E_{(1)}} \\ \bar{w}_2 &= w_{E_{(1)} \cup E_{(2)}} - w_{E_{(1)}} \\ \bar{w}_3 &= 1 - w_{E_{(1)} \cup E_{(2)}} \end{aligned} \tag{3}$$

We note that the main difference between CEU and SEU consists in the additive probability measure being replaced by a nonadditive capacity measure. If the capacities are actually probabilities (that is, if $w_{E_{(2)} \cup E_{(3)}} = w_{E_{(2)}} + w_{E_{(3)}}$, $w_{E_{(1)} \cup E_{(3)}} = w_{E_{(1)}} + w_{E_{(3)}}$, $w_{E_{(1)} \cup E_{(2)}} = w_{E_{(1)}} + w_{E_{(2)}}$ and $w_{E_{(1)}} + w_{E_{(2)}} + w_{E_{(3)}} = 1$) then (2) is equivalent to (1). We note that CEU is the same as Rank Dependent Expected Utility (which is not regarded by all as a theory of behaviour under ambiguity because it uses objective probabilities, but also uses, as we have already noted, to rescue it from that criticism, a weighting function, mapping objective probabilities into subjective probabilities) under an appropriate interpretation of that latter theory.[7] Similarly Cumulative Prospect Theory, with a fixed reference point, can be regarded in the same way as CEU.

ALPHA EXPECTED UTILITY THEORY

Alpha Expected Utility theory (AEU) was proposed by Ghirardato *et al* (2004) as a generalization of the theory proposed in Gilboa and Schmeidler (1989). Ghirardato *et al* (2004)'s model implies that, although the decision maker does not know the true probabilities, he or she acts as if he or she believes that the true probabilities lie within a set D of probabilities on different events. We can refer to each prior $p \in D$ as a "possible scenario" that the decision maker envisions. According to Ghirardato *et al*, the set D of probabilities represents formally the ambiguity that the decision maker feels in the decision problem (they introduce the concept of "revealed ambiguity"). In other words, the size of the set D measures the perception of ambiguity. The larger D is,

[7] In the context of our experiment, where there are three outcomes and hence 6 capacities, then the relationship between the two theories is given by the following, where $p_{(1)}, p_{(2)}, p_{(3)}$ are the objective probabilities and $w(.)$ is the weighting function, and the capacities for CEU are as denoted above:
$w_{E_{(i)}} = w(p_i)$ for $i = 1, 2, 3$ and
$w_{E_{(j)} \cup E_{(k)}} = w(p_j + p_k)$ for $j \neq k \in 1, 2, 3$

the more ambiguity the decision maker appears to perceive in the decision problem. In particular, no decision maker perceives less ambiguity than one who reveales a singleton set $D = \{p_1, p_2, p_3\}$. In this case the decision maker is a SEU maximiser with subjective probabilities p_1, p_2 and p_3.

According to Alpha Expected Utility Theory, decisions are made on the basis of a weighted average of the minimum expected utility over the set D of probabilities and the maximum expected utility over this set:

$$AEU = \alpha \min_{p \in D} \sum_{i=1}^{3} p_i u_i + (1-\alpha) \max_{p \in D} \sum_{i=1}^{3} p_i u_i \tag{4}$$

The parameter α can be interpreted as an index of the ambiguity aversion of the decision maker. The larger is α the larger is the weight the decision maker gives to the pessimistic evaluation given by $\min_{p \in D} \sum_{i=1}^{3} p_i u_i$.

In order to estimate this model we need to characterise the set D. The theory offers no advice and we chose the simplest: that the set is defined by three lower bounds $\underline{p}_1$, $\underline{p}_2$ and $\underline{p}_3$ (where $\underline{p}_1 + \underline{p}_2 + \underline{p}_3 \leq 1$) plus the condition that every element in the set has $p_1 \geq \underline{p}_1$, $p_2 \geq \underline{p}_2$ and $p_3 \geq \underline{p}_3$. In addition, of course $p_1 + p_2 + p_3 = 1$ for each element in the set. These conditions imply that the set D is a triangle properly within the Marschak-Machina Triangle. It reduces to a single point, and hence AEU reduces to SEU, if $\underline{p}_1 + \underline{p}_2 + \underline{p}_3 = 1$.

VECTOR EXPECTED UTILITY

The Vector Expected Utility (VEU) theory has been recently proposed by Siniscalchi (2009). In this model, an uncertain prospect is assessed according to a baseline expected utility evaluation and an adjustment that reflects the individual's perception of ambiguity and his or her attitude toward it. This adjustment is itself a function of the exposure to distinct sources of ambiguity, and its variability.

The key elements of the VEU model are a baseline probability and a collection of random variables, or adjustment factors, which represent acts exposed to distinct ambiguity sources and also reflect complementarity between ambiguous events.

The VEU model can be formally defined as follows:

$$VEU = \sum_{i=1}^{3} p_i u_i + A\left(\left(\sum_{i=1}^{3} p_i \zeta_{ji} u z_i\right)_{1 \leq j < 3}\right) \tag{5}$$

Here $p = (p_1, p_2, p_3)$ is the baseline prior; for $1 \leq j < 3$, each $\zeta_j = (\zeta_{j1}, .., \zeta_{j3})$ is an adjustment factor that satisfies $E_p[\zeta_j] = \sum_{i=1}^{3} p_i \zeta_{ji} = 0$; and A: $\mathbb{R}^n \to \mathbb{R}$ satisfies $A(0) = 0$ and $A(\phi) = A(-\phi)$. The function A is an adjustment function that reflects attitudes towards ambiguity. We need to specify the function $A(.)$ and also the values of the ζ.[8] After some helpful e-mail correspondence with Marciano Siniscalchi (though we do not imply that our modelling of the VEU model has his approval) and some simplification, we find that the VEU objective function takes the form, under these assumptions:

$$VEU = \sum_{i=1}^{3} p_i u_i - \delta(|u_1 - u_2| + |u_2 - u_3|) \tag{6}$$

This has intuitive appeal: decisions are made on the basis of expected utility 'corrected' for differences between the utilities of the various outcomes, weighted by a parameter δ that reflects the decision-maker's attitude to ambiguity. We should note that the most general version of VEU (as specified in equation (5)) has many implementable forms, depending upon the specification of the function A(.), since the restrictions on A(.) do not enable us to specify it precisely. It follows that equation (6) is just one of many possible specifications of the VEU preference functional. When we refer to "estimating the VEU model" this restriction should be taken into account.

The Contraction Model

Gajdos *et al* (2008) proposed a model (the "Contraction Model" or COM) in which it is possible to compare acts under different objective information structures. According to this theory, preferences are given by

$$COM = \lambda \min_{p \in D} \sum_{i=1}^{3} p_i u_i + (1 - \lambda) \sum_{i=1}^{3} P_i u_i \tag{7}$$

where λ measures imprecision aversion and P_1, P_2, P_3 is a particular probability distribution in the set D of possible distributions. It is what is called the 'Steiner Point' of the set - which is, in a particular sense, the 'centre' of the set (Neuberg, 1886). If we take the set D of possible distributions as all points (p_1, p_2, p_3) such that $p_1 + p_2 + p_3 = 1$ and $p_1 \geq \underline{p}_1$, $p_2 \geq \underline{p}_2$, $p_3 > \underline{p}_3$ then the Steiner point is the point (P_1, P_2, P_3) where $P_i = \underline{p}_i + (1 - \underline{p}_1 - \underline{p}_2 - \underline{p}_3)/3$ for $i = 1, 2, 3$. We note that we have characterised this set D (of possible probabilities) in the same way as we have done for the Alpha Expected Utility model - as a triangle properly within the Marschak-Machina Triangle. The Steiner point is the 'central' point of this triangle.

II. Our Experimental Design

As we have already noted, in our experiment ambiguity was implemented with a Bingo Blower and subjects were presented with a set of allocation problems, which were de-

[8] More details of the assumptions that we have made are available on the web-site: http://www.york.ac.uk/economics/research/research-clusters/experimental-economics/research/ongoing-projects/

termined after extensive Monte Carlo simulations. We implemented two separate treatments, which we describe below.

Subjects completed the experiment individually at screened computer terminals. They were given written instructions and then shown a PowerPoint presentation of the instructions. There was a Bingo Blower in action at the front of the laboratory throughout the experiment.[9] The Bingo Blower is a rectangular-shaped, glass-sided, object some 3 feet high and 2 feet by 2 feet in horizontal section. Inside the glass walls are a set of balls in continuous motion being moved about by a jet of wind from a fan in the base. In addition, images of the Blower in action were projected *via* a video camera onto two big screens in the laboratory. Subjects were free at any stage to go close to the Blower to examine it as much as they wanted. All the balls inside the Blower can at all times be seen by people outside, but, unless the number of balls in the Blower is low, the number of balls of differing colours can not be counted because they are continually moving around. Hence the balls in the bingo can be seen but not counted (unless the total number of balls is low), and the information available is not sufficient to calculate objective probabilities. This ensures that, while objective probabilities do exist, the decision-makers cannot know them. In this way, we have created a situation of genuine ambiguity which eliminates: the problem of suspicion; the problem of using directly a second-order probability distribution; and the problem of using real events, therefore keeping the problem more similar to the original Ellsberg one (Ellsberg, 1961). We note that a further advantage of this way of creating ambiguity in the laboratory is the fact that the information available is the same for all subjects. Hence there is no role for the so called 'comparative ignorance' (Fox and Tversky, 1995), and hence we can exclude such a factor as a possible explanation of behaviour.[10]

This Bingo Blower played a crucial role in representing ambiguity and in providing incentives. Inside the Bingo Blower were balls of three different colours: pink, yellow and blue. The number of each colour depended on the treatment:

	Treatment 1	Treatment 2
pink	2	8
yellow	3	12
blue	5	20

In Treatment 1 the pink and blue balls could be counted, though one might not be sure of the number of blue balls; this was *the least ambiguous treatment*. In Treatment 2 the balls of each colour could not be counted; this was *the most ambiguous treatment*. Note that in this latter treatment subjects could get some idea of the relative numbers of balls of the different colours but not count the numbers precisely. It was reasonably clear that there were more blue balls than yellow, and more yellow than pink, though precise

[9] On our site http://www.york.ac.uk/economics/research/research-clusters/experimental-economics/research/ongoing-projects/ can be found the instructions (both in Word and PowerPoint) and videos of the Bingo Blower, as well as screenshots of the experimental software.

[10] One criticism concerning the implementation of ambiguity in the lab using the Bingo Blower comes from Morone and Ozdemir (2011). The criticism consists of the observation that the ability of getting the right probabilities is subject specific; that is, subjects have different counting skills, or might have problems in the perception of colours. This criticism may be true but it is not clear how this could affect the validity of the Bingo Blower in generating ambiguity in the lab.

calculations could not be made.

Sixty six subjects completed Treatment 1 and sixty three completed Treatment 2. In both treatments, subjects were presented with a total of 76 questions. Each of these asked them to allocate a given quantity of tokens between the colours. There were two *types* of question. Type 1 asked them to allocate the tokens between *two* of the colours (with an implicit allocation of zero tokens to the third colour); Type 2 asked them to allocate the tokens between one of the three colours and the other two. In each problem subjects were told the *exchange rate* between tokens and money for each of the colours in the question. Thus an allocation of tokens implied an allocation of money to two or three of the colours.

We provided an incentive for carefully choosing the allocations with the following payment scheme. We told subjects that, after answering all 76 questions, one of the questions would be chosen at random, and the subject's allocation to the two or three colours for that problem retrieved from the computer. At that point the subject and the experimenter went over to the Bingo Blower, and the subject rotated the tube to expell one ball. The colour of the ball, the question picked at random and their answer to that question determined their payment. To be precise: if the problem chosen was one of Type 1, then they would be paid the money implied by their allocation to the colour of the ball expelled; if it was the colour not mentioned in that question they would be paid nothing; if the problem chosen was one of Type 2, then they would be paid the money implied by their allocation to the colour of the ball expelled. In addition they received a show-up fee of £5. They filled in a brief questionnaire, were paid, signed a receipt and were free to go. A total of 129 subjects participated at the experiments, 40 of them at CESARE at LUISS in Rome (Italy) and the remaining 89 at EXEC at The University of York (UK). In both cases, subjects were recruited using the ORSEE (Greiner 2004) software and the experiment was run using a purpose-written software written in Visual Basic 6.[11]

III. Related Experimental Literature

Having described our experimental implementation and motivation we are now in a position to survey the relevant experimental literature in more detail. We confine ourselves to recent important contributions to the literature; earlier literature is surveyed in Camerer and Weber (1992) and Camerer (1995) while more recent mainly theoretical literature is surveyed by Etner, Jeleva and Tallon (2012).

Hey *et al* (2010), using the same implementation of ambiguity in the laboratory as we use here, also with three possible outcomes, but asking a large number (162) of pairwise choice questions, examined the descriptive and predictive ability of twelve theories of behaviour under ambiguity: some very old and not using a preference functional (proceeding directly to a decision rule) such as the original MaxMin and MaxMax; and some very recent, such as the Alpha Expected Utility model. The findings were that the very

[11] Which can be found and downloaded from http://www.york.ac.uk/economics/research/research-clusters/experimental-economics/research/ongoing-projects/

old simple models (those without a preference function) were largely discredited, but that more modern and rather sophisticated models (such as Choquet) did not perform sufficiently better than simple theories such as Subjective Expected Utility theory. Estimation of the preference functions was done using maximum likelihood techniques with the stochastic specification determined by a model of how subjects made errors in their pairwise choices.

Ahn *et al* (2010) used allocation questions, like we do here, but implemented ambiguity by not telling the subjects the true objective probabilities of two of the three possible outcomes of the experiment. They did not look at the predictive ability of any models; neither did they examine the descriptive performance of any specific theory. Instead they examined two broad classes of functionals, smooth and kinked, which are special cases of various theoretical models that we specifically estimate. Econometrically, they estimated, subject by subject, the risk-aversion parameter of an assumed Constant Absolute Risk Aversion utility function, and a second parameter measuring ambiguity aversion, using Non Linear Least Squares (NLLS), that is by minimising the sum of squared differences between actual allocations and the theoretically optimal allocations for those risk and ambiguity aversion coefficients. They comment in a footnote that "...for simplicity, the estimation technique for both specifications is NLLS, rather than a structural model using maximum likelihood (ML). We favor the NLLS approach, because it provides a good fit and offers straightforward interpretation." They do not provide a formal justification to the choice of NLLS instead of ML.

Halevy (2007) implemented ambiguity in the laboratory using traditional Ellsberg Urns and asked reservation price questions. Because of the way that his 'Ellsberg Urns' were implemented, his set of models includes some models that we do not consider here, particularly two-stage-probability models such as Recursive Nonexpected Utility and Recursive Expected Utility. But we include some that he does not - making the two papers complementary. He used *reservation price* questions; we should describe and discuss these - as they are an alternative to pairwise choice questions and to allocation questions. Essentially he wants to know how much subjects value bets on various events. Let us consider a particular Ellsberg Urn and a particular colour. The subject is asked to imagine that he or she owns a bet which pays a certain amount of money ($2) if that coloured ball is drawn from that particular urn. Halevy wanted to elicit the subject's reservation price for this bet; this reservation price telling us about the subject's preferences. Halevy used the Becker-DeGroot-Marschak mechanism: "the subject was asked to state a minimal price at which she was willing to sell the bet... The subject set the selling price by moving a lever on a scale between $0 and $2. Then a random number between $0 and $2 was generated by the computer. The random number was the “buying price” for the bet. If the buying price was higher than the reservation price that the subject stated, she was paid the buying price (and her payoff did not depend on the outcome of her bet). However, if the buying price was lower than the minimal selling price, the actual payment depended on the outcome of her bet." This BDM technique is well-known in the literature, but is complicated to describe and difficult for subjects to understand. Moreover there are well-known problems, see Karni and Zafra (1979), with using this technique when preferences are *not* expected utility preferences - which, of course, is precisely the concern of the paper. Halevy did not use his data to estimate preference functionals and hence did not compare their descriptive and predictive power; instead he carried out an extensive set of tests of the various theories. Unfortunately this econometric procedure does not help to draw unique conclusions about the 'best' preference functional, even for individual subjects. Indeed Halevy concludes that his "...findings indicate that currently there is no unique theoretical model that universally captures ambiguity preferences".

Abdellaoui *et al* (2011) investigated only Rank Dependent Expected Utility theory. They did not explicitly examine its descriptive (nor predictive) ability, being more concerned with the effect on the estimated utility and weighting functions of different *sources* of ambiguity. As we have already noted, they implemented ambiguity in the laboratory in two ways: in one part of the experiment, using 8-colour 'Ellsberg Urns'; and in the other part using 'natural' events. They elicited certainty equivalents (or reservation prices) in order to infer preferences, not using the BDM mechanism (presumably because of the problems we have alluded to above), but instead using Holt-Laury price lists.[12] This mechanism seems to be a better way of eliciting certainty equivalents, even though the outcome does appear to be sensitive to the elements in the list – the number of them and their range. The resulting certainty equivalents are a valuation, just as Halevy's reservation prices, even though they come from a set of pairwise choice questions. However econometrically it must be the case that the valuation resulting from a list with *n* elements is less informative than *n* independent pairwise choice questions. They estimated utility functions (assumed to be power or CRRA) "using nonlinear least squares estimation with the certainty equivalent as dependent variable"; similarly they estimated the weighting function by "minimising the quadratic distance". They do not explain why.

Anderson *et al* (2009) use a technique similar to that used by Ahn *et al* (2010) in estimating two parameters (one a measure of risk aversion and the other a measure of ambiguity) in a minimalist non-EU model. They comment that this minimalist model comes either from the Source-Dependent Risk Attitude model or the Uncertain Priors model; in our termininology it is a two-stage-probability model[13] that looks exactly like Recursive Expected Utility. The bottom line is the following: suppose that there are I possible outcomes $i = 1, 2, ..., I$ with unknown probabilities. The decision-maker has a set of J possible values for these probabilities; we denote the j'th possible value $p_{1j}, p_{2j}, ..., p_{Ij}$ and the decision-maker considers that the probability that this is the correct set is π_j. The preference function is the maximisation of

$$\sum_{j=1}^{J} \pi_j v[\sum_{i=1}^{I} p_{ij} u(x_i)]$$

[12] In which subjects are presented with a set of pairwise choices arranged in a list. In each pair subjects are asked to choose between some ambiguous lottery and some certain amount of money. As one goes down the list, the certain amount increases. The subject's certainty equivalent is revealed by the point at which the subject switches from choosing the lottery to choosing the certain amount. See Holt and Laury (2002).

[13] Chambers *et al* (2010) also investigate a generic Multiple Priors model.

Note that there are two functions here: $u(.)$ which can be considered as a normal utility function, capturing attitude to risk; and $\upsilon(.)$ which can be considered as an ambiguity function; note that if $\upsilon(y) = y$ then this model reduces to Expected Utility theory. It is the non-linearity of $\upsilon(.)$ that captures aversion to ambiguity. Anderson et al (2009) assumed that both these functions are power functions - so that $u(x) = x^{\alpha}$ and $\upsilon(y) = y^{\beta}$ They estimated the two parameters α and β using maximum likelihood techniques (with careful attention paid to the stochastic specification) and assumptions[14] about the π's and p's.

IV. Technical Assumptions

Before proceeding to our estimates we need to make some technical assumptions. In particular we need to decide on our stochastic specifications and those concerning the form of utility function. These are interelated decisions. Both are important, as Wilcox (2007, 2008 and 2011) makes clear, though our context is different from his as the experimental task in our experiment requires subjects to make a series of *allocations*, rather than to make a series of *pairwise choices*. This has implications for the stochastic strucure, but this, in turn, depends upon the assumed utility function: as we have already noted the two issues are inter-related.

The two most commonly assumed utility functions are CARA (Constant Absolute Risk Aversion) and CRRA (Constant Relative Risk Aversion). If we take the CARA form[15]

$$(8) \qquad \begin{aligned} u(x) &= \frac{1 - \exp(-rx)}{1 - \exp(-75r)} \text{ if } r \neq 0 \\ &= x/75 \text{ if } r = 0 \end{aligned}$$

where r, the coefficient of absolute risk aversion, is subject-specific, then we can find the solution to the subjects' optimisation problem. This, for all the preference functionals, can be written as the maximisation of the expression

$$(9) \qquad w_1 u(e_1 x_1) + w_2 u(e_2 x_2)$$

subject to the constraint that $x_1 + x_2 = m$. Here m denotes the amount of tokens to allocate, x_i the tokens allocated to colour i and e_i the exchange rate between tokens allocated to colour i and money ($i = 1, 2$). The weights w_1 and w_2 depend upon the problem and the preference functional. The solution to this problem can be shown to be:

$$(10) \qquad \begin{aligned} x_1^* &= \frac{e_2 m + \{\ln[(w_1 e_1)/(w_2 e_2)]\}/r}{e_1 + e_2} \\ x_2^* &= \frac{e_1 m + \{\ln[(w_2 e_2)/(w_1 e_1)]\}/r}{e_1 + e_2} \end{aligned}$$

[14] We note that the authors admit that the assumptions were quite strong and discuss the serious identification problems with two-stage-probability models.

[15] £75 was the maximum payoff from the experiment.

We note that there is no guarantee that the optimal allocations are positive and less than m.

Before discussing the CRRA functional, we turn to the stochastic specification 'most naturally' associated with this CARA function: namely a normal distribution with zero mean. So while subjects, under this CARA specification *should* allocate the x_1^* (and x_2^*) above, we assume that their *actual* allocations are $x_1 = x_1^* + \varepsilon$ (and $x_2 = x_2^* - \varepsilon$) where ε is a normally distributed error term with zero mean. We follow precedent in assuming that the variance of ε is constant[16] and equal to $1/s^2$. We estimate s, the precision, along with the other parameters. We should, of course, note that, if $x_1^* + \varepsilon$ is larger than m, then the subject is forced to truncate the actual value to m; similarly if $x_1^* + \varepsilon$ is negative, then the subject is forced to truncate the actual value to 0. In addition subjects could specify their desired allocations to just two decimal places. All of this was taken into account in the maximum likelihood estimation program. We call this Specification 1.

Specification 2 differs from Specification 1 in two respects. First we assume a CRRA utility function of the following form:

$$
\begin{aligned}
u(x) &= \frac{x^{1-\frac{1}{r}} - 1}{1 - \frac{1}{r}} \text{ if } r \neq 1 \\
\ln(x) \text{ if } r &= 1
\end{aligned}
\tag{11}
$$

This particular parameterisation (the parameter r is the inverse of the usual coefficient of risk aversion) may appear unusual but it allows us to express the optimality conditions in a particularly elegant form. Given the same objective as above, the solution is:

$$
\begin{aligned}
x_1^* &= \frac{e_1^{r-1} w_1^r}{e_1^{r-1} w_1^r + e_2^{r-1} w_2^r} m \\
x_2^* &= \frac{e_2^{r-1} w_2^r}{e_1^{r-1} w_1^r + e_2^{r-1} w_2^r} m
\end{aligned}
\tag{12}
$$

In this case, the optimal allocations *are* bounded between 0 and m. Hence the proportions x_1^*/m and x_2^*/m, are bounded between 0 and 1. This suggests an alternative stochastic specification, which fits in more naturally with the boundedness of these optimal proportions, namely a Beta distribution. Specifically in Specification 2 we take the actual proportional allocation x_1/m to have a beta distribution with parameters $x_1^*(s-1)$ and $x_2^*(s-1)$. This guarantees that the mean of x_1 is x_1^* and its variance is $x_1^* x_2^*/s$. So the variance of $x_1$is not constant but is zero at 0 and m and reaches a maximum when $x_1^* = m/2$. It also follows (since $x_1^* + x_1^* = 1$) that x_2 has a beta distribution with pa-

[16]Wilcox discusses the possibility that the variance might be question-specific and gives good arguments why this might be so. But it is not clear how his arguments, relevant to his pairwise-choice context, carry over to our allocation context.

rameters $x_1^*(s-1)$ and $x_2^*(s-1)$. Of course we may still observe actual proportional allocations equal to 0 and 1 because of the rounding of subjects' choices.

Finally we consider Specification 3, in which we assume a CRRA utility function plus a zero mean normal heteroscedastic error term added to the absolute allocation (this latter part being the same as in Specification 1). While Specification 2 might appear more 'natural' there is nothing illogical about specification 3.

V. Results

We estimated each of the 5 preference functionals for each of the 129 subjects and for each of the three specifications on a subset of the data – namely a randomly chosen 60 of the 76 questions[17] – using the constrained maximum likelihood procedure in GAUSS. We thus have, for each preference functional, for each subject, and for each specification, estimates of the parameters of the functional, of s (the precision) and of r (the risk aversion parameter). In addition, we have the maximised log-likelihood.[18] We then used, for each subject and each preference the estimated parameters to predict behaviour on the remaining 16 questions. This gives us a prediction log-likelihood for each functional and for each subject - this is, of course, a measure of the predictive ability of the theory. In the tables that follow we break down some of the summary information by Treatment; recall that Treatment 1 (66 subjects) was essentially a case of risk; while Treatment 2 (63 subjects) was clearly one of ambiguity.[19]

If we are looking for a 'bottom line' at this stage, it is this: *there is much more variability across subjects and across specifications than across preference functionals and across treatments, though there is an interesting treatment effect*. We will explore the implications of this in our conclusions.

We start with Table 1 which shows the mean and standard deviation (across all subjects in each Treatment and in both Treatments) of the fitted log-likelihoods. It should be noted that, for each preference functional the means are consistently higher (the fit better) in Specification 2. However this Table does not allow us to compare the goodness of fit *across* preference functionals, for the simple reason that they have different degrees of freedom (SEU has 4 estimated parameters, CEU 8, AEU 6, VEU 5 and COM 6). If we correct the fitted log-likelihoods for the degrees of freedom by calculating the Bayesian Information Criterion (BIC)[20], we get Table 2; recall that the *lower* the BIC the better. Once again we get the same message: on average Specification 2 is better than the other two specifications. As these BIC values are comparable, we can also conclude that

[17] Because the subjects received the 76 questions in different orders (and with the colour on the left and the colour on the right randomly selected) this means that the position of the 60 estimation questions (and hence the 16 prediction questions) varied from subject to subject, but for each subject they were randomly positioned).

[18] Note that for specification 2, since the variables to be explained are the *proportions* of the endowment allocated the various colours, in order to make the log-likelihoods comparable with those in the other two specifications, we need to subtract from the maximised log-likelihoods the sum of the natural logarithms of the amounts to be allocated in the relevant problems.

[19] All this information is available at http://www.york.ac.uk/economics/research/research-clusters/experimental-economics/research/ongoing-projects/

[20] This is given by $k\ln(n) - 2LL$, where k is the number of estimated parameters, n the number of observations and LL the maximised log-likelihood.

there is a larger variation of the average BIC across specifications than across preference functionals.

In order to demonstrate that the means shown in Tables 1 and 2 hide rather large variations across subjects, we present histograms of the BIC across subjects in Figures 1, 2 and 3. Within any one specification, the shapes of the distributions of the subjects across preference functionals are very similar, and there are high correlations between the log-likelihoods over preference functionals. We note that the shapes of the distributions vary across specifications.

If we now rank the various preference functionals and specifications using the BIC, we get Table 3. Here we report the cumulative percentage in each ranking position: so for example, in Treatment 1, SEU is ranked first for 23% of our subjects; is ranked first or second for 29% of our subjects; and so on. It is very clear from this Table that SEU and VEU on specification 2 are particularly good. Indeed for 59% of subjects one or other of these two has the highest BIC in Treatment 1, and 63% for Treatment 2.

If we now turn to the prediction log-likelihoods, things are not so clear cut, though Specification 2 still emerges as the best on aggregate. If we look at Table 4, we see that the prediction log-likelihoods are, on average, higher for Specification 2 for all 5 preference functionals, though once again there are considerable variations across subjects. Table 5 gives cumulative rankings and is the counterpart of Table 3 for the fitted log-likelihoods. The findings are less clear-cut here; in Treatment 1, Specification 2 is the best predicting specification with 43% of first places; SEU is first 5% of the time and VEU 20% of the time; in Treatment 2, the corresponding figures are 60%, 5% and 32%. We note that we are using here *uncorrected* log-likelihoods.

Although Specification 2 is not always the best, on average it is, and from now on, to save space and repetition we report results just for Specification 2. We now return to the fitting and ask about *statistical significance* of our estimates. Because of the relationships between the preference functionals, we need to carry out two kinds of tests: nested tests and non-nested tests. We note that SEU is nested within all the other four preference functionals, but none of them (in the way that we have implemented them here) are nested inside any of the others. Hence for each of CEU, AEU, VEU and COM relative to SEU a likelihood ratio test is appropriate; for each of CEU, AEU, VEU and COM against the others a Clarke test is appropriate. The results are reported in Table 6, both at 5% (Table 6A) and 1% (Table 6B). Looking at the first column of Table 6A, we note that neither CEU nor COM do particularly well. Indeed, they are significantly better than SEU only for 18% and 6% of the subjects, respectively, in Treatment 1 and only for 5% and 2% of the subjects, respectively, in Treatment 2.

These statistical tests on the fitting of the various preference functionals tell us that the best seem to be AEU and VEU. Indeed, these preference functionals are significantly better than SEU for 45% and 64% of the subjects, respectively, in Treatment 1, and for 41% and 63% of the subjects, respectively, in Treatment 2 (see Table 6A). However, we prefer to look at predictions rather than statistical tests. We present Table 7 which is the counterpart of Table 5 and gives cumulative rankings. What is particularly striking is that while VEU still seems to be particularly good, the average rankings are much closer together than they were for the BICs. At the same time, there is a very strong correlation between the BICs (corrected log-likelihoods) and the prediction log-likelihoods, as Figure 4 makes clear: this suggests that the best (BIC) fitting preference functional is often also the best predicting log-likelihood, as Table 8 confirms. At the same time, Figure 5 emphasises that there is a very high correlation over preference functionals (for any one subject) and hence there are very small differences between the goodness of prediction of the various preference functionals. But this was also the case for the Bayesian Information Criteria: there is much more variation across subjects (and across Specifications) than across preference functionals. This latter point is emphasised by Table 9 and Figure 7. Table 9 presents the average prediction error (as measured, for any one subject and preference functional, by the square root of the mean squared difference between the actual allocation and the predicted allocation using that preference functional). Figure 7 presents the distribution across subjects of this measure of prediction error. What is noticeable is the big difference across subjects and the small difference across preference functionals.

We conclude that, in terms of predictions, there is not much difference between the preference functionals. One might as well use SEU.

VI. Conclusions

For a theorist, our conclusion that there is not much to choose between preference functionals when it comes to predictions must be disappointing: SEU is no worse than the others. Particularly disappointing is that there does not seem to be a treatment effect with respect to preference functionals: relatively the different preference functionals perform similarly in the two treatments. This indeed seems odd, as it is clear that Treatment 1 is almost a situation of risk, while Treatment 2 is clearly a situation of ambiguity. It seems, that in terms of preference functionals used by the subjects, SEU is as good as any of the other preference functionals.

In the light of the nature of the results, we can legitimately ask: is there no treatment effect? Is the experiment a failure? Fortunately for us we can give an answer to both questions.

In order to explore whether there was any treatment effect, we should look at Table 10.[21] This gives the averages of the estimated parameters for each preference functional for each treatment. Let us just look at SEU, as a similar message emerges (appropriately modified) for the other functionals[22]. The mean estimates of risk aversion are not significantly different in the two treatments; neither are the mean estimates of the noise (though the noise is slightly higher in Treatment 2). But examine the estimated probabilities: in Treatment 1 the mean estimates of the three probabilities (pink, yellow and blue) are

[21] Which is just for Specification 2.

[22] With the other preference functionals we note the following, as far as the mean paramters are concerned:

(1) with CEU the estimated mean capacities are almost additive, but get slightly less so in Treatment 2;

(2) with AEU, the mean lower bounds on the probabilities are close to the SEU subjective probabilities, and get closer to equality in Treatment 2;

(3) with VEU the mean δ parameter is close to 0 and similar in the two treatments;

(4) with COM the mean lower bounds on the probabilities are close to the SEU probabilities and the λ is close to 0.5.

0.241, 0.347 and 0.412; in Treatment 2, 0.255, 0.363 and 0.381; the true probabilities are 0.2, 0.3 and 0.5. There are significant differences between Treatment 1 and Treatment 2: in Treatment 1 the mean estimated probabilities are close to the true ones; in Treatment 2 they are significantly closer to equal probabilities for the three colours. *Subjects were responding to the ambiguity by working on the basis of almost equal probabilities.* So they were not working with a more sophisticated preference functional in the more complicated environment of Treatment 2; on the contrary, they responded by simplifying their decision problem. Perhaps their behaviour could suggest something to theorists?

In order to explore whether the experiment was a failure in terms of its ability to distinguish between the different preference functionals we should comment on the issue of *distinguishability* of the different preference functionals. We have already noted that the preference functionals "are" different (and thus lead to different decisions), which implies that the models are distinguishable: we are simply reinforcing here the theoretical point we have made earlier, at the end of section 1. We return to this point since the theory ignores the existence of "noise" in subjects' decisions. However there is noise in subjects' decisions and this noise could drown out the distinguishability that clearly exists in the absence of noise, unless the number of questions asked to the subjects was sufficiently high. So we carried out a simulation assuming a particular (realistic) level of noise and the questions that we asked. With our implementation and our characterisation of the preference functionals we can demonstrate with a simple example, shown in Table 11, that the preference functionals under analysis are fully distinguishable. In this example we have assumed a reasonable set of parameters for each functional, a particular specification (2) and a realistic value for the precision s.[23], and simulated estimation with 100 repetitions Each cell reports the mean Bayesian Information Criterion (the lower the better) of the column model when the row model is the true one. Table 11 shows that the preference functionals under analysis are fully distinguishable. Indeed, in each row the diagonal element is always a lot smaller than the off-diagonal elements. This means, for example, that if we know that a specific subject has CEU preferences, then CEU best fits behaviour in the experiment. Obviously this is for subjects who are 'clearly' CEU: if a subject has CEU preferences that are 'close' to SEU preferences and there is a lot of noise in that subject's behaviour, then distinguishability is more problematic. But, of course, if that is the case then SEU predictions will also be 'close' to CEU predictions.

This property of the experimental design (distinguishability) was not simply by chance as we carried out intensive pre-experimental simulations in order to select the set of questions to ask the subjects. The purpose of these simulations was precisely to select a number and a set of questions which would enable us to discriminate between the preference functionals, given the amount of noise in the subjects' responses. Clearly the greater the noise the more questions are required; this explains the relatively large number of questions in our experiment. We had carried out a pilot experiment to determine how much

[23]We run 100 replications using a coefficient of risk aversion r equal to 0.8 and a coefficient of precision equal to 12. For each preference funcionals we set the following parameters' values: SEU: $p_1 = 0.2$, $p_2 = 0.3$, $p_3 = 0.5$; CEU: $w_{E(1)} = 0.10$, $w_{E(2)} = 0.20$, $w_{E(3)} = 0.30$, $w_{E(2)\cup E(3)} = 0.85$, $w_{E(1)\cup E(3)} = 0.75$, $w_{E(2)\cup E(2)} = 0.65$; AEU: $\underline{p}_1 = 0.10$, $\underline{p}_2 = 0.15$, $\underline{p}_3 = 0.25$, $\alpha = 0.5$; VEU: $p_1 = 0.2$, $p_2 = 0.3$, $p_3 = 0.5$, $\delta = 0.10$; COM: $\underline{p}_1 = 0.10$, $\underline{p}_2 = 0.20$, $\underline{p}_3 = 0.30$, $\lambda = 0.75$.

noise there was in behaviour, and this informed the simulation and hence the number and choice of questions.

Before we conclude we should note that the fitting and prediction parts of the exercise give somewhat different results. In terms of *fitting*, some of the more general models (particularly AEU and VEU) do seem to fit better than SEU for many of the subjects. However, when it comes to *prediction*, as is clearly shown in Table 9, much of this superiority disappears (though AEU and VEU are still marginally better overall than SEU): *hence one does not lose a lot in predictive power in using SEU rather than one of the more general functionals*. An econometrician might regard this as an inevitable consequence of over-fitting: as is well-known, if one fits an *n*th-degree polynomial to *n* observations from a truly linear relationship (with noise), the fit is better than a linear fit, but extrapolative predictions are almost certainly worse. It should be noted, however, that our prediction questions were a randomly chosen subset of all the questions, and cannot be considered as extrapolative. Indeed it is not clear in our case what "extrapolative questions" means.

So the bottom line appears to be that SEU performs as well as some of the recent, more sophisticated models of behaviour: when we move from Treatment 1 (almost a case of risk) to Treatment 2 (clearly a situation of ambiguity) subjects do not respond by moving to a more sophisticated preference functional. Instead they seem to respond by using SEU with subjective probabilities further away from the true probabilities and nearer to equality. In our view, this is a rational response: if a situation is ambiguous, and hence complicated, why complicate it further by using a more complicated preference functional?

References

Abdellaoui, M., Baillon, A., Placido, L. and P. Wakker (2011), “The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation”, *American Economic Review*, 101: 695-723.

Ahn, D.S., Choi, S., Gale, D., and S. Kariv (2010), “Estimating Ambiguity Aversion in a Portfolio Choice Experiment”, Working Paper.

Anderson, S., Fountain, J., Harrison, G.W. and E.E. Rutström (2009), “Estimating aversion to uncertainty”, Working Paper.

Andreoni, J., and J. Miller (2002), “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism”, *Econometrica*, 70, 2: 737-753.

Blavatskyy, P. (2011), “A Model of Probabilistic Choice Satisfying First-Order Stochastic Dominance”, *Management Science*, 57, 3: 542-548.

Camerer, C., and M. Weber (1992), “Recent Development in Modeling Preferences: Uncertainty and Ambiguity”, *Journal of Risk and Uncertainty*, 5,4: 325-370.

Camerer, C. (1995), “Individual Decision Making” in J. Kagel and A. Roth (Eds.), *Handbook of Experimental Economics*, Princeton University Press, 587-703.

Chambers, R.G., Melkonyan, T. and D. Pick (2010), “Experimental Evidence on Multiple Prior Models in the Presence of Uncertainty”, Working Paper.

Choi, S., Fisman R., Gale, D., and S. Kariv (2007), "Consistency and Heterogeneity of Individual Behavior under Uncertainty", *American Economic Review*, 97, 4: 1921-1938.

Clarke, K.A. (2007), "A Simple Distribution-Free Test for Non-nested Model Selection", *Political Analysis*, 15: 347-363.

Ellsberg, D. (1961), "Risk, Ambiguity and the Savage Axiom", *Quarterly Journal of Economics*, 75: 643-669.

Etner, J., Jeleva, M., and J.M. Tallon (2012), "Decision Theory under Ambiguity", *Journal of Economic Surveys*, 26, 2: 234-270.

Fox, C.R. and A. Tvesky (1995), "Ambiguity Aversion and Comparative Ignorance", *Quarterly Journal of Economics*, 110, 3: 585-603.

Gajdos, T., Hayashi T., Tallon J.M. and J.C. Vergnaud (2008), "Attitude toward imprecise information ", *Journal of Economic Theory*, 140, 27-65.

Ghirardato, P., Maccheroni F., and M. Marinacci (2004), "Differentiating Ambiguity and Ambiguity Attitude", *Journal of Economic Theory*, 118, 2: 133-173.

Gilboa, I. and D. Schmeidler (1989), "Maxmin Expected Utility with Non-Unique Prior", *Journal of Mathematical Economics*, 18: 141-153.

Greiner, B (2004), "The Online Recruitment System ORSEE 2.0 – A Guide for the Organization of Experiments in Economics", University of Cologne Discussion Paper (www.orsee.org).

Halevy, Y. (2007), "Ellsberg Revisited: An Experimental Study", *Econometrica*, 75, 2: 503-536.

Hey, J.D., Lotito G. and A. Maffioletti (2010), "The Descriptive and Predictive Adequacy of Theories of Decision Making Under Uncertainty/Ambiguity", *Journal of Risk and Uncertainty*, 41, 2: 81-111.

Holt, C.A. and S.K. Laury (2002), "Risk aversion and incentive effects", American Economic Review, 92, 5: 1644-1655.

Karni, E., and Z. Safra (1987), "Preference reversals and the observability of preferences by experimental methods", *Econometrica*, 55: 675-685.

Loomes, G. (1991), "Evidence of a new violation of the independence axiom", *Journal of Risk and Uncertainty*, 4, 1: 91-108.

Maccheroni, F., Marinacci, M. and A. Rustichini (2006), "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences", *Econometrica*, 74: 1447-1498.

Morone, A., and O. Ozdemir (2011), "Displaying uncertainty information about probability: experimental evidence", *Bulletin of Economic Research*, Forthcoming. DOI: 10.1111/j.1467-8586.2010.00380.x

Neuberg, J. (1886), "Sur le point de Steiner", *Journal de mathématiques spéciales*: 29.

Schmeidler, D. (1989), "Subjective Probability and Expected Utility without Additivity", *Econometrica*, 57, 3: 571-587.

Siniscalchi, M. (2009), "Vector Expected Utility and Attitudes Toward Variation", *Econometrica*, 77, 3: 801-855.

Wilcox, N. (2007), “Predicting Risky Choices Out-of-Context: A Monte Carlo Study”, University of Houston Working Paper.

Wilcox, N. (2008), “Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison”, In J. C. Cox and G. W. Harrison, eds., *Research in Experimental Economics Vol. 12: Risk Aversion in Experiments* pp. 197-292. Bingley, UK: Emerald.

Wilcox, N. (2011), "Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk", *Journal of Econometrics*, 162: 89-104.

Tables and Figures

Uncorrected log-likelihoods						
Specification 1	SEU	CEU	AEU	VEU	COM	Obs.
All	-176.94	-168.94	-171.90	-175.66	-171.01	129
	(43.83)	(44.75)	(43.52)	(44.21)	(42.83)	
Treatment 1	-179.43	-170.99	-174.12	-178.16	-173.24	66
	(38.48)	(40.06)	(38.84)	(38.92)	(38.30)	
Treatment 2	-174.33	-166.79	-169.57	-173.04	-168.68	63
	(48.99)	(49.42)	(48.15)	(49.33)	(47.32)	
t-stat diff.	-0.656	-0.529	-0.589	-0.652	-0.600	
Specification 2	SEU	CEU	AEU	VEU	COM	Obs.
All	-168.92	-166.55	-165.56	-163.02	-168.27	129
	(57.21)	(56.79)	(56.93)	(56.59)	(57.15)	
Treatment 1	-171.54	-168.77	-168.11	-165.72	-170.71	66
	(48.93)	(48.59)	(48.99)	(48.78)	(48.91)	
Treatment 2	-166.19	-164.22	-162.88	-160.19	-165.72	63
	(65.06)	(64.60)	(64.50)	(64.03)	(64.97)	
t-stat diff.	-0.526	-0.450	-0.516	-0.550	-0.491	
Specification 3	SEU	CEU	AEU	VEU	COM	Obs.
All	-183.15	-179.13	-181.05	-179.52	-169.16	129
	(34.32)	(33.29)	(33.76)	(33.41)	(43.08)	
Treatment 1	-182.76	-178.59	-180.57	-179.32	-170.48	66
	(32.43)	(31.54)	(31.94)	(31.73)	(39.55)	
Treatment 2	-183.57	-179.69	-181.55	-179.73	-167.78	63
	(36.46)	(35.28)	(35.81)	(35.33)	(46.77)	
t-stat diff.	0.134	0.186	0.163	0.069	-0.352	

TABLE 1—AVERAGE FITTED LOG-LIKELIHOODS (STANDARD DEVIATION IS IN PARENTHESIS)

Bayesian Information Criterion						
Specification 1	SEU	CEU	AEU	VEU	COM	Obs.
All	370.26	370.62	368.36	371.79	366.59	129
	(87.66)	(89.49)	(87.04)	(88.42)	(85.67)	
Treatment 1	375.24	374.73	372.80	376.79	371.04	66
	(76.95)	(80.11)	(77.67)	(77.84)	(76.60)	
Treatment 2	365.03	366.33	363.70	366.56	361.93	63
	(97.99)	(98.84)	(96.29)	(98.66)	(94.65)	
t-stat diff.	0.656	0.529	0.589	0.652	0.600	
Specification 2	SEU	CEU	AEU	VEU	COM	Obs.
All	354.23	365.86	355.68	346.51	361.11	129
	(114.43)	(113.59)	(113.85)	(113.17)	(114.29)	
Treatment 1	359.45	370.30	360.78	351.91	365.99	66
	(97.87)	(97.18)	(97.97)	(97.57)	(97.81)	
Treatment 2	348.75	361.20	350.33	340.85	356.00	63
	(130.12)	(129.20)	(129.00)	(128.06)	(129.95)	
t-stat diff.	0.526	0.450	0.516	0.550	0.491	
Specification 3	SEU	CEU	AEU	VEU	COM	Obs.
All	382.69	391.01	386.67	379.52	362.89	129
	(68.64)	(66.58)	(67.51)	(66.82)	(86.15)	
Treatment 1	381.89	389.94	385.71	379.12	365.52	66
	(64.86)	(63.08)	(63.88)	(63.47)	(79.09)	
Treatment 2	383.52	392.14	387.67	379.93	360.13	63
	(72.91)	(70.56)	(71.63)	(70.67)	(93.55)	
t-stat diff.	-0.133	-0.186	-0.163	-0.069	0.352	

TABLE 2—BAYESIAN INFORMATION CRITERION (STANDARD DEVIATION IS IN PARENTHESIS)

	Treatment 1																
	Model	1st	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-10	1-11	1-12	1-13	1-14	1-15	Aver. Rank
Spec. 1	SEU	8	14	20	23	27	39	42	52	52	61	67	73	82	98	100	8.44
	CEU	8	8	12	15	23	33	33	33	36	39	42	55	67	73	100	10.23
	AEU	2	8	15	20	29	36	38	42	44	53	65	85	91	100	100	8.73
	VEU	9	21	23	26	29	32	39	44	45	53	59	61	73	82	100	9.05
	COM	6	11	21	32	35	39	45	52	56	62	70	83	95	100	100	7.92
Spec. 2	SEU	23	29	53	58	59	59	67	79	86	88	91	97	100	100	100	5.12
	CEU	0	2	8	17	39	42	45	47	48	55	61	65	76	80	100	9.15
	AEU	3	24	36	47	52	61	64	67	73	85	92	94	97	100	100	6.06
	VEU	36	55	62	65	70	73	85	92	95	97	98	100	100	100	100	3.71
	COM	0	2	8	29	41	47	52	55	70	77	86	88	91	98	100	7.58
Spec. 3	SEU	2	5	9	14	18	36	45	62	68	73	74	94	100	100	100	8.00
	CEU	0	0	2	3	5	9	12	15	30	39	50	52	58	74	100	11.52
	AEU	0	5	6	11	14	18	29	35	47	53	64	67	73	94	100	9.86
	VEU	5	14	17	21	26	36	55	65	71	77	89	94	100	100	100	7.30
	COM	0	6	9	21	35	38	48	61	77	88	91	94	98	100	100	7.33

	Treatment 2																
	Model	1st	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-10	1-11	1-12	1-13	1-14	1-15	Aver. Rank
Spec 1	SEU	10	14	22	22	30	41	46	49	57	62	70	78	79	100	100	8.19
	CEU	10	11	16	16	16	29	30	40	43	48	54	62	75	78	100	9.75
	AEU	2	6	11	22	27	35	38	46	52	59	67	89	97	100	100	8.49
	VEU	6	17	22	24	25	33	38	46	52	57	62	70	79	83	100	8.84
	COM	3	8	17	25	41	46	52	60	62	75	83	83	92	100	100	7.52
Spec 2	SEU	22	35	51	56	60	60	68	79	89	94	97	98	100	100	100	4.90
	CEU	0	0	3	5	40	41	46	46	48	52	57	59	73	84	100	9.46
	AEU	0	19	49	54	63	65	67	68	75	84	87	94	98	98	100	5.78
	VEU	41	67	67	70	70	70	84	92	92	95	98	100	100	100	100	3.54
	COM	0	0	2	43	48	52	54	54	65	71	81	86	89	98	100	7.57
Spec 3	SEU	5	6	8	13	14	24	41	49	63	71	78	95	100	100	100	8.32
	CEU	0	0	0	2	2	3	5	11	17	32	40	46	52	68	100	12.22
	AEU	0	0	2	8	10	16	24	33	40	48	52	59	68	90	100	10.51
	VEU	2	13	21	24	25	37	51	62	68	71	83	89	100	100	100	7.56
	COM	0	3	10	17	29	48	56	63	76	81	92	94	97	100	100	7.35

TABLE 3—RANKINGS BASED ON THE BAYESIAN INFORMATION CRITERION (ALL VALUES REPRESENT CUMULATIVE PERCENTAGES- LAST COLUMN SHOWS THE AVERAGE RANKING)

Prediction Log-Likelihoods						
Specification 1	SEU	CEU	AEU	VEU	COM	Obs.
All	-49.19	-50.83	-49.16	-49.67	-48.68	127
	(15.84)	(21.28)	(16.51)	(18.97)	(16.53)	
Treatment 1	-49.09	-49.46	-48.92	-49.35	-48.45	65
	(11.20)	(11.76)	(11.79)	(11.68)	(11.56)	
Treatment 2	-49.31	-52.26	-49.41	-50.02	-48.91	62
	(19.66)	(28.04)	(20.42)	(24.49)	(20.59)	
t-stat diff.	0.076	0.728	0.167	0.195	0.154	
Specification 2	SEU	CEU	AEU	VEU	COM	Obs.
All	-46.70	-47.69	-46.38	-45.74	-46.61	127
	(17.99)	(18.96)	(18.02)	(18.08)	(18.09)	
Treatment 1	-47.41	-48.24	-47.07	-46.57	-47.36	65
	(14.89)	(15.93)	(14.97)	(15.05)	(15.15)	
Treatment 2	-45.96	-47.11	-45.64	-44.86	-45.84	62
	(20.85)	(21.82)	(20.85)	(20.88)	(20.84)	
t-stat diff.	-0.278	-0.194	-0.274	-0.326	-0.288	
Specification 3	SEU	CEU	AEU	VEU	COM	Obs.
All	-50.77	-51.31	-50.86	-50.62	-50.78	127
	(11.69)	(12.99)	(11.81)	(12.71)	(11.89)	
Treatment 1	-50.22	-50.41	-50.26	-49.80	-50.18	65
	(9.30)	(9.79)	(9.56)	(9.63)	(9.41)	
Treatment 2	-51.36	-52.26	-51.51	-51.48	-51.40	62
	(13.83)	(15.70)	(13.87)	(15.33)	(14.09)	
t-stat diff.	0.517	0.681	0.557	0.642	0.531	

TABLE 4—AVERAGE PREDICTION LOG-LIKELIHOODS (STANDARD DEVIATION IS IN PARENTHESIS)

Treatment 1																	
	Model	1st	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-10	1-11	1-12	1-13	1-14	1-15	Aver. Rank
Spec. 1	SEU	6	11	18	21	27	32	35	45	50	61	68	71	77	91	100	8.86
	CEU	8	18	21	29	41	45	52	55	58	67	74	80	88	91	100	7.74
	AEU	6	9	17	24	33	39	47	53	59	70	74	80	85	92	100	8.11
	VEU	2	12	18	27	27	33	44	52	61	64	65	67	77	89	100	8.62
	COM	14	18	30	39	42	52	62	68	71	73	76	85	89	94	100	6.86
Spec. 2	SEU	5	12	17	35	39	41	45	50	56	68	79	86	92	98	100	7.76
	CEU	6	12	20	21	41	45	52	59	64	70	76	79	89	89	100	7.77
	AEU	3	23	33	42	45	47	55	61	70	74	79	88	92	95	100	6.92
	VEU	20	26	35	42	50	65	68	73	76	79	86	89	91	100	100	6.00
	COM	9	17	29	35	42	50	53	58	70	74	79	85	92	97	100	7.11
Spec. 3	SEU	5	9	11	15	20	29	36	45	52	59	64	68	80	88	100	9.20
	CEU	8	8	11	14	20	32	42	48	55	61	73	82	85	86	100	8.77
	AEU	0	8	14	20	24	27	33	41	47	59	64	73	86	94	100	9.11
	VEU	8	15	21	23	27	33	42	53	59	65	77	88	92	97	100	7.99
	COM	3	3	6	12	20	29	33	39	55	58	67	79	82	97	100	9.18

Treatment 2																	
	Model	1st	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-10	1-11	1-12	1-13	1-14	1-15	Aver. Rank
Spec 1	SEU	3	8	11	16	25	25	33	41	54	62	71	81	90	98	100	8.79
	CEU	8	16	17	21	29	41	44	48	57	68	70	71	78	78	100	8.54
	AEU	5	11	19	24	29	43	46	51	62	67	78	86	87	100	100	7.94
	VEU	10	14	19	24	32	33	43	48	57	63	68	78	83	90	100	8.38
	COM	6	17	29	33	41	46	49	57	63	71	76	79	94	97	100	7.40
Spec 2	SEU	5	16	30	46	52	52	56	59	68	75	84	89	94	100	100	6.75
	CEU	10	10	21	24	41	48	54	57	59	68	73	78	79	84	100	7.95
	AEU	2	27	33	49	57	57	68	76	79	83	86	87	90	97	100	6.08
	VEU	32	41	48	60	67	79	83	84	86	86	89	89	95	100	100	4.62
	COM	11	17	33	46	52	59	62	70	78	83	86	95	97	98	100	6.13
Spec 3	SEU	0	3	8	11	14	17	33	44	48	52	62	73	81	89	100	9.64
	CEU	2	2	3	5	6	21	27	35	40	51	63	73	83	87	100	10.03
	AEU	2	3	6	11	16	24	31	44	48	56	66	73	84	98	100	9.37
	VEU	2	3	6	11	16	24	31	44	48	56	66	73	84	98	100	9.37
	COM	2	2	3	10	13	24	33	43	49	56	60	68	79	89	100	9.70

TABLE 5—RANKINGS BASED ON PREDICTION LOG-LIKELIHOODS (ALL VALUES REPRESENT CUMULATIVE PERCENTAGES- LAST COLUMN SHOWS THE AVERAGE RANKING)

Treatment 1					
Model	SEU	CEU	AEU	VEU	COM
CEU	18	n.a.	0	0	0
AEU	45	0	n.a.	0	0
VEU	64	0	5	n.a.	0
COM	6	0	5	0	n.a.
Treatment 2					
Model	SEU	CEU	AEU	VEU	COM
CEU	5	n.a.	2	0	0
AEU	41	0	n.a.	0	0
VEU	63	0	13	n.a.	2
COM	2	0	0	0	n.a.

TABLE 6—A) SIGNIFICANCE TEST FOR SUPERIORITY OF PREFERENCE FUNCTIONALS - 5 PERCENT LEVEL

Treatment 1					
Model	SEU	CEU	AEU	VEU	COM
CEU	8	n.a.	0	0	0
AEU	38	0	n.a.	0	0
VEU	52	0	2	n.a.	0
COM	2	0	0	0	n.a.
Treatment 2					
Model	SEU	CEU	AEU	VEU	COM
CEU	2	n.a.	0	0	0
AEU	33	0	n.a.	0	0
VEU	56	0	6	n.a.	2
COM	0	0	0	0	n.a.

TABLE 6 — B) SIGNIFICANCE TEST FOR SUPERIORITY OF PREFERENCE FUNCTIONALS - ONE PERCENT LEVEL

Treatment 1						
Model	1st	2nd	3rd	4th	5th	Average Ranking
SEU	9	26	44	85	100	3.36
CEU	17	29	48	58	100	3.48
AEU	8	52	67	85	100	2.89
VEU	50	58	68	82	100	2.42
COM	17	36	73	91	100	2.83
Treatment 2						
Model	1st	2nd	3rd	4th	5th	Average Ranking
SEU	8	22	51	89	100	3.30
CEU	16	22	37	38	100	3.87
AEU	2	46	63	87	100	3.02
VEU	59	71	78	94	100	1.98
COM	16	38	71	92	100	2.83

TABLE 7—RANKINGS BY PREDICTION LOG-LIKELIHOODS WITHIN SPECIFICATION 2 (ALL VALUES REPRESENT CUMULATIVE PERCENTAGE- THE LAST COLUMN SHOWS THE AVERAGE RANKING)

Treatment 1					
	Bayesian Information Criterion				
Prediction Log-Likelihoods	SEU	CEU	AEU	VEU	COM
SEU	-0.821	-0.820	-0.818	-0.814	-0.819
CEU	-0.770	-0.767	-0.768	-0.766	-0.768
AEU	-0.815	-0.813	-0.813	-0.810	-0.813
VEU	-0.804	-0.802	-0.805	-0.807	-0.802
COM	-0.808	-0.807	-0.806	-0.801	-0.806
Treatment 2					
	Bayesian Information Criterion				
Prediction Log-Likelihoods	SEU	CEU	AEU	VEU	COM
SEU	-0.939	-0.938	-0.939	-0.935	-0.933
CEU	-0.918	-0.915	-0.918	-0.915	-0.912
AEU	-0.929	-0.927	-0.929	-0.926	-0.923
VEU	-0.921	-0.919	-0.923	-0.921	-0.915
COM	-0.939	-0.937	-0.938	-0.934	-0.933

TABLE 8—CORRELATIONS BETWEEN PREDICTION LOG-LIKELIHOODS AND BICS- SPECIFICATION 2

	SEU	CEU	AEU	VEU	COM
All	8.05	8.25	8.04	7.91	8.04
	(4.00)	(4.17)	(3.95)	(3.95)	(4.00)
Treatment 1	7.05	7.21	7.10	6.99	7.03
	(3.29)	(3.48)	(3.47)	(3.40)	(3.29)
Treatment 2	9.10	9.35	9.02	8.87	9.10
	(4.41)	(4.57)	(4.21)	(4.27)	(4.41)

TABLE 9—AVERAGE DEPARTURE FROM BEST PREDICTION (STANDARD DEVIATION IS IN PARENTHESES)

Specification 2					
Theory	Parameter	All	Treatment1	Treatment2	t-test diff.
SEU	p_1	0.248	0.241	0.255	-1.422
	p_2	0.355	0.347	0.363	-1.955
	p_3	0.397	0.412	0.381	3.129
	r	0.801	0.815	0.788	0.655
	s	11.821	12.846	10.747	1.145
CEU	$w_{E(1)}$	0.249	0.235	0.263	-2.177
	$w_{E(2)}$	0.354	0.345	0.364	-1.990
	$w_{E(3)}$	0.381	0.396	0.365	2.785
	$w_{E(2)\cup E(3)}$	0.764	0.761	0.766	-0.242
	$w_{E(1)\cup E(3)}$	0.611	0.622	0.601	1.459
	$w_{E(1)\cup E(2)}$	0.604	0.568	0.642	-3.115
	r	0.803	0.818	0.787	0.746
	s	12.536	13.762	11.252	1.303
AEU	$\underline{p}_1$	0.230	0.222	0.239	-1.723
	$\underline{p}_2$	0.334	0.325	0.344	-1.920
	$\underline{p}_3$	0.375	0.388	0.361	2.536
	α	0.220	0.241	0.198	0.862
	r	0.790	0.807	0.772	0.820
	s	12.709	13.836	11.528	1.150
VEU	p_1	0.248	0.241	0.256	-1.433
	p_2	0.354	0.346	0.363	-1.984
	p_3	0.397	0.413	0.381	3.206
	δ	-0.039	-0.033	-0.045	1.398
	r	0.786	0.804	0.767	0.847
	s	13.178	14.187	12.120	0.982
COM	$\underline{p}_1$	0.242	0.232	0.252	-1.805
	$\underline{p}_2$	0.346	0.335	0.359	-2.912
	$\underline{p}_3$	0.389	0.401	0.378	2.309
	λ	0.517	0.524	0.509	2.055
	r	0.803	0.817	0.788	2.055
	s	11.935	13.063	10.753	1.238

TABLE 10—DESCRIPTIVE STATISTICS AVERAGE ESTIMATED PARAMETERS - SPECIFICATION 2

	Bayesian Information Criterion				
True Model	SEU	CEU	AEU	VEU	COM
SEU	471.21	484.64	479.03	474.46	479.40
CEU	529.03	440.82	537.69	449.09	530.93
AEU	561.17	496.39	469.24	514.19	569.83
VEU	487.69	456.30	496.35	446.31	478.52
COM	485.44	471.19	493.40	475.53	465.80

TABLE 11—BAYESIAN INFORMATION CRITERION WHEN THE ROW MODEL IS THE TRUE ONE

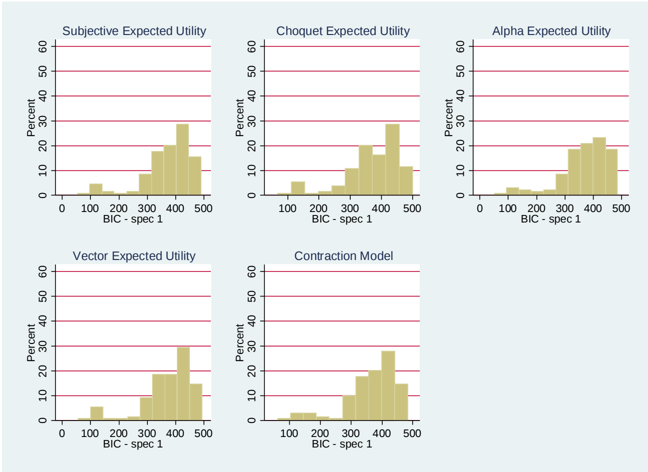

FIGURE 1. HISTOGRAMS BIC BY PREFERENCE FUNCTIONALS - SPECIFICATION 1

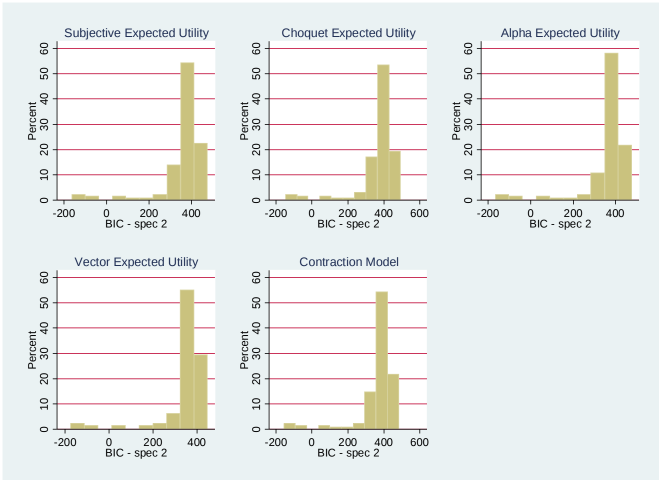

FIGURE 2. HISTOGRAMS BIC BY PREFERENCE FUNCTIONALS - SPECIFICATION 2

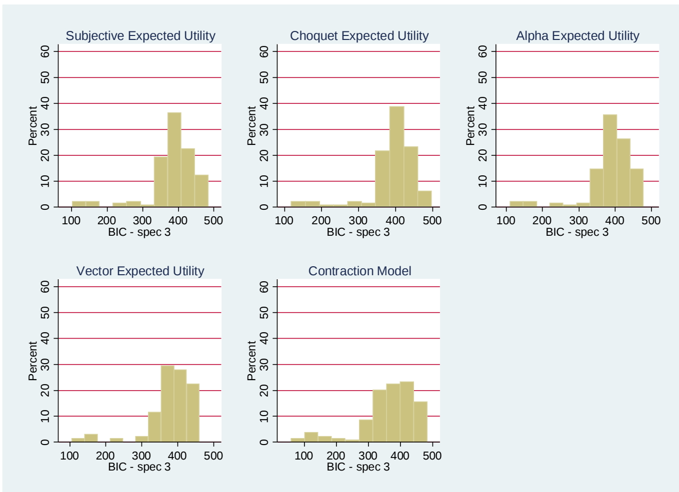

FIGURE 3. HISTOGRAMS BIC BY PREFERENCE FUNCTIONALS - SPECIFICATION 3

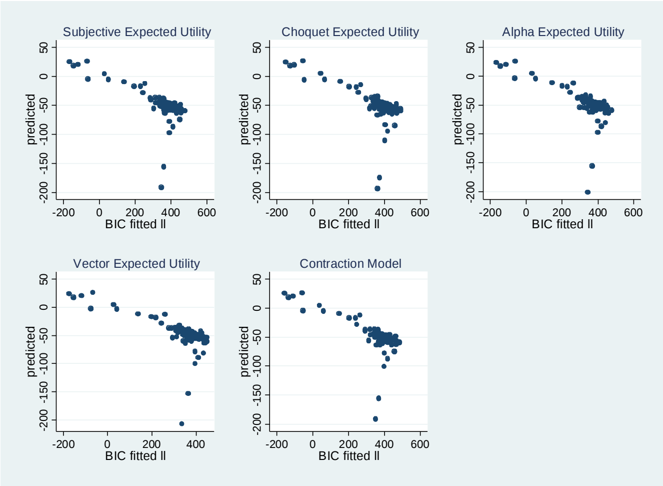

FIGURE 4. SCATTER PREDICTION LOG-LIKELIHOODS VS BICS - SPECIFICATION 2

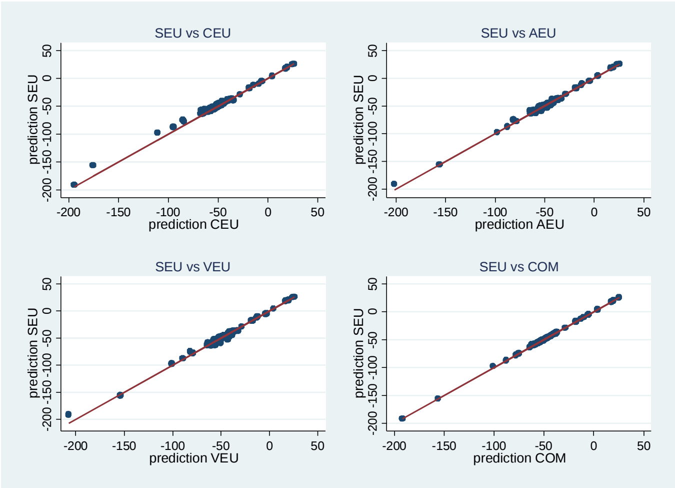

FIGURE 5. SCATTER PREDICTION LOG-LIKELIHOODS SEU VS OTHER PREFERENCE FUNCTIONALS – SPECIFICATION 2

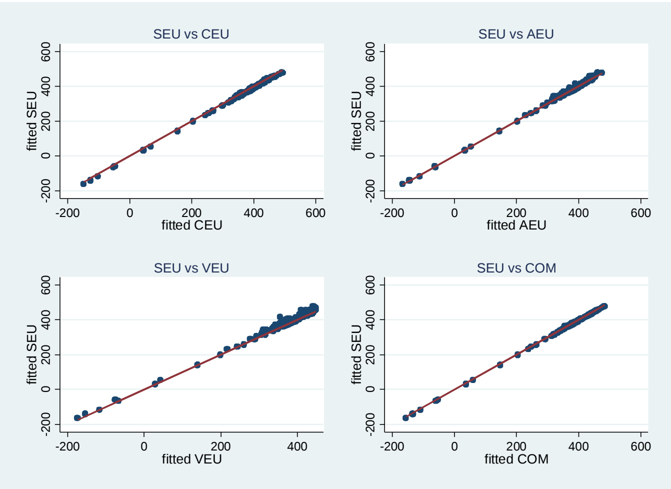

FIGURE 6. SCATTER BICS SEU VS OTHER PREFERENCE FUNCTIONALS - SPECIFICATION 2

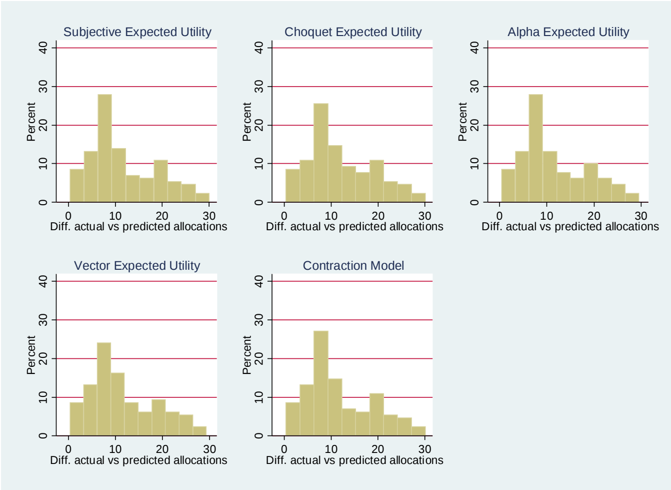

FIGURE 7. DIFFERENCES ACTUAL VS PREDICTED ALLOCATIONS - SPECIFICATION 2