

# Assessing Multiple Prior Models of Behaviour under Ambiguity

Anna Conte<sup>1</sup> and John D. Hey<sup>2</sup>

## Abstract

The recent spate of theoretical models of behaviour under ambiguity can be partitioned into two sets: those involving multiple priors (in which the probabilities of the various events are not known but probabilities can be attached to the various possible values for the probabilities) and those not involving multiple priors. This paper concentrates on the first set and provides an experimental investigation into recently proposed theories. Using an appropriate experimental interface, in which the probabilities on the various possibilities are explicitly stated, we examine the fitted and predictive power of the various theories. We first estimate subject-by-subject, and then we estimate and predict using a mixture model over the contending theories. The individual estimates suggest that 24% of our 149 subjects have behaviour consistent with Expected Utility, 56% with the Smooth Model (of Klibanoff *et al*, 2005), 11% with Rank Dependent Expected Utility and 9% with the Alpha Model (of Ghirardato *et al* 2004); these figures are close to the mixing proportions obtained from the mixture estimates. If we classify our subjects through the posterior probabilities (given all the evidence) of each of them being of the various types: using the estimates we get 25%, 50%, 20% and 5% (for EU, Smooth, Rank Dependent and Alpha); while using the predictions 22%, 53%, 22% and 3%. Interestingly the Smooth model seems to fare the best.

**JEL numbers:** D81, C91, C23

**Keywords:** Alpha Model, Ambiguity, Expected Utility, Mixture Models, Rank Dependent Expected Utility, Smooth Model.

<sup>1</sup>University of Westminster, London, UK, and the Max-Planck-Institute of Economics, Jena, Germany.

<sup>2</sup>University of York, UK.

## 1. Introduction

Ambiguity differs from risk in that, under ambiguity, events are not certain but probabilities are not known. Decision-making under ambiguity is, in principle therefore, different from decision-making under risk. Over the past few years, and intensively so recently, theorists have proposed many new theories of behaviour under ambiguity. The purpose of this paper is to report on the empirical adequacy of a subset of these new theories. The subset that we concentrate on here arises naturally because of a partition of the literature. Part of the literature envisages the decision-maker as realising that he or she does not know the values of the relevant probabilities (even when they exist), or is not prepared to make subjective judgements about the possible probabilities; while another part envisages the decision-maker as being able to list the various possibilities for the various probabilities and, moreover, as being able to attach probabilities to the various possibilities. If you like, it is second-order probabilities, or probabilities of probabilities. It is this second subset that we investigate here<sup>1</sup>.

We compare the comparative empirical performance of four theories of decision-making under ambiguity which use a multiple prior approach. In order to provide a fair evaluation of these theories, we use an experimental interface which reproduces exactly a situation of 'probabilities of probabilities'. Moreover, in order to avoid identification issues concerning the correct estimation of an underlying utility function (all four theories embody such a function), our experimental design involves just two final outcomes for the subject; we normalise on these two and hence do not need any functional form for the utility function. Subjects in the experiment were given a show-up fee of €7.50 and could add to that another €40 if the ambiguity resolved itself in their favour.

Subjects were presented with a total of 49 tasks. Each task started off with a choice between two two-stage lotteries. Each two-stage lottery consisted of a set of one-stage lotteries. A one-stage lottery was composed of a known number of red balls and a known number of blue balls. At the end of the experiment a one-stage lottery was played out: this was done by picking one of the balls at

---

<sup>1</sup> We should note that we share the doubts of those who wonder whether this is the appropriate characterisation of a situation of ambiguity – though, at the end of the day, this is an empirical issue.

random and determining the colour of the drawn ball. Depending on which colour the subject had designated earlier as their 'winning colour', they either got €40 in addition to their show-up fee or nothing. In each of the 49 tasks, one of the two two-stage lotteries was designated by the experimenter the "unchanging lottery" and the other the "changing lottery". After indicating which they preferred, one of the one-stage lotteries in the changing two-stage lottery was chosen at random and removed from the two-stage lottery; subjects were then asked to choose again, and this procedure continued until the changing two-stage lottery was reduced to a one-stage lottery. After all 49 tasks were completed, one of them was chosen at random; one of the stages in playing out that task was chosen at random; the subject's choice recovered from the computer; and the chosen two-stage lottery was played out in the obvious fashion; resulting in a payment of €40 or of nothing to the subject.

The experimental procedure made clear that the winning colour is subject to a second-order probability. Thus we have a direct test of the various theories. We give details of the theories that we are investigating in the next section, restricted to the relevant situation in which the final payoff is one of two possibilities. We then give more detail of the experimental procedure in the following section, and we then present our econometric results, first estimating on a subject-by-subject basis before estimating and predicting using a mixture model. We then conclude.

## **1. Theories under investigation**

We consider those theories for which we can explicitly derive a preference functional in our experimental context. Remember that the context is that of probabilities of probabilities. Clearly Expected Utility theory is one of these. We consider also the Smooth Model of Klibanoff *et al* (2005), the well-known Rank Dependent Expected Utility model, and the Alpha Model of Ghirardato *et al* (2004), which is a generalisation of the Maximin Expected Utility model of Gilboa and Schmeidler (1989). Before we can specify these models, we need to describe the context of our experiment and introduce some notation.

As we have already noted, our experimental set up was such that the final payoff (over and above the show-up fee) was either €40 or nothing. We normalise the utility function (present in all theories) on these two values and hence put  $u(€40)=1$  and  $u(€0)=0$ . All the theories we investigate envisage the choice between any two lotteries as being dependent on the difference between the evaluations of the two lotteries. We ignore for the time being the issue of error. We now specify the evaluations of an arbitrary two-stage lottery under each of the theories we are examining.

We start with some notational definitions, first defining a generic one-stage lottery of the type used in the experiment, and which we denote by  $O(m,M)$ . This has  $M$  balls inside it, of which  $m$  ( $0 \leq m \leq M$ ) are of the 'winning colour' (as chosen by the subject), and  $M-m$  of the other (non-winning) colour. Each of these  $M$  balls is equally likely to be drawn if this one-stage lottery is to be played out, so that the probability of drawing a winning ball out of the one stage lottery  $O(m,M)$  is  $m/M$ .

We now define a generic two-stage lottery of the type used in the experiment, and which we denote by  $T(m_1, m_2, \dots, m_N; M)$ , where the  $m_n$  are distinct integers. We write this in such a way that  $m_1 < m_2 < \dots < m_N$ . This consists of  $N$  one-stage lotteries, each of which has  $M$  balls in them. The  $n$ 'th of these has  $m_n$  winning balls and  $M-m_n$  non-winning balls in it. So the generic two-stage lottery consists of  $N$  one-stage lotteries as follows:  $O(m_1, M), O(m_2, M), \dots, O(m_N, M)$ . Each of these  $N$  one-stage lotteries is equally likely to be drawn if this two-stage lottery is to be played out, so that the probability of drawing the one-stage lottery  $O(m_n, M)$  is  $1/N$ . As we have already noted, if  $O(m_n, M)$  is the one-stage lottery drawn, then the probability of drawing a winning ball from that is  $m_n/M$ .

We can now explain how each of the theories we are considering evaluates the generic one-stage and the generic two-stage lotteries. Let  $V_{PF}[O(m, M)]$  and  $V_{PF}[T(m_1, m_2, \dots, m_N; M)]$  denote these valuations respectively for preference functional  $PF$ .

## 2.1 Expected Utility Theory (EU)

Since there are  $m$  winning balls, each of which leads to utility 1, and  $(M-m)$  non-winning balls, each of which leads to utility 0, and all balls are equally likely to be drawn, the expected utility of the generic one-stage lottery is simply given by

$$(1) \quad V_{EU}[O(m, M)] = \frac{m}{M}$$

Since EU is linear in the probabilities, and each of the  $N$  one-stage lotteries by which it is formed are equally likely, it follows that the valuation of the generic two-stage lottery is

$$(2) \quad V_{EU}[T(m_1, m_2, \dots, m_N; M)] = \left[ \frac{m_1}{M} + \frac{m_2}{M} + \dots + \frac{m_N}{M} \right] \frac{1}{N}$$

## 2.2 Smooth Model (SM)

This model, proposed by Klibanoff *et al* (2005), is quintessentially a multiple prior model<sup>2</sup>. We describe its application in the context of the experiment that we have conducted. Essentially, in evaluating any two-stage lottery, this model proceeds by taking the Expected Value (with respect to the lotteries composing a two-stage lottery) of some function of the Expected Utility of each of the one-stage lotteries from which the two-stage lottery is composed. Denoting this function by  $\phi(\cdot)$  and noting that the one-stage lottery  $O(m_n, M)$  has expected utility  $m_n/M$  (as in the above) and hence has the same value as in EU, it follows that the valuation by the Smooth Model of the generic two-stage lottery is given by

$$(3) \quad V_{SM}[T(m_1, m_2, \dots, m_N; M)] = \left[ \phi\left(\frac{m_1}{M}\right) + \phi\left(\frac{m_2}{M}\right) + \dots + \phi\left(\frac{m_N}{M}\right) \right] \frac{1}{N}$$

It remains to specify the function  $\phi(\cdot)$ . Klibanoff *et al* suggested the particular form  $\phi(x) = -\frac{e^{-\alpha x}}{\alpha}$ .

However, we prefer a specification for which a particular parameter value reduces the Smooth

---

<sup>2</sup> In the general form of the theory, the decision-makers themselves are supposed to specify the set of possible probabilities and the probabilities attached to them; in our experiment these were objectively specified and we assume that the subjective set and the subjective probabilities are the same as the objective ones.

model to EU; this is the case when  $\phi(x) = x$  for all  $x$ . So one way of characterising the  $\phi(\cdot)$  function is to put

$$(4) \quad \phi(x) = \frac{1 - e^{-sx}}{1 - e^{-s}}$$

Note that when  $s \rightarrow 0$  the Smooth Model reduces to EU. We need to estimate the parameter  $s$ .

### 2.3 Rank Dependent Expected Utility (RD)

This model was originally introduced by Quiggin (1982) and called by him Anticipated Utility. Since then it has been further developed by Segal (1987) and Kahneman and Tversky (1979) and has now become more commonly known as Rank Dependent Expected Utility theory, or, equivalently in our context<sup>3</sup>, as Cumulative Prospect Theory (Tversky and Kahneman 1992). It values a one-stage lottery differently from Expected Utility theory since probabilities are transformed. Let  $f(\cdot)$  denote the transformation function, where  $f(0)=0$ ,  $f(1)=1$  and  $f$  is non-decreasing everywhere. Then the generic one-stage lottery is valued as

$$(5) \quad V_{RD}[O(m, M)] = f\left(\frac{m}{M}\right)$$

In evaluating the generic two-stage lottery, since the one-stage lotteries within it are naturally ranked in order by our notation (note that  $m_1 < m_2 < \dots < m_N$ ) it follows that its valuation<sup>4</sup> is

$$(6) \quad V_{RD}[T(m_1, m_2, \dots, m_N; M)] = \sum_{i=1}^N f\left(\frac{m_i}{M}\right) \left[ f\left(\frac{N-i+1}{N}\right) - f\left(\frac{N-i}{N}\right) \right]$$

We need to specify the function  $f(\cdot)$ . We follow<sup>5</sup> the suggestion of Tversky and Kahneman (1992) and use a form which allows for realistic shapes of the function:

$$(7) \quad f(p) = \frac{p^g}{(p^g + (1-p)^g)^{1/g}}$$

We need to estimate the parameter  $g$ . When it takes the value 1, RD reduces to EU.

---

<sup>3</sup> Because with just two outcomes, the notion of a *reference point* is irrelevant.

<sup>4</sup> Here the usual convention that  $f(x)=0$  if  $x < 0$  never applies.

<sup>5</sup> We also tried the power form  $f(p) = p^g$  but this did not appear to represent an improvement.

## 1.4 Alpha Model (AM)

Ghirardato *et al* (2004) proposed a generalisation of the Gilboa and Schmeidler (1989) Maximin Expected Utility model. Here there is no use made of the probabilities attached to the members of the set of possible distributions, though use is made of knowledge of the best and worst members of this set. So this is a multiple prior model, but not one that makes use of actual values of the second-order probabilities. The model's valuation of a one-stage lottery is exactly as in EU. The worst of the one-stage lotteries within the generic two-stage lottery, in our notation, is  $O(m_1, M)$ , and the best is  $O(m_N, M)$ . The Alpha model values the generic second-order lottery as a weighted average of the value of the worst and the value of the best one-stage lotteries within the two-stage lottery. Hence we have:

$$(8) \quad V_{AM}[T(m_1, m_2, \dots, m_N; M)] = a \frac{m_1}{M} + (1-a) \frac{m_N}{M}$$

We note a special case: when  $a=1$  this reduces to Gilboa and Schmeidler's Maximin model (Gilboa and Schmeidler 1989).

## 2. Our Experimental Implementation

We have already sketched the main features of the experiment; we now give detail. Subjects completed the experiment individually at screened and separated computer terminals. During the experiment each of the subjects was presented with 49 tasks. Each task started out with two two-stage lotteries being portrayed on the computer screen, and the subject informed as to which of the two was the changing lottery, and hence which was the unchanging lottery. An example of the opening screen of a task is given in Figure 1. Each two-stage lottery was composed of a number of one-stage lotteries – which are the columns in the two two-stage lotteries in Figure 1. The subject was then asked to choose a winning colour (from blue and red) for that task.

After choosing the winning colour the task started. The following process was followed until the changing lottery was reduced to a single one-stage lottery. The subject was asked to select one

of the two stage-lotteries as his or her preferred lottery. Then one of the one-stage lotteries from the changing two-stage lottery was eliminated at random – leaving a column gap in the visual presentation. The subject was again asked to state his or her preferred two-stage lottery. After the last choice for a given task (when the changing lottery had been reduced to a single one-stage lottery) a new task was presented, until all 49 tasks had been completed. Different subjects got the 49 tasks in different, randomised, orders.

The natural incentive mechanism was used. Subjects were told that all their choices on each of the pairwise choice problems in each of the 49 tasks would be recorded. Then, at the end of the experiment, one of the 49 tasks would be randomly selected, one of the pairwise choice problems would be randomly selected, and the subject's preferred choice on that particular problem would be played out. To do this, one of the one-stage lotteries (if there were more than one), or the one-stage lottery (if there was only one), of the chosen two-stage lottery would be picked at random (with all one-stage lotteries within it having the same chance of being picked), and one of the balls in the one-stage lottery picked at random. If the colour of the ball picked was the subject's chosen Winning Colour on that task, the subject would get paid €40 in addition to the show-up fee.

We should note that the tasks used in the experiment – their number and the tasks themselves – were selected after extensive pre-experimental simulations. We were anxious to ensure that we would have both enough data, and enough informative data, to be able to distinguish between the different types. Obviously the answer to this question depends upon the *noise* in subjects' responses, so we first carried out some pilot studies to get some idea as to the magnitude of this noise. Equipped with this information, we then simulated a large number of experimental questions, selecting those that appeared most discriminatory. The design and results of these simulations are available on request.

The experiment was conducted in, and financed by, the experimental lab of the Max-Planck Institute of Economics, directed by Professor Werner Güth, with subjects recruited using ORSEE (Greiner 2004).



### 3. Related Experimental Literature

We provide here a brief summary of recent experimental work investigating theories of behaviour under ambiguity. Earlier literature is surveyed in Camerer and Weber (1992) and Camerer (1995). A fuller survey can be found in Hey and Pace (2011), which also gives more bibliographical information.

Halevy (2007) marks the start of new experimental work investigating the recent theories; subsequent contributions are those of Anderson *et al* (2009), Ahn *et al* (2010), Hey *et al* (2010), Hey and Pace (2011) and Abdellaoui *et al* (2011). To avoid duplication, we present here only the essential features of these experiments, specifically the following: the kinds of questions asked to the subjects: the way that ambiguity was implemented in the laboratory setting: and information about the theories (or class of theories) under test. Table 1 gives an overview.

In these experiments, there were three different types of questions asked to the subjects: (1) reservation price questions; (2) pairwise choice questions; (3) allocation questions. The advantage of the first and last types is that there is more information contained in the answer to each question than in the second type; the disadvantage is that it becomes necessary to estimate a utility function. Because we wanted to reduce the number of parameters that we needed to estimate, we took the decision not to estimate a utility function and thus we restricted the payments to the subjects to one of two values. This implied immediately that we could not ask allocation questions nor ask reservation price questions; and thus were restricted to pairwise choice questions. We compensated by having relatively many of them: 49 tasks and 256 pairwise choice questions in total; this is many more questions than is usually the case.

In contrast, the papers by Ahn *et al* (2010) and Hey and Pace (2011) used allocation questions, in which subjects were asked, in each of a series of questions, to allocate a given sum of tokens to various events, with given exchange rates between tokens and money for each of the events. If a particular question was selected to determine the payment to the subject, then the

question was played out and the resulting event, combined with the allocation that the subject had made to that event and the pre-specified exchange rate between tokens and money for that question and that event, determined the payment that the subject received for his or her participation (plus any given participation fee). As a result the actual payment might take one of a range of values, and hence a utility function over this range would have to be inferred from the subject's answers. This was also the case in the experiments of Halevy (2007) and Abdellaoui *et al* (2011), though both of these papers asked subjects to state *certainty equivalents* for certain events. In the first of these papers, the Becker-DeGroot-Marschak (BDM) method was used to get subjects to state their certainty equivalents; in the second Holt-Laury price lists were used. Both of these mechanisms have their problems: the BDM, in that appears to be the case that subjects find it difficult to understand what the mechanism entails; and the Holt-Laury, in that it appears that the answers given by subjects are sensitive to the bounds on the lists. Our method avoids both these problems.

The implementation of ambiguity in the laboratory also varies from paper to paper. Hey *et al* (2010) and Hey and Pace (2011) used a Bingo Blower – in which balls of differing colours (which define the events) are blown around inside the Blower in such a way that the balls can be seen but not counted. Anderson *et al* (2009) used a Bingo Cage, which is similar to a Bingo Blower in that the balls cannot be counted, but differs from it in that the balls are stationary and not being blown about continuously. Anderson *et al* also used bets on natural events (for example the temperature in Paris at a precise time in the future; see Baillon (2008) for more detail). Halevy (2007) used Ellsberg-type urns, described in much the same way as Ellsberg did:

“Urn 2: The number of red and black balls is unknown, it could be any number between 0 red balls (and 10 black balls) to 10 red balls (and 0 black balls)”

as well as urns which would be better described as two-stage lotteries:

“Urn 3: The number of red and black balls is determined as follows: one ticket is drawn from a bag containing 11 tickets with the numbers 0 to 10 written on them. The number written on the drawn ticket will determine the number of red balls in the third urn. For example, if the ticket drawn is 3, then there will be 3 red balls and 7 black balls.”

Abdellaoui *et al* (2011) also used Ellsberg-like urns, but with 8 colours:

“The known urn *K* contained eight balls of different colours: red, blue, yellow, black, green, purple, brown, cyan. The unknown urn contained eight balls with the same eight colours, but the composition was unknown in the sense that some colours might appear several times and others might be absent.”

This has the same feature as the original unknown Ellsberg urn – subjects were not informed about the process of the formation of this urn. It might therefore be the case that they regard this as the “suspicious urn.”

In contrast, and particularly because our experiment was specifically designed to test multiple-prior models of ambiguity, we used two-stage lotteries (like Halevy’s Urns 2 and 3). These are exactly the type of ambiguity referred to in the theories under test. Moreover, we control the probabilities of the various possibilities, and they are therefore immune to subjects’ subjective interpretation and assessment.

Partly because of the different types of questions and the different implementations of ambiguity, the theories under test also differ from paper to paper. But there is a second key difference between the theories under test. Three of the papers do not investigate *specific* models but rather ‘generic classes’ of models. This is particularly true of Ahn *et al* (2010) who investigate special cases of two classes: smooth and kinked. The kinked class essentially consists of those theories that are rank dependent in some sense: obviously Choquet EU (see Schmeidler 1989) is in this set, as is Rank Dependent EU. There is a kink in the preference function when the ranking of the outcomes changes. In contrast there is no kink in the members of the smooth set, because ranking is not important. A key member of this latter set is Klibanoff *et al*’s Smooth Model. So, in a sense, the Ahn *et al* paper investigates two *generic* classes, though it should be noted that it does not investigate any *specific* model in either class. Anderson *et al* do a similar investigation of the smooth class, while Abdellaoui *et al* do the same for the kinked class. It follows that none of these theories investigate specific theories. This is in contrast to the other papers in the table.

One key remaining difference between these various papers is their objectives. The Hey *et al* and Hey and Pace papers *estimate* preference functionals and see how well the data fits the various theories. So does Anderson *et al*. In contrast Halevy *tests* between the various theories. Abdellaoui *et al*, by design, do not compare different theories but instead describe the way that ambiguity and risk enter the decision process. While Ahn *et al* adopt a different methodological approach to that of Abdellaoui *et al*, their interest is similar; they too are interested in how attitude to ambiguity can be characterised and how it enters into the decision making process. In contrast, given the recent activity of the theorists in producing theories of decision-making under two-stage ambiguity, *our* objective in this paper is to discover which of these is ‘best’ and hence worth pursuing.

#### **4. Stochastic assumptions**

We assume that decisions are made with error. Our specific stochastic assumption is that the difference in the value of the objective function is calculated with error. Specifically we follow precedent and assume that, while the decision on any of the pairwise choice questions in the experiment *should* be made on the basis of the difference between  $V(L)$  and  $V(R)$  (where  $L$  and  $R$  are the two lotteries involved in the decision), it is actually made on the basis of  $V(L) - V(R) + u$  where  $u$  is  $N(0, \sigma^2)$  and is independent across questions. We also add a probabilistic ‘tremble’ (see Moffat and Peters 2001) of magnitude  $\omega$ . We estimate the parameters of the various models ( $s$  in the Smooth Model,  $g$  in the Rank Dependent Model and  $a$  in the Alpha Model), as well as  $\sigma$  and  $\omega$ .

#### **5. Econometric procedures**

We have four models: Expected Utility (EU), the Smooth Model (SM), the Rank Dependent model (RD) and the Alpha Model (AM). Our goal is to try and assign subjects to models/types – to see if, and which of, these models describe the behaviour of our subjects. We start by fitting the various models subject-by-subject. We then use a mixture model, pooling the data from all our 149 subjects, and employ the Wilcoxon estimation/prediction methodology: using just over half our data for

estimation, and the remaining data for prediction – hence assessing the relative predictive ability of the theories. We provide details in the next two sub-sections.

### 6.1 Estimation subject-by-subject and model-by-model

We have a total of 149 subjects. Each of them completed 49 tasks, composed of a grand total of 256 pairwise choice questions. We summarise the subject-by-subject results in Table 2. In this table we assign subjects to a particular type according to pair-wise model testing: likelihood-ratio tests for nested pairs (EU  $\nu$  SM and EU  $\nu$  RD) and BIC tests for non-nested pairs (EU  $\nu$  AM, SM  $\nu$  RD, SM  $\nu$  AM, RD  $\nu$  AM).

We see that, using this method to classify subjects to models/types we assign 24% to EU, 56% to SM, 11% to RD and 9% to AM. We note that this assignment of subjects to type is not very different from the assignment we obtain from the mixture estimates as we shall see later: respectively 26%, 49%, 20% and 5% (see Table 3).

Figure 2 displays histograms of parameter estimates. The panes on the left show histograms of parameter estimates over all subjects in the sample. The panes on the right contain histograms of parameter estimates only from subjects who are classified as being of the type,  $h$ , with  $h \in \{SM, RD, AM\}$ , that is indicated in the corresponding row (see also Table 2). It may be of interest to recall that SM reduces to EU when the parameter  $s$  takes the value 0; while RD reduces to EU when  $g$  takes the value 1.

### 6.2 Mixture Estimation

The individual estimates reported above are very extravagant in terms of numbers of parameters. Moreover, this subject-by-subject approach ignores any smoothness in the distribution of parameters that exists in the population from which this sample was drawn. To avoid this extravagance, and provide a more parsimonious specification, we now introduce a mixture model. In

this context, a mixture model<sup>6</sup> assumes that a certain proportion of the population from which the subject pool was drawn is of each of the four types under consideration (EU, SM, RD and AM): these proportions in the population are termed the *mixing proportions*; further, and this is the way a mixture model reduces the extravagance, within each type the approach assumes that each parameter specified in that particular type has a distribution over the population. Estimation involves pooling the data over all subjects and estimating the mixing proportions and the parameters of the distributions (over the population) of the parameters of the various models. So, for example, instead of having 149 different estimates of the smooth parameter (one for each subject) this mixture estimation provides just two: the mean and the variance of the distribution of the smooth parameter over the population. The same is true for all the other parameters. The histograms in Figure 2 are effectively replaced by the distributions in Figure 3. Instead of getting a picture of 149 separate individuals, we are getting a picture of the set of individuals in the population from which these 149 individuals were drawn. Our results are more generalisable, and enable us to make predictions on samples different from the one used in this paper, while individual-by-individual analysis does not.

So we use a mixture model. We add an additional methodological component: we follow the Wilcoxon estimation-then-prediction approach (Wilcox 2007) to assess the comparative empirical plausibility of the various models. Basically this approach involves splitting the data up into two parts, one used for estimation and the other (using the resulting estimates) for prediction. He advocates this approach because of the greater challenge of the test of good prediction.

Wilcox (2007) advocates this approach not in a mixture model context but in an individual context. In this individual context the approach requires that data are estimated subject-by-subject and model-by-model on part of the data, and that the identification of a subject's type is based on the predictive performance of each model on the remaining data. We have to modify appropriately his approach in a mixture model context. As we have already noted the mixture model approach

---

<sup>6</sup> For further details on the mixture approach in the context of choice under risk, see Conte et al. (2011).

pools data over subjects and hence deals with subjects' heterogeneity in two ways: it allows subjects to be of different types; and to differ in terms of the parameters for the functional that they have. The former point is taken care of by assuming that each subject is of one type, and that he or she does not change type throughout the experiment, and by estimating the proportion of the population who are of each type. The latter is dealt with by assuming a distribution of the relevant parameters over the population and by estimating the parameters of that distribution.

The reason we decided to combine the two approaches is twofold. We find the Wilcoxon approach extremely elegant and powerful, but we realised that it does not work well on an individual basis with our data in the context of our experiment. Suppose, for example, that a subject behaves exactly as EU without any mistake but one, and that the wrong choice is included in the sample of choices used for prediction. From the sample used for estimation, we infer that our subject is EU with an error term with zero variance. When we try to calculate the likelihood of the prediction, the only mistaken choice results in a log-likelihood of  $-\infty$ . Consequently, we might be induced to reject the hypothesis for that subject of being EU in favour of another model, when that subject is in fact EU. The mixture approach assumes that the parameters of the relevant models follow a certain distribution but, at the same time, provides consistent estimates of the proportion of the population who are of each type. In a particular sense, it introduces a parameter smoothness assumption, but uses information about the population (mixing proportions and distribution of the relevant parameters for the functionals included in the mixture) in assigning subjects to types.

In order to implement our approach, we split our observations into two samples: (1) an estimation sample,  $I^e$ , that includes the observations from 25 tasks (selected at random), corresponding to 132 binary choices, and we use these to estimate the mixture model; (2) a prediction sample,  $I^p$ , that includes the remainder 24 tasks, corresponding to 124 binary choices, and we use these to make predictions. More precisely, we proceed as follows: we use the estimation sample to estimate the mixture model; we calculate the likelihood of the prediction from the prediction sample using the parameter estimates obtained in the first step. We finally compute by

Bayes' rule the posterior probability for each subject of being of each type from the likelihood of the prediction and assign subjects to types according to the highest posterior probability of the prediction.

As we have already noted, we start by assuming that a proportion  $\pi_k$  of the population from which the experimental sample is drawn is of type  $k$ , with  $\pi_k \leq 1$ ,  $\sum_k \pi_k = 1$  and  $k \in (EU, SM, RD, AM)$ . These *mixing proportions*  $\pi_k$ , with  $k \in (EU, SM, RD, AM)$ , are estimated along with the other parameters of the mixture model. The likelihood contribution of subject  $i$  is

$$(10) \quad L_i = \sum_{k \in (EU, SM, RD, AM)} \pi_k \times \tilde{l}_i^k,$$

where  $\tilde{l}_i^k$  is the penalised likelihood contribution of individual  $i$  under the hypothesis of her being of type  $k$ , with  $k \in (EU, SM, RD, AM)$ . We will define what we mean by a "penalised" likelihood contribution after having specified the likelihood contribution of each type. For this purpose, let us use the binary variable  $y_{it}$  that takes the value 1 if subject  $i$  chooses the left-hand lottery in problem  $t$ , it takes the value -1 if subject  $i$  chooses the right-hand lottery in problem  $t$ . The likelihood contribution for subject  $i$ 's choice in problem  $t$ , given that subject  $i$  is of type  $EU$ , is

$$(11) \quad l_i^{EU} = \prod_{t \in I^e} \left\{ (1 - \omega) \Phi \left[ y_{it} \times \frac{V_{EU(L)} - V_{EU(R)}}{\sigma_{EU}} \right] + \frac{\omega}{2} \right\} \quad y_{it} \in \{1, -1\}.$$

Here,  $\Phi(\cdot)$  is the unit normal cumulative distribution function;  $\sigma_{EU}$  is the standard deviation of the Fechner error term for the EU model; and  $\omega$  is a tremble probability (the probability that the subject chooses at random between the two lotteries). The likelihood contribution for subject  $i$ 's choice in problem  $t$ , given that subject  $i$  is of type  $h \in (SM, SQ, AM)$ , is

$$(12) \quad l_i^h = \int_{\underline{b}}^{\bar{b}} \prod_{t \in I^e} \left\{ (1 - \omega) \Phi \left[ y_{it} \times \frac{V_h(L) - V_h(R)}{\sigma_h} \right] + \frac{\omega}{2} \right\} g(r) dr \quad y_{it} \in \{1, -1\}.$$

Here,  $g(r)$  is the density of the parameter  $r$ . In the case  $h = SM$ ,  $r$  represents the parameter  $s$ , with  $s \sim N(\mu, \theta^2)$ ,  $\underline{b} = -\infty$  and  $\bar{b} = +\infty$ . In the case<sup>7</sup>  $h = RD$ ,  $r$  represents the parameter  $g$ , with  $\ln(g - 0.279095) \sim N(\delta, \epsilon^2)$ ,  $\underline{b} = 0.279095$  and  $\bar{b} = +\infty$ . In the case<sup>8</sup>  $h = AM$ ,  $r$  represents the

<sup>7</sup> In this case the parameter  $g$  has to be bounded below to make the function  $f(\cdot)$  monotonically increasing over  $[0, 1]$ .

<sup>8</sup> In this case the parameter  $a$  has to be between 0 and 1.



parameter  $a$ , with  $a \sim \text{Beta}(\vartheta, \varphi)$ ,  $\underline{b} = 0$  and  $\bar{b} = 1$ .  $\sigma_h$  are the standard deviations of the Fechner error term attached to each model in  $h$ .  $\Phi$  and  $\omega$  are the same as defined for type EU. We assume that  $\omega$  is independent of type because there is no reason to assume that any type should tremble more than another. The sum over all  $i$  in the sample for  $t \in I^e$  of the logarithm of (10) is maximised by maximum simulated likelihood; the integrations in (12) are performed by four sets of Halton sequences.

Likelihood contributions are penalised depending upon the number of observations used in the estimation and the number of free parameters estimated.<sup>9</sup> Thus, the relationship between  $l_i^k$  and  $\tilde{l}_i^k$  is

$$(13) \quad \tilde{l}_i^k = \frac{l_i^k}{T^{\frac{j}{2}}} \quad k \in (EU, SM, RD, AM)$$

Here,  $T = 132$  is the number of choices per individual in the estimation sample.  $j$  is the number of free parameters estimated. We have that  $j$  equals one for type EU and three for types SM, RD and AM. The penalisation of the likelihood contributions used here corresponds to the Schwarz-Bayesian Information Criterion (BIC).

Table 3 reports parameter estimates of the mixture model derived above. The estimates of the mixing proportions suggest that 26% of the population is EU, 49% is SM, 20% is RD and 5% is AM. Figure 3 shows density plots of the mixture model parameters according to the estimated results in Table 3. We note, in passing, that these mixing proportions are close to the proportions assigned to types that we obtained from the individual estimates, and are also close to the posterior probabilities that we will present later in Table 4.

### 6.2.1 Prediction results

Now that we have obtained the estimation results reported in Table 3, we can proceed with calculating the likelihoods of the choice prediction using equations (10), (11), (12) and (13) but

---

<sup>9</sup> See, for example, Preminger and Wettstein (2005).

where we replace parameters to be estimated with parameter estimates, which are assembled in the vector  $\mathbf{Y} = (\hat{\sigma}_{EU}, \hat{\mu}, \hat{\theta}, \hat{\sigma}_{SM}, \hat{\delta}, \hat{\epsilon}, \hat{\sigma}_{RD}, \hat{\vartheta}, \hat{\varphi}, \hat{\sigma}_{AL}, \hat{\pi}_{EU}, \hat{\pi}_{SM}, \hat{\pi}_{RD}, \hat{\pi}_{AM})$ . Tasks and choices are those contained in the prediction sample,  $y_{it}$  with  $t \in I^P$ . In equations (13)  $T = 124$ . Using Bayes' rule we compute the posterior probabilities of types for each  $i$  in our sample as follows:

$$(14) \quad Pr[i \text{ is of type } k | y_{it} \text{ with } t \in I^P, \mathbf{Y}] = \frac{Pr[\text{type } k] \times Pr[y_{it} \text{ with } t \in I^P | i \text{ is of type } k]}{Pr[y_{it} \text{ with } t \in I^P]}$$

$$= \frac{\hat{\pi}_k \times \hat{l}_i^k(y_{it} \text{ with } t \in I^P, \mathbf{Y})}{L_i(y_{it} \text{ with } t \in I^P, \mathbf{Y})} \quad k \in (EU, SM, RD, AM).$$

Basically, we re-calculate equations (10), (11), (12) and (13) by replacing parameters with parameter estimates, and, for each subject, by replacing tasks and choices from the estimation sample with tasks and choices from the prediction sample.<sup>10</sup> We can also calculate the posterior probabilities based on the *estimation* sample; these are given by equation (14) with  $I^P$  replaced by  $I^e$ ; of course here we use the maximised likelihood over the estimation sample. We can use these posterior probabilities on both the estimates and the predictions to assign subjects to types, though, following Wilcox (2007) we prefer assignment on the predictions. Interestingly, however, in this context the proportions assigned to types are not very different whether we use estimates or predictions (see Table 4 which we will explain shortly). First, however, we explain the assignment with the help of Figure 4.

Figure 4 displays six simplexes summarising the results from analysing the posterior probabilities; the first row of three simplexes refers to the estimation sample, and the second row to the prediction sample. This figure is crucial as it shows very clearly the power of our experiment in classifying the subjects by type. Each simplex represents the posterior probabilities of types for each subject in our sample computed following the procedure explained above. Subjects are points in the simplexes: the closer a subject is to the vertex of a simplex, the higher the posterior probability for that subject of the type represented on that vertex. The smallest circles represent single subjects, larger circles represent concentration of subjects in that area of the simplex: the larger the circle, the

---

<sup>10</sup> Obviously, we do not maximise the resulting likelihood at this stage, because we use parameter estimates.

higher the number of subjects represented by that circle. Since we realised that for all subjects in our sample no more than three posterior probabilities at a time were larger than .005, we grouped subjects in three simplexes that display on each vertex the three types for which the posterior probabilities are larger than .005. The simplexes in the first (estimation) row involve 61, 44 and 44 subjects respectively; those in the second (prediction) row 72, 31 and 46 respectively. We see that most of the subjects are concentrated at the vertices of the simplexes, showing that our mixture assigns them to type with very high posterior probability. This is a really crucial finding: it implies that, except for a handful of subjects for whom there is some confusion as to their type, the vast majority of the subjects can very clearly be assigned to a specific type.

This statement is confirmed by Figure 5 (which shows the cumulative percentage of subjects assigned to a type with maximum posterior probability less than the probability indicated on the horizontal axis) both for the estimates (the black step function) and the predictions (the grey step function). It shows that the predictive power of our mixture model is quite impressive: all subjects are assigned to a type with posterior probability larger than .5; more than 95% of subjects are assigned to a type with posterior probability .75 or more; almost 90% of subjects in our sample are assigned to a type with posterior probability .90 or more. The prediction is slightly less impressive (as the step function is above that for the estimations) but only marginally so.

Table 4 presents the results of assignment to types according to the posterior probabilities of the estimates (the final column) and of the predictions (the bottom row); it also presents a cross-tabulation (in the cells of the matrix). Using the posterior probabilities from the estimates (predictions) 25% (22%) are assigned to EU, 50% (53%) to SM, 20% (22%) to RD and 5% (3%) to AM. The bulk of the subjects appear down the main diagonal in Table 4 (indicating the same assignment on both the estimation and the prediction) though 26.8% are classified differently on the two methods. This 'instability' is a feature pointed out by Wilcox (2007) as inevitable unless one has a lot of data: in our case it would seem that 49 tasks or 256 observations are not enough for some subjects and that these subjects are simply unstable and do not have a fixed preference function. In

conclusion it would appear to be the case that we can assign virtually all subjects to a type with very high probability.

## 6. Conclusions

We have examined the empirical plausibility (both in estimation and prediction) of four multiple prior models of choice under ambiguity (these being an important subset of all models of behaviour under ambiguity). We carried out an appropriate experiment on 149 subjects, involving 49 tasks and 256 pairwise decision problems, and used the resulting data to assign subjects to models/types. We started with estimation subject-by-subject, and then we turned to a mixture model, pooling all our 149 subjects together. The mixture model enables us to classify subjects to models with a high degree of accuracy, with posterior probabilities of subjects being of one type or another being over 0.90 for around 90% of our subjects. These posterior probabilities suggest that around 22% of our subjects follow Expected Utility (EU) theory, 53% the Smooth Model (SM) of Klibanoff *et al*, some 22% the Rank Dependent (RD) model and around 3% the Alpha Model (AM) of Ghirardato *et al*. Interestingly the assignments based on the estimated data do not differ very much from those based on prediction data – unlike in Wilcox (2007).

In a sense the poor showing of the Alpha model is not surprising as the model does not use all the information given to the subjects; indeed one might not even call it a genuine multiple prior model. Both EU and RD appear to perform equally well. However, the clear ‘winner’ in terms of the number of subjects with that preference functional is the Smooth model – suggesting that over 50% of subjects do not reduce two-stage gambles to their one-stage equivalents analysing the decision problems: they do indeed view two-stage trees as different from the reduced one-stage trees.

## 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.
- Anderson, S., Fountain, J., Harrison, G.W. and E.E. Rutström (2009), "Estimating Aversion to Uncertainty", *Working Paper*.
- Ahn, D.S., Choi, S., Gale, D., and S. Kariv (2010), "Estimating Ambiguity Aversion in a Portfolio Choice Experiment", *Working Paper*.
- Baillon, A. (2008), "Eliciting Subjective Probabilities Through Exchangeable Events: An Advantage and a Limitation," *Decision Analysis*, 5, 76-87.
- Camerer, C. (1995), "Individual Decision Making" in J. Kagel and A. Roth (Eds.), *Handbook of Experimental Economics*, Princeton University Press, 587-703.
- Camerer, C. and M. Weber (1992), "Recent Development in Modelling Preferences: Uncertainty and Ambiguity", *Journal of Risk and Uncertainty*, 5,4: 325-370.
- Choquet
- Conte, A., Hey J.D and P.G. Moffatt (2011), "Mixture Models of Choice under Risk", *Journal of Econometrics*, 162(1), 79-88.
- Ghirardato, P., Maccheroni F., and M. Marinacci (2004), "Differentiating Ambiguity and Ambiguity Attitude", *Journal of Economic Theory*, 118, 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](http://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.

Hey, J.D. and M. Pace (2011), "The Explanatory and Predictive Power of Non Two-Stage-Probability Theories of Decision Making Under Ambiguity", *University of York Department of Economics and Related Studies Discussion Paper 11/22*.

Kahneman D. and Tversky A. (1979) Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, 67, 263-291.

Klibanoff P., Marinaci M. and S. Mukerji (2005), "A Smooth Model of Decision Making Under Ambiguity", *Econometrica*, 73, 1849-1892.

Moffat P.G. and S.A. Peters (2001), "Testing for the Presence of a Tremble in Economic Experiments", *Experimental Economics*, 4, 221-228.

Preminger A. and D. Wettstein (2005), "Using the Penalized Likelihood Method for Model Selection with Nuisance Parameters Present only under the Alternative: An Application to Switching Regression Models", *Journal of Time Series Analysis* 26(5), 715-741.

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

Segal U. (1987), "The Ellsberg Paradox and Risk Aversion: an Anticipated Utility Approach", *International Economic Review*, 28, 175-202.

Tversky A. and D. Kahneman (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty*, 5, 297-323.

Wakker, P.P. (2010), *Prospect Theory: For Risk and Ambiguity*, Cambridge University Press, Cambridge, UK.

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

Table 1: Recent related experimental work

Paper	Type of questions	Implementation of ambiguity	Theories being investigated
Halevy (2007)	Certainty equivalents/Reservation prices (using BDM Mechanism)	Ellsberg-type urns	SEU, Maximin, Anticipated Utility <sup>ac</sup> , Smooth <sup>d</sup>
Anderson <i>et al</i> (2009)	Pairwise Choices	Bingo cage and real events	Minimalist non-EU model (a special case of Smooth <sup>d</sup> )
Ahn <i>et al</i> (2010)	Allocations	Probabilities of two of the three states not stated	Two broad classes: smooth and kinked; special cases of more general models
Hey, Lotito and Maffioletti (2010)	Pairwise choices	Bingo Blower	EV, SEU, Choquet, Prospect, Cumulative Prospect, Decision Field Theory, Alpha <sup>b</sup> plus some older theories
Hey and Pace (2011)	Allocations	Bingo Blower	SEU, Choquet, Alpha <sup>b</sup> , Vector EU, Contraction <sup>e</sup>
Abdellaoui <i>et al</i> (2011)	Certainty equivalents/Reservation prices (using Holt-Laury price lists)	8-colour Ellsberg type urns	RDEU
Conte and Hey (2011) <sup>f</sup>	Pairwise choices	Multiple priors	SEU, Smooth, Alpha <sup>b</sup> , Anticipated Utility

Notes: <sup>a</sup>Referred to as Recursive Nonexpected Utility; <sup>b</sup>Ghirardato et al (2004) (Including the special cases Maxmin and Maxmax); <sup>c</sup>It is not clear to me what weighting function (Quiggin or Power) is used; <sup>d</sup>Klibanoff et al (2005); <sup>e</sup>Gajdos et al (2008); <sup>f</sup>this paper.

Table 2: Individual by individual analysis  
Subjects are assigned to types by likelihood-ratio and BIC tests

type	Freq.	%
EU	36	24.16
SM	83	55.70
RD	17	11.41
AM	13	8.72
Total	149	100.00



Table 3: Estimates results of mixture model (10).

type	parameters	mixing proportions
EU	$\sigma_{EU}$ .28998354 (.0130494)	$\pi_{EU}$ .2604322 (.0376641)
SM	$\mu$ 1.37699 (.1325007) $\theta$ 1.084952 (.1045272) $\sigma_{SM}$ .0877766 (.0025284)	$\pi_{SM}$ .492443 (.043891)
RD	$\delta$ -.1799474 (.0495131) $\epsilon$ .2424905 (.040733) $\sigma_{RD}$ .0245302 (.0023192)	$\pi_{RD}$ .1981231 (.0340864)
AM	$\vartheta$ 7.994903 (2.936609) $\varphi$ 6.762214 (2.335167) $\sigma_{AL}$ .115276 (.0109145)	$\pi_{AM}$ .0490017 (.0216843)
	$\omega$ .0149681 (.0027092)	
	Log-likelihood	-8926.8476
	number of observations	19668
	number of subjects	149
	number of observations per subject	132

Table 4: Cross tabulation of subjects assigned to types according to highest posterior probabilities of the estimate (rows) and subjects assigned to types according to the highest posterior probability of the prediction (columns.)

		Assignment from prediction sample				
		EU	SM	RD	AM	total
Assignment from estimation sample	EU	25	11	0	2	38
	SM	5	59	8	2	74
	RD	0	6	24	0	30
	AM	3	3	0	1	7
	total	33	79	32	5	149

Figure 1: The experimental interface

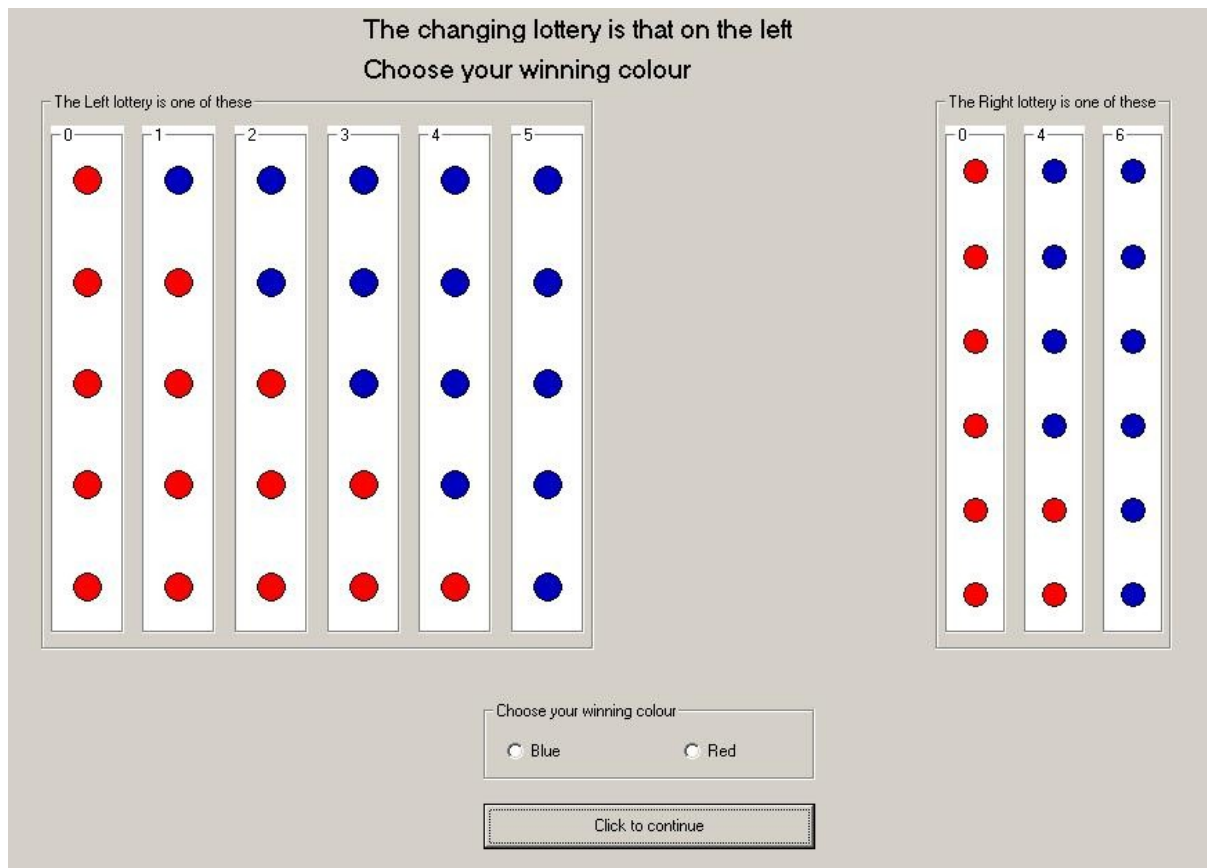


Figure 2: Histograms of parameter estimates obtained by individual by individual and model by model analysis of the data.

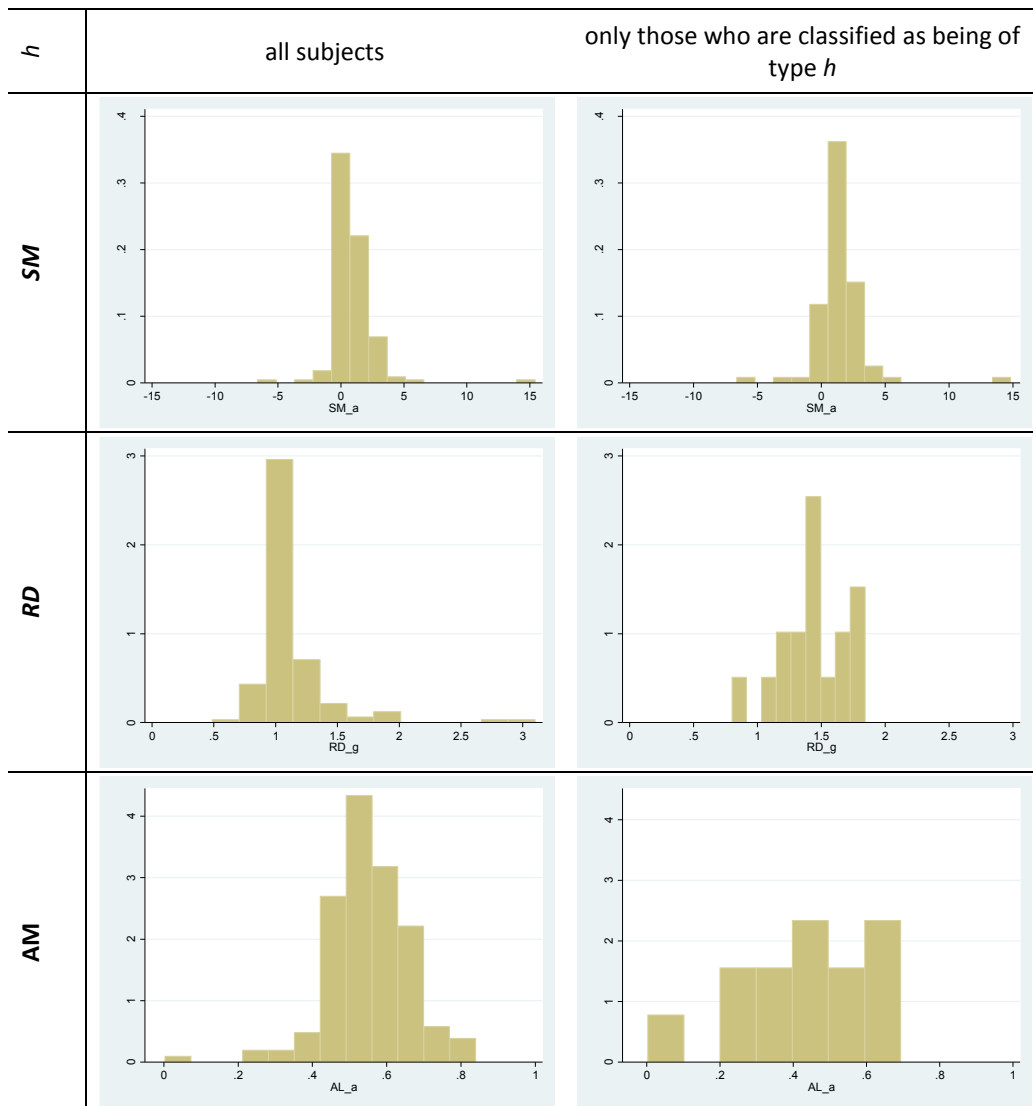


Figure 3: Density plots of the relevant parameters of the functional in mixture model (10) from parameter estimates in Table 3.

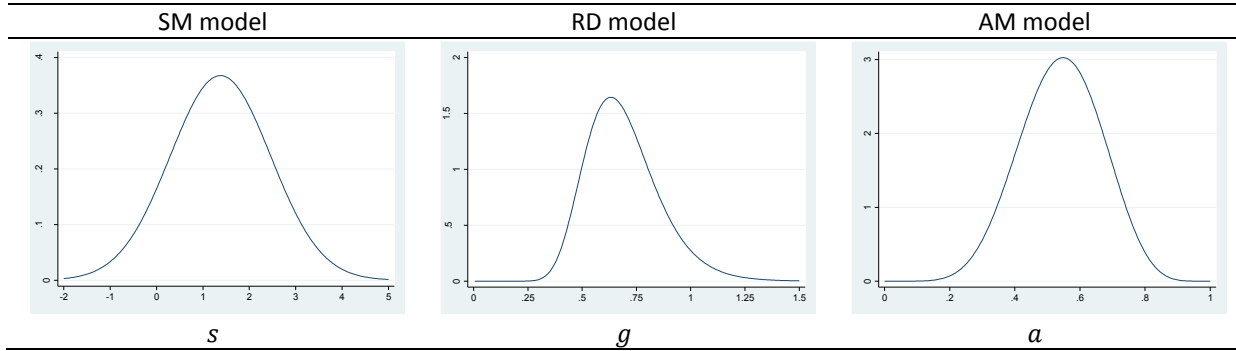


Figure 4: Posterior probabilities distributions of the 4 types. In each simplex, the posterior probability of the missing type is  $< .005$ . For the estimation sample, the three graphs (left to right) involve 61, 44 and 44 subjects, respectively. For the prediction sample, 72, 31 and 46, respectively. To produce the graphs, posterior probabilities have been rounded to the closest .05.

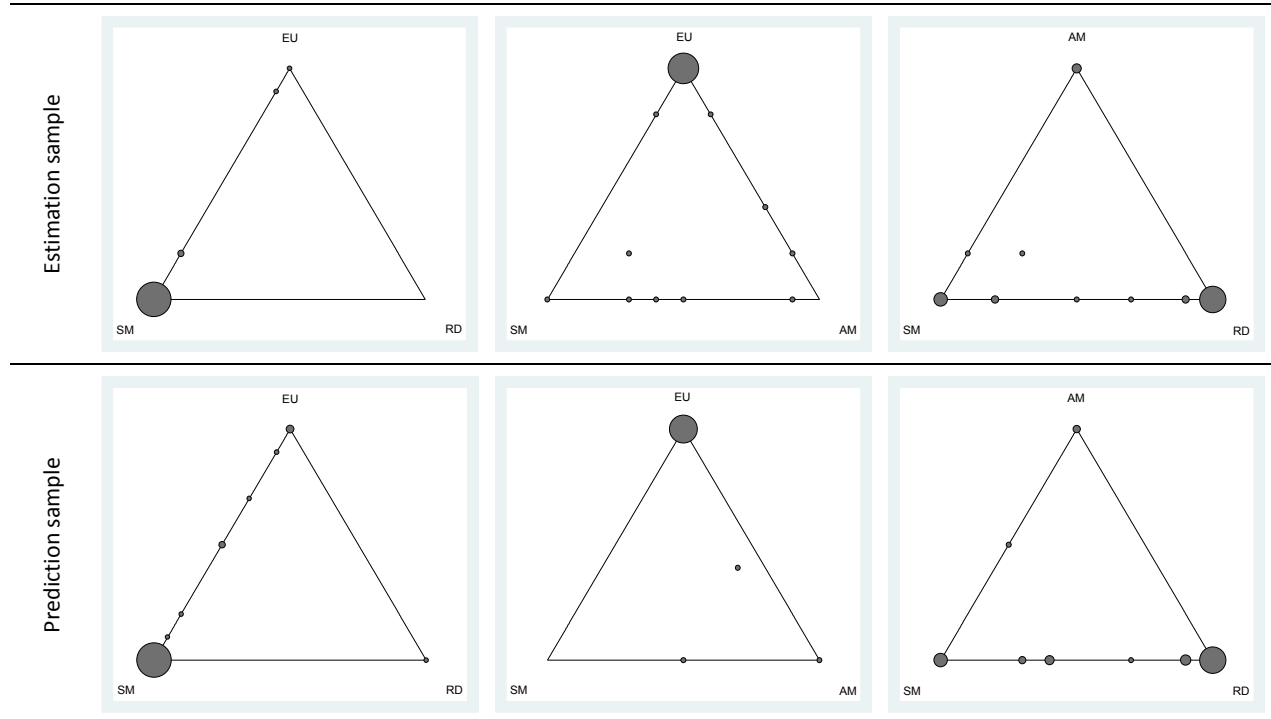


Figure 5: Cumulative percentage of subjects assigned to a type with maximum posterior probability < posterior probability indicated on the horizontal axis.

