Real monetary incentives and chained questions: An experimental study investigating the validity of risk estimates elicited via exchangeability method

Simone Cerroni
Ph.D. candidate at the Department of Economics. University of Trento.
email: simone.cerroni@unitn.it

Sandra Notaro
Researcher at the Department of Economics. University of Trento.

W. Douglass Shaw
Professor at the Department of Agricultural Economics and Research Fellow at the Hazards Reduction and Recovery Center. Texas A&M University.

Paper prepared for presentation at the EAAE 2011 Congress
Change and Uncertainty
Challenges for Agriculture,
Food and Natural Resources

August 30 to September 2, 2011
ETH Zurich, Zurich, Switzerland

1. Introduction

During the last decade, the lack of precise scientific measures of future environmental and agricultural risks has often induced many economists to investigate people's perceptions of those risks. The study of risk perceptions becomes even more important when it is apparent that people behave and make decisions according to what they believe risks to be, and when these beliefs might not coincide with science-based estimates of risk.

Since the main critique about the use of these subjective estimates in economic analyses is about their reliability, we aim to test the validity of subjective estimates of risks elicited via an innovative risk elicitation technique, the exchangeability method (EM). In particular, we focus on consumers' perceptions of the probability that given levels of pesticide residues will be present in apples produced in the future. These residues pose a health risk to people who eat the apples, and, thus, their presence on apples can affect economic behaviours of consumers.

Although the EM has many advantages compared to other risk elicitation techniques, many doubts have been raised about its incentive compatibility. In fact, the chained structure of its experimental design is accused of undermining the incentive compatibility of the process, especially when respondents are provided with real monetary incentives (Harrison, 1986). Previous studies have overcome this issue by presenting people with peculiar experimental designs that partially hide the chained structure of the game (Baillon, 2008; Abdellaoui et al., 2010) but, to our knowledge, no study has ever tested Harrison's hypothesis.

Our lab experiment aims to define a clear-cut valuation method for measuring the validity of subjective risk estimates elicited via the EM. This valuation method based on de Finetti's notion of "*coherence*" examines the "*validity rates*" of subjective risk estimates, which is the number of valid risk measures obtained from each specific experimental design we use during the experiment. Through this method we also aim to exam the potential effect of real monetary incentives and chained questions on subjective risk estimates elicited via EM. In particular, we study whether these factors affect the validity of elicited subjective risks or not[1].

2. The Exchangeability game

Considering a random variable under study, for instance g, the EM uses a series of binary questions to reveal an individual's underlying cumulative distribution function (CDF) over an event x that is drawn from an event space, $S_G = G_1^1$. The first step of the EM establishes the lower and upper bounds of the event space, say g_0 and g_1. Each subject is asked the bounds on outcomes outside of which they are essentially certain the outcome cannot happen at all — i.e., the bounds that pertain to a non-zero probability of an outcome.

The second step involves asking a series of questions that establish the value of $g_{1/2} \in S_G$ that corresponds with the 50th percentile of the subjective CDF, the median estimate. This series of questions asks the subject to choose between binary prospects. In the first binary question, S_G is divided at a point g_a into two prospects, say $G_a=\{g_0<x\leq g_a\}$ and $G_a'=\{g_a<x<g_1\}$, where $g_a=\{g_0 + [(g_1-g_0)/2]\}$ (see Example 1 in Appendix A). If G_a was chosen by the individual, the implication is that the individual believes $P(G_a)\geq P(G_a')$, so that $g_a\geq g_{1/2}$. A follow-up binary question is then asked of this same individual, using a new value g_b and two new prospects G_b and G_b'. If G_a was chosen in the first question, then $g_a<g_b$. However, if G_a' was chosen in the

[1] Only expected results are presented in this version of the manuscript because the experiment will be run in the first week of February at the Computable and Experimental Economics Laboratory (CEEL) of the University of Trento (Italy). Here, we present the theoretical framework and the testable hypotheses we use to test the validity of the EM. However, these theoretical hypotheses will be empirically tested using data coming from the experiment.

first question, then $g_b > g_a$. This process is repeated until the individual reaches a value g_z such that she is indifferent between G_z and G_z'. When this point is reached, it follows that $g_z = g_{1/2}$, $G_z = G_2^1$, $G_z' = G_2^2$, and $P(G_z) = P(G_z')$.

A similar process can be followed to determine other points for the individual's subjective CDF; in theory as many as the researcher wants, but of course limited by exhaustion of the subject. For instance, to determine the value of $g_{1/4} \in S_G$ that corresponds with the 25th percentile, a gamble is proposed that is contingent on a value of x that is lower than $g_{1/2}$, obtained in the previous step. Once again, a sequence of values, g_a, g_b, …, g_z is used, but in this case the initial upper bound is $g_{1/2}$. In the first new binary question, subjects choose between the following binary prospects, $G_a = \{g_0 < x \leq g_A\}$ and $G_a' = \{k_1 < x < g_{1/2}\}$. As above, this process is repeated until the individual is indifferent between G_z and G_z', so that $g_z = g_{1/4}$ $G_z = G_4^1$, $G_z' = G_4^2$, and $P(G_z) = P(G_z')$ (Figure 1).

3. Strengths and limitations of Exchangeability Method

The reliability of many risk elicitation techniques has been tested in the literature reporting on laboratory experiment results, especially concerning financial risks, but, to some extent for health and environmental risks. Defining the risk as a probability that given outcomes occur (or that given magnitudes of an outcome occur), these techniques are used to obtain subjective probability estimates of the occurrence of those outcomes (or magnitudes). Many investigations have elicited risks through "*direct methods*", that is by asking people to directly state either the chance that a specific magnitude of the outcome will happen in the future or, the other way round, the specific magnitude of the outcome that will happen with a certain probability (Spetzler and Von Holstein, 1975). Although this approach is very much straightforward, many doubts have been raised about the reliability of risk estimates because of the unfamiliarity of laypeople with the notion of probability (e.g., Jakus et al., 2009; Riddel and Shaw, 2006 for health risks; and Baker et al., 2009; Botzen et al., 2009 for environmental risks).

More sophisticated methodologies, eliciting risk measures via respondents' choices over lotteries and bets, have been implemented to overcome limitations of "*direct methods*". In this case, probability measures are indirectly estimated at the point in which people show indifference between lotteries or gambles. Even if these techniques, called "*indirect*", have been mostly used for financial risks (e.g., Andersen et al., 2009; Offerman et al., 2009), nowadays scholars start considering them for estimating health and environmental risks (e.g., see Fiore et al., 2009; Cerroni and Shaw, 2011 for environmental risks)[2].

The most popular "*indirect methods*" are the so-called "*external reference events*" in which people are asked to bet on, either a lottery characterized by an uncertain event whose probability needs to be estimated, or a lottery characterized by an external reference event whose probability is known. The probability related to the latter lottery is often visually presented through probability wheels or scroll bars. During experiments using these methods, people are asked to express their preferences for one of the two lotteries, and, when respondents are indifferent between them, this means that they attach to the uncertain outcome the same probability with which the familiar outcome will happen (Spetzler and Von Holstein, 1975). Although, these techniques are widely used, they have a crucial drawback related to the notion of "*source*

[2] The limited use of "*indirect methods*" for eliciting health and environmental risks is only due to the fact that health outcomes and very long term environmental outcomes cannot be played out at the end of experiments in the lab setting.

dependence”. Some experimental studies have recently shown that individual choices depend on the source of uncertainty[3] respondent have been told to consider (Kilka and Weber, 2001; Abdellaoui et al, 2010). When individuals have to process more than one source of uncertainty at the same time, their risk estimates might be biased because of source dependence. This is likely to occur in “*external reference events*” approaches in which subjects have to deal with uncertainties related to both outcomes and probabilities represented through external devices.

“*Source dependence*” does not appear in another class of “*indirect methods*”, called “*internal events*”, in which people deal with magnitudes of the outcome, but not with probabilities that are not even mentioned to them. In fact, respondents are only asked to bet on disjoint subspaces of the whole state space of the variable under study. When they are indifferent between the two subspaces, then they are assumed to perceive them as equally likely (Spetzler and Von Holstein, 1975). The EM, formally described by Raiffa (1968) and recently implemented by Baillon (2008) and Abdellaoui et al. (2010), is based on this idea.

Although this technique bypasses the problem of “*source dependence*”, it has a much disputed technical drawback related to its incentive compatibility. Its peculiar experimental design, based on chained questions, is accused to undermine the incentive compatibility of the game, in particular when respondents are provided with real monetary incentives. In particular the EM’s experimental design is characterized by a set of binary questions in which the two disjoint subspaces characterizing each question depend on which subspace each respondent decided to bet on in the previous question. Given this design, respondents may behave strategically and prefer not answering in line with their real preferences to get new stimuli from the following questions (Harrison, 1986).

To our knowledge, previous investigations, using games with chained structures and real monetary incentives, have taken this issue very seriously. For instance, some of them have validated their results by using respondents’ statements of unawareness about the presence of chained questions in the game (Van de Kuilen et al., 1981 and Abdellaoui et al., 2010). In his application of exchangeability method, Baillon (2008) dealt with this problem by randomizing the order of questions and making the chaining unclear to respondents, in the sense that they are not longer aware of the relation between the disjoint subspaces they face in one question with those of the previous question.

It is clear that all previous studies, using chained games in presence of real monetary incentives, have struck at the root of the problem by avoiding the use of identifiable chained questions in their experimental designs without even thinking to investigate their presumed negative effect on people’s choice-behaviours. Hence, our study aims to test the presence of a potential “chaining effect” by comparing the validity rate of subjective risk estimates elicited via exchangeability games where the chained structure of the experimental design is hidden or not to respondents.

A very interesting point about the incentive compatibility of exchangeability method have been made by Baillon (2008), who claimed that telling the truth is the simplest and efficient strategy respondents can use when they play the game. A direct consequence of this reasoning is that respondents may not care about real monetary incentives because the game may be considered incentive compatible by itself. At this regard, in their application of exchangeability game, Abdellaoui et al. (2010) have tested the effect of real monetary incentives on people’s

[3] Taking inspiration from Heath and Tversky (1991), Baillon (2008) defined a source of uncertainty as “…a set of events that are generated by a common mechanism of uncertainty”.

choice-behaviours by comparing subjective risk estimates provided by two groups of respondents, one provided with monetary incentives and the other not. They conclude that the former group provides less noisy risk estimates than the latter group, however, the figures showing the risk estimates do not in fact illustrate much difference. In addition, given that this analysis uses a between–subjects investigation, this discrepancy of results may be due to different compositions of samples. Moreover, the assumption under which rewarded people provide more reliable estimates of the other is only a speculation, in fact the opposite may be true.

Below, we define a clear-cut valuation method to test the presumed superiority of subjective risk estimates elicited via exchangeability game when people are rewarded with real monetary incentives and when they are presented with unchained questions. Our valuation method is based on the comparison among "*validity rates*" of subjective risk estimates obtained by using various versions of exchangeability game.

4. The notion of validity rate

The "*validity rate*" is defined as the number of valid subjective risk estimates provided by people, while a valid risk measure is the one that de Finetti defines as "*coherent*", in the sense that it obeys to all axioms and theorems of probability theory (de Finetti, 1937, 1974)[4]. The choice of using the de Finetti's notion of "*coherence*" to define valid risk measures is not groundless, but relies on the fact that "exchangeability method" is based on the assumption of "*exchangeability-based probabilistic sophistication*" (Chew and Sagi, 2006), that in turn is based on the notion of "*exchangeability events*" (de Finetti, 1937).

The idea of "*Probabilistic sophistication*" as originally formulated by Machina and Schmeidler (1992) implies that decision makers' choices depend upon their probabilistic beliefs. Assuming a lottery $L = (E_1{:}x_1,\ldots, E_n{:}x_n)$ where a decision maker achieves the monetary outcome x_i only if the event E_i happens, she/he is "*probabilistically sophisticated*" when her/his preferences over lotteries are only determined by the subjective probabilities of occurrence she/he attaches to each event E_i. This notion of probabilistic sophistication relies on axioms of Savage's subjective utility theory (SEU). Chew and Sagi (2006) provide a much more intuitive notion of probabilistic sophistication based on the idea of "*exchangeable events*" by de Finetti (1937). *Exchangeability*, as formulated by de Finetti, implies that the probability that each event belonging to the set, occurs is the same without depending on the order of the events, but only on the number n of events. Hence, even the joint probability of all events belonging to a set of n events is always the same and does not depend on the order of the events (de Finetti, 1937).

Based on de Finetti's idea of equal likelihoods of exchangeable events, Chew and Sagi defined two events as "*comparable*", under a probabilistic point of view, only when a sub-event of one is exchangeable with the other event. This way of comparison is intuitively straightforward considering that a sub-event is logically less likely than the event in which it is contained. It is clear that, under "*exchangeability-based probabilistic sophistication*", the choice-behaviour of the decision maker is based on probabilistic beliefs deriving from the notion of exchangeable events. In conclusion, for a probabilistically sophisticated respondent playing exchangeability game, two disjoint sub-events are "*exchangeable*", and thus they have the same

[4] de Finetti's (1937, 1974) definition of "*coherence*" is related to the notion of probability. We extend his definition to the notion of risk because we define risk as the probability that a given event (or a given magnitude of an event) occurs.

probability of occurrence, when she/he is indifferent in betting on one sub-event rather than on the other one.

Given this theoretical background, it is clear that "*exchangeability*" assumption enforces that probability estimates elicited via the EM are valid measures. In fact, all definitions, axioms and theorems of probability theory are satisfied under the assumption of "*exchangeability*" implying that probabilities of each disjoint event G_j^i are equal (see Appendix A). Therefore, we aim to test the validity of subjective risk estimates elicited via exchangeability method by investigating whether respondents' choice behaviours are consistent with the "*exchangeability*" assumption or not.

5. Hypotheses

We study the validity of subjective risk estimates by testing whether subjects' choice behaviours satisfy "*exchangeability*" assumption or not. In particular, considering two disjoint sub-events are exchangeable when the probability related to the occurrence of one is equal to the probability of occurrence of the other.

H_0: $P(G_j^i) = P(G_j^k), \forall k \neq i, k \leq n$

H_1: $P(G_j^i) \neq P(G_j^k), k \neq i, k \leq n$

Specific tests for exchangeability will be presented in Section 7.1.

After having investigated the validity of risk estimates elicited via exchangeability method, we test whether diverse experimental settings affect the "*validity rate*" (V) of subjective risk measures. We remind that the "*validity rate*" is the number of subjective risk measures that satisfy the "*exchangeability* assumption" and, thus, all axioms and theorems of probability theory. In particular we test the following hypotheses:

a. The provision of real monetary incentives to respondents does not affect the "*validity rate*", that is the number of valid risk measures. In this case, the "*validity rate*" of subjective estimates of risk is equal when respondents are provided with real monetary incentives (V_m) and when they are not (V_h).

H_0: $V_m = V_h$

H_1: $V_m \neq V_h$

b. The awareness of the chaining structure of the exchangeability game does not influence the "*validity rate*", that is the number of valid risk measures. In this case, the "*validity rate*" of subjective estimates of risk is equal when respondents are aware of the chaining structure of the game (V_c) and when they are not (V_u).

H_0: $V_c = V_u$

H_1: $V_c \neq V_u$

c. The provision of real monetary incentives to respondents who are aware of the chained structure of the game does not affect the "*validity rate*", that is the number of valid risk measures. In this case, the "*validity rate*" of subjective estimates of risk when respondents are provided with real monetary incentives and when respondents are not aware of chaining structure of the game (V_{mu}) is equal to the validity rate when respondents are provided with real monetary incentives, but they are aware of chaining structure of the game (V_{mc}).

H_0: $V_{mc} = V_{mu}$

H_1: $V_{mc} \neq V_{mu}$

6. The specific application and experiment.

This study covers the investigation of subjective risks related to fire blight, a bacterial disease threatening apple orchards in the Province of Trento since 2005 (IASMA, 2006). This phytopathology damages and kills apple plants provoking losses in the production of apples, one the most important agricultural products of this Italian province. Scientists at the Edmund Mach Foundation predicts a future spread of the disease in many apple orchards of the Province of Trento because suitable climatic conditions for the biology of the bacterium *Erwinia amylovora* are likely to occur in future.

Italian farmers currently control Fire Blight and its negative consequences on apple production by only using a small number of not much toxic pesticides. Given that these measures might be not efficient enough to prevent the future spread of fire blight, farmers are expected to start implementing new adaptation strategies against fire blight. The only strategy that is currently available to farmers is the introduction of new active principles for control of fire blight such as the antibiotic streptomycin that is currently forbidden by the Italian legislation, but that has been already used in U.S., and other European countries for controlling fire light (personal communication from Edmund Mach Foundation).

In the context presented above, we decided to investigate three diverse random variables: the percentage (or number) of days in which the infestation will occur during the blossoming period in 2030 (g), the number of apples containing residues in a sample of 100 apples in 2030 (a), and the number of apples containing more than one residue in a sample of 100 apples in 2030 (r). These variables have been selected after having interviewed approximately 20 focus group subjects.

The sample of laboratory subjects consists of 100 individuals who were randomly recruited outside the main supermarkets of Trento and asked to come in the experimental laboratory of the University of Trento under a compensation of 25€ (show-up fee). Selected participants are divided in four subsamples that we call treatments, each is characterized by a different experimental design: “real incentives-unchaining”, “real incentives-chaining”, “hypothetical incentives-unchaining”, and “hypothetical incentives-chaining”.

In the “hypothetical incentives” treatments, respondents are not provided with real monetary incentives in addition to the show-up fee, while, in the “real incentives” ones, they are told that one randomly selected individual from each group has the chance to win additional 100€ based on her/his choices during the experiment. In particular they are informed that one respondent in each treatment will be randomly selected at the end of the experiment as well as one of the questions she/he will answer during the experiment. The participant, who might win the additional 100€, will be selected through the draw of a numbered chip from a cage that contains as many numbered chips as the number of computer positions that are in the laboratory. Each numbered chips is related to a numbered computer position. The question, that might make the drawn respondents earning, will be selected through the draw of a numbered chip from a cage that contains as many numbered chips as the number of questions that the respondent answer during the experiment. Each numbered chips is related to a numbered question. The drawn participant win the additional 100€ if and only if the event she/he had chosen in the drawn question is consistent with the simulation about the random variable under study provided by the Edmund Mach Foundation.

In the “chained” treatments, respondents play the exchangeability game facing chained questions where the outcomes presented in one questions clearly depend on their choice in the previous one. In particular, we ask them to answer questions that allow us to elicit the percentiles of their CDFs in the following order: $g_{1/2}$, $g_{1/4}$, $g_{3/4}$, $a_{1/2}$, $a_{1/4}$, $a_{3/4}$, $r_{1/2}$, $r_{1/4}$, and $r_{3/4}$. In the "unchained" treatments, this chained structure of the game is hidden through a particular randomization of questions. In fact, we elicit the percentiles of respondents'' CDFs in the following order: $g_{1/2}$, $a_{1/2}$, $r_{1/2}$, $g_{1/4}$, $a_{1/4}$, $r_{1/4}$, $g_{3/4}$, $a_{3/4}$,and $r_{3/4}$.

Analyzing three different random variables, it follows that each respondent, whatever treatment she/he belongs to, plays exchangeability games and lotteries three times, one for each random variable under study.

7. Methodology

The assessment of the validity rate of subjective risk estimates for each treatment of our experiment represents the first step to take for investigating the presence of "chaining effect" and "real incentive effect". The validity rate is the number of valid subjective estimates provided by respondents, which is the number of individual estimates satisfying exchangeability. Since the exchangeability is a crucial requirement for identifying valid subjective risk estimates, we implement three diverse tests to investigate whether respondents' behaviours are consistent with this notion. While one of them, say Test 3, only investigates whether exchangeability is satisfied for each respondent's risk estimate within the sample, the others, say Test 1 and Test 2, also checks whether exchangeability is globally satisfied in the whole sample.

Then, we test the presence of "chaining effect" and "real incentive effect" by using different techniques. First, we compare validity rate of subjective risk estimates obtained for each particular experimental setting characterizing each treatment. Validity rates are assumed to be proxies of the goodness of the used experimental setting. Second, we econometrically test "chaining effect" and "real incentive effect" by estimating a model in which the discrete dependent variable captures the validity of subjective risk estimates and these effects are captured by dummy variables representing the characteristics of each experimental setting. The estimation of this model takes into consideration the panel nature of our dataset.

7.1 Exchangeability tests

The first test (Test 1) is implemented by eliciting a new measure of the median value of individual CDFs, say $g_{1/2}$', through a second round of exchangeability game. This round differs from the first one because the lower and upper bounds of the event space are not anymore g_0 and g_1, but the subjective estimates of the quartiles $g_{1/4}$ and $g_{3/4}$ elicited via the first round of EM. This investigation tests whether the assumption of exchangeability is satisfied or not by comparing the subjective estimates of $g_{1/2}$ and $g_{1/2}$'. In particular, if they are equal, then exchangeability is satisfied, otherwise not:

H_0: $g_{1/2} = g_{1/2}'$

H_1: $g_{1/2} \neq g_{1/2}'$

We test this hypothesis both at sample level and at individual level. In the first case, we use nonparametric tests such as Mann-Whitney U (MWU) and Kolmogorv-Smirnov (KS) tests, in the second case, we simply check if $g_{1/2}' = g_{1/2}$.

The second test (Test 2) is based on the notion of "*certainty equivalent*" (CE) defined as the sure amount of money that makes people indifferent to gamble. Respondents are presented with choice tasks where they are asked to choose between a lottery, in which they win a monetary outcome x if the real outcome G_j^i will happen in the future (or a null monetary outcome otherwise), and a sure payment z. According to "*exchangeability*" assumption, the certainty equivalent respondents are willing to accept in order to giving up the possibility to play these lotteries should be equal for those characterized by real outcomes G_j^i that have been judged to be equally likely by respondents themselves during the first round of EM. For this reason, this test of exchangeability checks the following hypotheses:

H_0: $CE[L(x:G_j^i)] = CE[L(x:G_j^k)]$, with $k \neq i$, $k \leq j$

H_1: $CE[L(x:G_j^i)] \neq CE[L(x:G_j^k)]$

We test this hypothesis both at sample level by using Mann-Whitney U (MWU) and Kolmogorv-Smirnov (KS) tests, and at individual level by checking if $CE[L(x:G_j^i)] = CE[L(x:G_j^k)]$.

The <u>third test</u> (Test 3) is based on the creation of a multiple choice card where respondents, for each choice task, have the possibility to choose between lotteries that they have implicitly defined to be equivalent according to their choices during the first round of EM. A crucial point is that respondents have also the possibility to show their indifference among these lotteries by choosing neither. In each choice task, respondents are presented with two lotteries, both characterized by a monetary outcome x and by a real outcome G_j^i. Respondents gain the same monetary outcome x if the real outcome G_j^i of the lottery they have chosen will happen in the future or a null monetary outcome otherwise. Under the assumption of "exchangeability", respondents are assumed to be indifferent between lotteries characterized by real outcomes which have been defined as equally likely by respondents themselves during the first round of exchangeability game. In fact those lotteries are equivalent, in the sense that they exactly provide the same payoff. Hence, exchangeability holds if and only if respondents decide to bet on neither confirming their indifference between the two prospects they have shown in the first round of EM. In particular, we test the following hypotheses:

H_0: $L(x:G_j^i) \approx L(x:G_j^k)$, with $k \neq i$, $k \leq j$

H_1: $L(x:G_j^i) \not\gtrless L(x:G_j^k)$

This hypothesis can be tested only at individual level by simply checking if $L(x:G_j^i) \approx L(x:G_j^k)$.

7.2 Chaining effect and real monetary incentive effect

The hypotheses related to "chaining" and "real incentive" effects are tested by using two diverse approaches; while the first is very much based on intuition, the second is based on econometric analyses.

In the first approach, we simply calculate the validity rate of subjective risk estimates that is the number of valid risk measures elicited via exchangeability method for each treatment of the experiment. We assume that subjective risk estimates are valid if and only if the following null hypotheses are not rejected:

a. H_0: $g_{1/2} = g_{1/2}$' at sample and individual levels
b. H_0: $CE[L(x:G_j^i)] = CE[L(x:G_j^k)]$, with $k \neq i$, $k \leq j$ at sample and individual levels
c. H_0: $L(x:G_j^i) \approx L(x:G_j^k)$, with $k \neq i$, $k \leq j$ at individual level

The treatment obtaining the highest validity rate represents the most efficient experimental design that can be used to elicit subjective risk estimates via exchangeability method, and its characterization in terms of chained/unchained structure and real/hypothetical incentives allows us to investigate the presence of "chaining" and "real incentive" effects.

In particular, if the "real incentives-unchaining" treatment obtains the highest validity rate, we can conclude that, as the literature in economic experiments predicts, real monetary incentives increases the reliability of risk estimates. Moreover, as predicted by Harrison (1986), chained structures of experimental games create problems for risk elicitation, especially if these structures are combined with a system of incentives based on real payments.

Assuming that the highest validity rate is obtained by "real incentives-chaining" treatment, we can conclude that, real monetary incentives improves reliability of results, and their combination with chained structures does not have the presumed negative effect predicted by Harrison (1986).

The hypothesis under which real incentives increase the reliability of obtained risk estimates is rejected if "hypothetical incentives-unchaining" and "hypothetical incentives-chaining" treatments obtain the highest validity score. Moreover, we might claim that EM is incentive compatible by itself, and that real incentives have a misleading effect on respondents. However, in case "unchaining" treatment gets the highest validity rate, we deduce that the chained structure of games negatively affects the validity of results as predicted by the literature in economic experiments, while, in the other case, we argue that chained structure of EM helps respondents to provide better risk estimates.

In the second approach, we build a model in which the discrete dependent variable, which describes if subjective risk estimates are valid or not, is explained by a set of dependent variables taking into account socio-economic characteristics of respondents, their attitude towards the consumption of apples and food contamination, and, the treatment they belong to. The last information is captured by a set of dummy variables in the sense that each treatment is represented by a dummy variable which takes value one if the respondent belong to this treatment and value zero otherwise.

We estimate this model taking into account both the discrete nature of the dependent variable under study and the panel nature of our data set (random and fixed effects logit models). In fact, each respondent provides us three different estimates, $g_{1/4}$, $g_{1/2}$, and $g_{3/4}$, that are basically provided in three different times.

The statistical significance and the signs obtained by coefficients of dummy variables interest our analysis, because they allow us to understand which characteristics of the experimental design affect the probability that subjective risk estimates are valid or not. The interpretation of results is very much similar to that described for the first approach presented in this paragraph.

8. Expected results

We aim to define a clear-cut valuation method to investigate the validity of subjective risk measures elicited via exchangeability method. This method allows us to investigate whether real monetary incentives and chained questions affect the validity of risk estimates.

To our knowledge, previous investigations using EM have been always conducted by using real monetary incentives as suggested by the literature in experimental economics. Nevertheless, taking inspiration from a Baillon's sentence (2008), we hypothesize that real monetary incentives do not really matter in EM because the simplest strategy respondents can use to play the game is telling the truth.

If this is the case, the presumed negative effect of the chained structure of EM's experimental design hypothesized by Harrison (1986) should almost disappear because it is mostly due to strategic behaviours of incentivized respondents. On the contrary, we hypothesize that a negative effect of chained questions still exists because the chained structure of EM induce respondents to answer meaningless questions. In fact, the implementation of an experimental design in which the chained structure is not hidden by a randomization of questions lead respondents to face binary questions describing two prospects that they have already ruled out in previous questions.

In conclusion, we hypothesize that the highest "*validity rate*" might be reached in the groups of people who faces "unchained" experimental designs. Moreover, given that telling the truth seems to be the respondents' easiest way of playing the game, we expect that real monetary incentives do not have any effect on the reliability of estimates. In conclusion, we predict that "real monetary incentives-unchained questions" and "hypothetical monetary incentives-unchained questions" will obtain the highest "*validity rates*".

9. Conclusion

In the last few years, scholars investigating environmental and health risk perceptions have started looking at "*indirect methods*" of risk elicitation. The most used techniques relying on the use of "*external reference events*", but recent works on risk elicitation have re-discovered methodologies based on "*internal events*", an example is the Exchangeability Method (EM). Although this technique seems to overcome some limitations of "*external reference events*"-based techniques, it suffers from other limitations. In this paper we focus on very technical and methodological issues related to the EM by using the very intriguing statistical findings by the Italian mathematician Bruno de Finetti. In particular we investigate the effect of real monetary incentives and chained experimental designs on the validity of subjective risk estimates elicited via EM. We expect to obtain empirical data from our lab experiment suggesting that, while real incentives do not play any role since the simplest strategy respondents can use to play the game is telling the truth (Baillon, 2008), chained questions might have a negative effect on the validity of risk estimates. This work is one of the few attempts to investigate peoples' perceptions of agricultural risks such as food contamination by using a laboratory experiment and real monetary incentives.

References

Abdellaoui M, Baillon A, Placedoa L, Wakker PP (2010) The Rich Domain of Uncertainty. Amer Econ Rev (forthcoming)

Andersen S, Fountain J, Harrison GW, Rutström (2009) Estimating Subjective Probabilities. Working Paper 09-01, Department of Economics, College of Business Administration, University of Central Florida

Baillon A (2008) Eliciting Subjective Probabilities Through Exchangeable Events: an Advantage and a Limitation. Decis Anal 5(2):76-87

Baker J, Shaw WD, Riddel M, Woodward RT (2009) Explaining Changes in Subjective Hurricane Risks as Time Passes: An Analysis of a Sample of Katrina Evacuees. J Risk Res 12(1):59-74

Botzen, WJW, van den Bergh JCJM, (2009) Monetary Valuation of Insurance against Climate Change Risk. Paper presented at FUR Conference

Cerroni S, Shaw WD (2011) Does Climate Change Information Affect Elicited Risks Related to Pine Beetle Impacts on Forests? An Application of the Exchangeability Method. Revised discussion paper (currently unpublished), University of Trento.

Chew SH, Sagi J (2006) Event exchangeability: Probabilistic sophistication without continuity or monotonicity. Econometrica 74: 771–786 de Finetti B (1937) La prévision: Ses lois logiques, ses sources subjectives. Annales de l'IHP 7:1–68

de Finetti B (1974) The Value of Studying Subjective Evaluations of Probability. In: The Concept of Probability in Psychological Experiments Stael Von Holstein C.-A.S, D. Reidel Publishing Company, Dordrecht-Holland/Boston-U.S.A. pp. 1-14

Fiore SM, Harrison GW, Hughes CE, Ruström EE (2009) Virtual experiments and environmental policy. J Environ Econ Manage 57 (1): 65-86

Harrison GW (1986) An experimental test for risk aversion. Econom. Lett. 21 7–11

Jakus PM, Shaw WD, Nguyen TN, Walker M (2009) Risk Perceptions of Arsenic in Tap Water and Bottled Water Consumption. Water Resour. Res 45. Doi:10.1029/2008WR007427

Kilka M, Weber M (2001) What Determines the Shape of the Probability weighting function under uncertainty. Management Sci 47:1712–1726

Machina MJ and Schmeidler D (1992) A More Robust Definition of Subjective probability. Econometrica 60: 745-780

Offerman T, Sonnemans J, Van de Kuilen G, Wakker PP (2009) A Truth Serum for NON-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes. Review of Economic Studies 76: 1461-1489

Raiffa H (1968) Decision Analysis. Addison-Wesley, London.

Riddel M, Shaw WD (2006) A Theoretically-Consistent Empirical Non-Expected Utility Model of Ambiguity: Nuclear Waste Mortality Risk and Yucca Mountain. J Risk Uncertainty 32(2): 131-150

Spetzler CS, Stael Von Holstein C.-A.S (1975) Probability encoding in decision analysis. Management Sci. 22: 340–358.

Tversky A, Koehler DJ (1994) Support Theory: A Nonextensional Representation of Subjective Probability. Psychological review 4: 547-567

Van de Kuilen G, Wakker PP, Zou L (2006) A Midpoint Technique for Easily Measuring Prospect Theory's Probability Weighting.

Appendix A. Definition, axioms and theorems of probability theory

Let G_j^i be disjoint events with $i = \{1,..,n\}$ and $j = n$ and S_G be a sample space, then:

Statement 1. $P(G_j^i) \geq 0$

Statement 2. $P(S_G) = 1$

Statement 3. If $\{G_j^i\}$ is a sequence of disjoint sets in S_G, then

$$P\left(\bigcup_{i=1}^{n} G_j^i\right) = \sum_{i=1}^{n} P(G_j^i)$$

Statement 4. $P(G_j^i) = 1 - P\left({G_j^i}^C\right)$

Statement 5. $P(\phi) = 0$

Statement 6. For each $G_j^i \in S_G$, then $0 \leq P(G_j^i) \leq 1$

Statement 7. If $G_j^i \subset G_n^i$ with $n = jk, k \in N, k \neq 0$, then $P(G_j^i) \geq P(G_n^i)$

Figure 1: Structure of the experimental design