Investors’ risk attitude and risky behavior: a Bayesian approach with imperfect information

by Stefano Iezzi
The purpose of the Temi di discussione series is to promote the circulation of working papers prepared within the Bank of Italy or presented in Bank seminars by outside economists with the aim of stimulating comments and suggestions.

The views expressed in the articles are those of the authors and do not involve the responsibility of the Bank.


Editorial Assistants: Roberto Marano, Nicoletta Olivanti.
INVESTORS’ RISK ATTITUDE AND RISKY BEHAVIOR: A BAYESIAN APPROACH WITH IMPERFECT INFORMATION

by Stefano Iezzi*

Abstract

In a choice model of risky assets the role of risk aversion is analyzed. The measure of risk preference comes from a direct subjective survey question and it is considered as an imperfect information about the true risk attitude of investors. Misclassification between the true and the observed risk aversion is explicitly taken into account in the empirical model. A Data Augmentation approach, a Bayesian procedure for incomplete-data problems, is applied on data from the 2006 Survey of Household Income and Wealth by the Bank of Italy. Results indicate that when misclassification of investors is taken into account model estimates show the good performance of the subjective question when used as a control in a portfolio choice models. Moreover risk aversion emerges as a strong predictor of the probability to hold risky assets. The analysis also shows that probability of misclassification decreases as latent risk aversion increases, that means that more risk tolerant investors tend to be classified erroneously more often than less risk tolerant investors.

JEL Classification: C11, C25, G11.
Keywords: portfolio choice, risk attitude, misclassification error, Bayesian analysis.

Contents

1. Introduction ................................................................. 3
2. The model specification .................................................. 5
   2.1 The demand model for risky assets ................................ 5
   2.2 The assumption of independent misclassification .......... 6
   2.3 The assumption of misclassification dependence on explanatory variables ............ 8
3. A Bayesian approach to model estimation ..................... 9
   3.1 The DA algorithm ................................................... 10
   3.2 Posterior analysis .................................................. 10
4. Empirical analysis ...................................................... 12
   4.1 The data .................................................................. 12
   4.2 The model estimate ................................................ 13
   4.3 Results ................................................................. 14
5. Concluding remarks .................................................. 17
References ................................................................. 25

* Bank of Italy, Economics, Research and International Relations.
1 Introduction

This paper applies discrete econometric Bayesian techniques to microeconomic data on household portfolio composition taken from the Bank of Italy’s 2006 Survey of Household Income and Wealth, in order to shed light on the role of risk aversion in the individual choice to hold risky assets.

Empirical analyses of the portfolio choices of households or individuals indicate that observed choices are often inconsistent with standard asset allocation models. As a consequence, several studies have focused on empirical failures of portfolio theory. The greatest failure is perhaps the fact that the majority of individuals do not hold fully diversified portfolios, although the percentage of households holding risky assets has increased over the last decade (Haliassos and Hassapis, 1998).

Although the rational model of choice is unable to explain several empirical phenomena, it is often hard to determine in more detail what the underlying cause of disparities between theory and empirical facts may be. For this reason some authors have turned to more direct, subjective evidence on preferences to reduce the distance between theory and empirical facts. A prominent example is the role of risk preference.

In order to provide evidence on the role of risk aversion one needs to be able to measure it at the individual level, but individual risk aversion is not normally observable. One way to estimate the degree of risk preference is to utilize survey questions concerning hypothetical choices between uncertain income streams. Guiso and Paiella (2001) use a sample of 8,135 heads of households from the Survey on Household Income and Wealth and measure risk preferences with an abstractly-framed, hypothetical lottery. Diaz-Serrano and O’Neill (2004) use the same sample but also add the next wave of the survey. Donkers et al. (2001) use a sample of 4,000 individuals living in the Netherlands and measure risk preferences with a series of abstract lotteries. Barsky et al. (1997) use an especially large sample, 14,000 individuals living in the US drawn from the Health and Retirement Survey, and measure risk preferences using a hypothetical lottery involving different income streams.

However, all these measures have theoretical and empirical problems. First of all, the wording of the question is usually quite complicated and many respondents may have a hard time understanding the exact meaning, thus producing considerable missing values or scarce reliability. Secondly, the answer is conditioned by the respondent’s current situation. For instance, a risk tolerant individual with a risky portfolio may be induced to choose a safe income stream since he is already exposed to considerable risk. Conversely, a risk averse individual with a safe portfolio can afford to choose a riskier income path. In both cases the observed relationship between the measured risk tolerance and portfolio choice is attenuated (Kapteyn and Teppa, 2002).

Dohmen et al. (2005) use a new set of survey measures, collected for a representative sample of 22,000 individuals: one question is about the attitude towards risk in general, five questions refer to risk attitude in specific contexts (car driving, financial matters, leisure and sport, career, and health), and the last question corresponds to the lottery measures used in previous studies. They find that the general risk question is the best all-round predictor of risky behavior, outperforming a lottery measure

---

1 The author wishes to thank Giovanni D’Alessio, Leandro D’Aurizio, Ivan Faiella, Massimo Sbracia, Alessandro Secchi and Luigi Federico Signorini for their very useful comments.
All the measures of risk aversion are unavoidably afflicted with errors that, in some cases, may lead to serious misclassification of investors. In general, estimated associations between an outcome variable and a mismeasured covariate tend to be biased when the methods of estimation that ignore the classification error are applied. The identification and estimation of models that are non-linear in mismeasured variables is a notoriously difficult problem (Carroll, Ruppert and Stefanaski, 1995). There are three possible approaches to this: (1) the parametric approach; (2) methods that use additional sample information, such as a validation sample or replicate measurements; (3) the instrumental variable method.

The parametric approach makes strong and untestable distributional assumptions. In particular, it is assumed that the distribution of the measurement error is in some parametric class (Hsiao, 1989).

The second approach is to use additional sample information in the form of replicate measurements or in the form of a validation sample. The sample contains replicate measurements if there are at least two mismeasured variables that correspond to the same latent true value. Li and Vuong (1998) show that if the measurement errors in the two measurements are stochastically independent, the distribution of the latent true value is non-parametrically identified.

The third approach is the instrumental variable method. In an errors-in-variables model, a valid instrument is a variable that (a) can be excluded from the model, (b) is correlated with the latent variable, and (c) is independent of the measurement error. The IV method was developed for linear models, and, in general, IV estimators are biased in nonlinear models (Amemiya, 1985). However Carroll and Stefanaski (1990) obtain a consistent IV estimator in nonlinear models under the assumption that the measurement error vanishes if the sample size increases. Newey (2001) considers the nonlinear regression model, but he notes that there are no general results on the non-parametric identification of nonlinear models with mismeasured regressors by instrumental variables.

When the mismeasured covariate is categorical, measurement error in nonlinear models introduces difficulties whose solutions require techniques that are quite different from those usually called for in linear models. The independence assumption between the measurement error and the latent variable, invoked by the classical model for measurement error, is particularly untenable. More generally, the phenomenon of negative correlation between the errors and the true values (referred to as "mean reversion") has been found to exist for a number of quantities of interest, for example in earnings data by Bound and Krueger (1991) and Bollinger (1998).

This paper exploits a subjective measure of risk preference in a choice model of risky assets. The available measure of risk preference is assumed to be an imperfect information so that in the empirical model the true risk attitude variable is considered as a latent class variable. A Bayesian approach through a Monte Carlo Markov chain procedure is developed in order to estimate the empirical model of ownership of risky assets where the latent risk attitude variable is one of the determinants. The key idea of the approach is to introduce the values of the latent variable explicitly into the estimation process as missing data. The problem thus involves two types of unknown quantities, the parameters to be estimated and the missing data, unobserved because of measurement error. This distinction
seems well suited to misclassification problems where it simplifies the analysis both conceptually and computationally with respect to other methods of estimation.

This paper is organized as follows. Section 2 describes the choice model of risky assets and the distributional assumptions needed for its identification. Section 3 illustrates the Bayesian procedure for model estimation. Section 4 describes the data-set, the model estimated and interprets the results. In section 5 the main conclusions are drawn.

2 The model specification

In this section I first present a standard model of risky assets ownership where the role of risk aversion is emphasized (section 2.1). Next, under the hypothesis that the risk aversion variable is affected by misclassification, I introduce three basic assumptions in order to achieve model identification (section 2.2). In particular, one of the assumptions implies that the misclassification probability is independent from any other explanatory variables. This is a strong but necessary assumption when only one surrogate for the latent variable is available. Finally, the model is developed by relaxing the assumption of independent misclassification under the hypothesis that two surrogates for the latent risk aversion are available (section 2.3).

2.1 The demand model for risky assets

My interest is in household ownership of risky assets. I do not observe the value of the desired level of risky assets \( y_i^* \) for each household \( i \) but I know whether households hold risky assets and I want to study the ownership probabilities \( \Pr (y_i = 1) \) where

\[
y_i = \begin{cases} 
1 & \text{if } y_i^* > 0 \\
0 & \text{if } y_i^* \leq 0 
\end{cases}
\]

Now let’s assume that the desired level of risky assets is a linear function of a set of exogenous variables \( z_i \):

\[
y_i^* = \gamma' z_i + \varepsilon_i \tag{1}
\]

where \( \gamma \) is its associated vector of parameters and \( \varepsilon_i \) is the error term. Therefore our model can be rewritten as:

\[
y_i = \begin{cases} 
1 & \text{if } \varepsilon_i > -\gamma' z_i \\
0 & \text{if } \varepsilon_i \leq -\gamma' z_i 
\end{cases}
\]

I assume \( F() \) is the cumulative distribution function of the random variable \( \varepsilon_i \). This implies

\[
E (y_i|z_i) = \Pr (y_i = 1|z_i) = F (\gamma' z_i) \tag{2}
\]

Among the exogenous variables \( z_i \), an important role is played by the risk preference variable. For the sake of simplicity I consider only two groups of investors characterized by two degrees of risk attitude:
\( I_i = 0 \) when household \( i \) has a low aversion to risk; \( I_i = 1 \) when household \( i \) has a high aversion to risk. Generalization to more than two groups of investors is straightforward.

Model (2) can then be expressed as:

\[
E \left( y_{i|z_i} \right) = \Pr (y_i = 1|x_i, I_i) = F \left( \beta^\prime x_i + \delta I_i \right)
\]

where \( x_i \) denotes sociodemographic and economic characteristics of household \( i \) and \( \delta \) measures the marginal effect of the degree of risk aversion on the desired level of risky assets.

Suppose I know exactly whether household \( i \) comes from the low risk aversion group or the high risk aversion group. In this case I have a perfect sample separation information and equation (3) can be estimated straightforwardly.

### 2.2 The assumption of independent misclassification

I now consider the situation where information about individual risk aversion is available but imperfect. In particular, suppose I know a dichotomous variable \( w_i \) for each household \( i \) resulting from the known sample separation information. In this case, \( I_i \) is the latent true discrete variable which is subject to misclassification error, and \( w_i \) is known as a "surrogate" for \( I_i \). Following the current literature, the indicator \( w_i \) can be obtained from survey subjective questions regarding household risk preferences.

I first consider the problem in a non-parametric setting with no functional form assumption for the binary choice model. I do, however, place two restrictions on the nature of the measurement error. The first assumes that the outcome \( y \) is independent of \( w_i \) conditional on the correctly measured random variable \( I_i \) and the other explanatory variables \( x_i \). Formally:

\[
\Pr (y_i = 1|I_i, w_i, x_i) = \Pr (y_i = 1|I_i, x_i) \tag{Assumpt. 1}
\]

In the literature this is usually referred to as the assumption of non-differential measurement error, and implies that the measurement error is uninformative about the response given information on the truth. The conditional statement is important since the misclassification rates may in fact be informative about responses through their correlation with other variables in the model.

The second restriction limits the extent of the measurement error by requiring that the probability of a correct classification be greater than that of an incorrect one, i.e.,

\[
\Pr (w_i = 1|I_i = 1, x_i) > \Pr (w_i = 1|I_i = 0, x_i) \tag{Assumpt. 2}
\]

This ensures that the unobserved variable \( I_i \) is positively connected with its surrogate \( w_i \). This basically means that the direction of the effect of the surrogate on the response \( y_i \) is the same as the effect of the latent variable.

Manski and Tamer (2002) show that if \( x_i \) are binary variables there is no finite bound for \( \beta \) at all, whereas if the support of \( x_i \) is unbounded \( \beta \) is point identified. In contrast, the sign of \( \delta \) is always identified regardless of the assumption on the support of \( x_i \). This implies that model (3) under Assumpt. 2 is only partially identified. We need to add one more assumption to achieve identification.
The assumption is that the probability of misclassification is independent of the other explanatory variables in the model. Formally:

\[ \Pr (w_i = 1 | I_i, x_i) = \Pr (w_i = 1 | I_i) \quad \text{(Assumpt. 3)} \]

Given that the relationship between the latent variable and its surrogate is defined by the following transition probability matrix:

\[
\begin{array}{c|cc}
I_i & w_i = 1 & w_i = 0 \\
\hline
1 & p_{11} & p_{10} \\
0 & p_{01} & p_{00}
\end{array}
\]

where \( p_{rs} = \Pr (w_i = s | I_i = r) \), with \( r, s = 0, 1 \), is the probability that a household belonging to group \( r \) records \( w_i = s \), Assumpt. 3 implies that the misclassification rates are characterized completely by the two constants \( p_{01} \) the "false positive" rate and \( p_{10} \) the "false negative" rate. Moreover, under Assumpt. 2 the condition

\[ p_{01} + p_{10} < 1 \]

is satisfied.

In such a way, misclassification probability is treated as additional unknown parameters to be estimated in the model. The identification of the parameters of interest together with \( p_{01} \) and \( p_{10} \) is proved by contradiction in Mahajan (2006) and hence is not constructive in the sense that it leads directly to an estimator. In this paper I construct an estimator by following a Bayesian approach.

I assume that the probability that a household has a high aversion to risk is constant and equal to \( \Pr (I_i = 1) = \lambda \), where \( \lambda \in (0, 1) \). Then, the conditional probability function for \( y \) and \( w \), \( \Pr (y, w | x_i) \), can be easily developed by integrating out the latent variable \( I \):

\[
\Pr (y, w | x_i) = \Pr (y | I_i = 1, x_i) \cdot \Pr (w | I_i = 1) \cdot P (I = 1) + \Pr (y | I_i = 0, x_i) \cdot \Pr (w | I_i = 0) \cdot P (I = 0)
\]

that can be expressed as:

\[
\Pr (y, w | x_i) = (yF_{1i} + (1 - y)(1 - F_{1i}))[wp_{11} + (1 - w)(1 - p_{11})] \lambda +
+yF_{0i} + (1 - y)(1 - F_{0i})[wp_{01} + (1 - w)(1 - p_{01})](1 - \lambda)
\]

where \( F_{0i} \) and \( F_{1i} \) are, respectively, the probabilities for a low risk aversion investor and a high risk aversion investor to hold risky assets.

The observed values of \( y_i \) are just realizations of a binomial process with probabilities given by (3). Hence, the likelihood function is:
\[ L = \prod_{i=1}^{n} \{ F_{i} \lambda I_{i} [w_{i} p_{11} + (1 - w_{i}) (1 - p_{11})] + F_{0i} (1 - \lambda) (1 - I_{i}) [w_{i} p_{01} + (1 - w_{i}) (1 - p_{01})] \}^{y_{i}} \cdot \{(1 - F_{1i}) \lambda I_{i} [w_{i} p_{11} + (1 - w_{i}) (1 - p_{11})] + (1 - F_{0i}) (1 - \lambda) (1 - I_{i}) [w_{i} p_{01} + (1 - w_{i}) (1 - p_{01})] \}^{(1-y_{i})} \]  

(5)

Conditional on \( w_{i} = 1 \), the probability of realization \( y_{i} = 1 \) is

\[ \Pr (y_{i} = 1|w_{i} = 1, x_{i}) = F_{1i} \frac{\lambda p_{11}}{p} + F_{0i} \frac{(1 - \lambda) p_{01}}{p} \]

while, conditional on \( w_{i} = 0 \), the probability of realization \( y_{i} = 1 \) is

\[ \Pr (y_{i} = 1|w_{i} = 0, x_{i}) = F_{1i} \frac{\lambda (1 - p_{11}) (1 - p)}{(1 - p)} + F_{0i} \frac{(1 - \lambda) (1 - p_{01})}{(1 - p)} \]

where \( p = \Pr (w_{i} = 1) = \lambda p_{11} + (1 - \lambda) p_{01} \).

Finally I can define the conditional probability that the latent variable indicator is equal to 1 conditional on \((y_{i}, w_{i}, x_{i})\). Given that \( \Pr (I_{i} = 1|y_{i}, w_{i}, x_{i}) = \Pr (I_{i} = 1|y_{i}, w_{i}|x_{i}) / \Pr (y_{i}, w_{i}|x_{i}) \), from the joint probability in (4), the conditional probability that \( I_{i} = 1 \) conditional on \((y_{i}, w_{i}, x_{i})\) is

\[ \Pr (I_{i} = 1|y_{i}, w_{i}, x_{i}) = \left[ \frac{\lambda p_{11} F_{1i}}{\lambda p_{11} F_{1i} + (1 - \lambda) p_{01} F_{0i}} \right]^{y_{i} w_{i}} \cdot \left[ \frac{\lambda (1 - p_{11}) F_{1i}}{\lambda (1 - p_{11}) F_{1i} + (1 - \lambda) (1 - p_{01}) F_{0i}} \right]^{y_{i} (1-w_{i})} \cdot \left[ \frac{\lambda p_{11} (1 - F_{1i})}{\lambda p_{11} (1 - F_{1i}) + (1 - \lambda) (1 - p_{01}) (1 - F_{0i})} \right]^{(1-y_{i}) w_{i}} \]

\]  

(6)

2.3 The assumption of misclassification dependence on explanatory variables

Results achieved in the previous section are obtained by imposing three assumptions. In particular the latest one implies that the probability of misclassification does not depend upon the other explanatory variables in the model (see Assumpt. 3). However, there is a strong feeling that misclassification error of risk attitude is correlated with individual characteristics of investors, but, as underlined in the previous section, once we allow for such dependence the model is no longer identified without further assumptions.

In order to relax the assumption of misclassification independence I need to use at least one more surrogate for the latent risk attitude. Let \( \{w_{1}, w_{2}\} \) denote a couple of replicated measurements for the latent variable \( I \). I require that, conditional on the truth and the other explanatory variables, the surrogates provide no further information about the response variable. Specifically, Assumpt. 1 is modified to

\[ \Pr (y_{i} = 1|I_{i}, w_{1i}, w_{2i}, x_{i}) = \Pr (y_{i} = 1|I_{i}, x_{i}) \]  

(Assumpt. 4)
I also limit the probability of misclassification, now conditional on the other regressors, by requiring that for each replication

$$\Pr (w_{ji} = 1|I_i = 1, x_i) > \Pr (w_{ji} = 1|I_i = 0, x_i) \quad j = 1, 2 \quad \text{(Assumpt. 5)}$$

Finally, I require that the two surrogates are independent conditional on the truth and the other explanatory variables:

$$\Pr (w_1 | w_2, I, x) = \Pr (w_1 | I, x) \quad \text{(Assumpt. 6)}$$

I do not assume that the surrogates are identically distributed conditional on \((I, x)\). While Assumpt. 6 restricts the nature of the conditional dependence between the surrogates, it allows them to be unconditionally dependent one on the other. Under the above weaker assumptions, Mahajan (2006) shows that the model is point identified, and hence a Bayesian estimator can be constructed in the same fashion as in section 4.

We indicate the probability of misclassification conditional on the other regressors as \(p_{j11} (x) = \Pr (w_j = 1|I = 1, x)\) and \(p_{j01} (x) = \Pr (w_j = 1|I = 0, x), j = 1, 2\). Then the likelihood function becomes

$$L = \prod_{i=1}^{n} \{ (F_{i1} \lambda_i [w_{1i} p_{111} (x) + (1 - w_{1i}) (1 - p_{111} (x))] [w_{2i} p_{211} (x) + (1 - w_{2i}) (1 - p_{211} (x))] \\
+ F_{0i} (1 - \lambda) (1 - I_i) [w_{1i} p_{101} (x) + (1 - w_{1i}) (1 - p_{101} (x))] [w_{2i} p_{201} + (1 - w_{2i}) (1 - p_{201} (x))] \}^{y_i} \\
\cdot \{(1 - F_{i1}) \lambda_i [w_{1i} p_{111} (x) + (1 - w_{1i}) (1 - p_{111} (x))] [w_{2i} p_{211} + (1 - w_{2i}) (1 - p_{211} (x))] \\
+ (1 - F_{0i}) (1 - \lambda) (1 - I_i) [w_{1i} p_{101} (x) + (1 - w_{1i}) (1 - p_{101} (x))] \\
[w_{2i} p_{201} (x) + (1 - w_{2i}) (1 - p_{201} (x))])^{(1-y_i)}
$$

The misclassification models for the two surrogates can be expressed as:

$$p_{j11} (x) = \Pr (w_j = 1|I = 1, x) = G_j (\zeta_j x + \eta_j)$$

$$p_{j01} (x) = \Pr (w_j = 1|I = 0, x) = G_j (\zeta_j x) \quad (7)$$

where \(G\) is a link function. Note that under Assumpt. 5 the condition \(\eta_j > 0 (j = 1, 2)\) is satisfied. In this way, misclassification probability is treated as additional unknown parameters, \((\zeta_j, \eta_j) j = 1, 2\), to be estimated in the model.

### 3 A Bayesian approach to model estimation

The most common approach to the estimation of models with latent class variables is the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). The EM algorithm does not routinely produce standard errors. Standard errors can be obtained on the basis of the likelihood function, but they are accurate only when the likelihood has a smooth, quadratic shape. For complex latent class models, this is rarely the case. In addition, when parameters are estimated at or near a boundary the associated standard errors may be a poor reflection of inferential uncertainty or may be unavailable (Chung, 2003).
The Data Augmentation (DA) approach (Schafer, 1997), a Bayesian procedure for incomplete data problems, avoids three of the most prominent difficulties associated with the EM approach. First, the Bayesian procedure does not require maximization of any function that can be numerically very difficult in this case. Second, DA provides a highly flexible way of obtaining estimates and standard errors of any desired combination of parameters. Finally, desirable estimation properties, such as consistency and efficiency, can be obtained at the cost of weak conditions on prior distributions of parameters. Here, however, only vague prior information is assumed and the method is used mainly to summarize the likelihood function and to obtain maximum likelihood estimates and their standard errors from it. Estimate results, in fact, are interpreted in a purely classical way (Gelman et al. 2004).

3.1 The DA algorithm

Data Augmentation can be thought of as a Bayesian analog of the EM algorithm. Both EM and DA treat latent variables as missing data. Whereas EM computes maximum-likelihood estimates of model parameters, DA simulates random draws of parameters from their posterior distribution given the observed data. In DA the user specifies starting values for the parameters. The algorithm alternates between the following steps:

a) Simulate the missing data, i.e. the latent variables, given the current parameter estimates and the observed data. This produces a complete data set, identical to the observed data with the addition of the imputed latent variables.

b) Using the complete data set from the previous step, the prior distributions and the model, draw a new set of parameters.

Alternating between these two steps creates a Markov chain, a sequence of random variables in which the distribution of each element depends on that of the previous one. As iterations continue, however, the dependence on the starting values eventually burns off and becomes negligible, at which point the process is said to have achieved stationarity. For this reason, DA can be regarded as a special case of Gibbs sampling (Casella and George, 1992) for incomplete-data problems.

The number of DA iterations required to achieve stationarity must be large enough for the lag-$k$ autocorrelation in every parameter to die down to zero. Thus, examining time-series and autocorrelation plots that are based on the series of random parameters resulting from the DA run is a good way to get a feel for the size of parameter $k$. Once we have identified a value for $k$ we can base our inference on simulated data at lag $k$ only. In fact, the simulated data from successive iterations are correlated but parameters’ draws spaced $k$ or more cycles apart are not. Therefore, we retain the output of DA at iterations $k, 2k, \ldots, Mk$ and use these data as $M$ independent simulations.

3.2 Posterior analysis

As with all Bayesian models, one begins by postulating suitable prior distributions for all parameters $(\beta, \delta, \lambda, p_{11}, p_{01})$, and then deriving the corresponding conditional posterior distributions given the observed data.

The distribution function for the error term in (1) is assumed to be a logistic function. The most
widely used Bayesian approach to the logistic regression model is to impose a normal prior with mean $c$ and covariance matrix $\Sigma$ on parameters $\beta$ and $\delta$, which can be made “almost diffuse” by centering at $c = 0$ and setting $\Sigma = \sigma I$ for some sufficiently large variance $\sigma$. The effect of using normal priors with means of 0 is that parameter estimates are smoothed toward zero. However, since this smoothing toward zero effect is determined by the variance, it can be decreased by increasing the variances (Congdon, 2001).

In the case of independent misclassification, because $I_i$, $w_i|I_i = 1$ and $w_i|I_i = 0$ follow a multinomial distribution, it is convenient to apply three prior distributions from the Dirichlet family (Schafer, 1997) to their parameters, $\lambda, p_{11}, p_{01}$.

In order to implement the DA sampling approach I need to derive the complete conditional distributions for all parameters in the model.

From (5) it is straightforward to identify the conditional posterior distributions of parameters. For $\lambda, p_{11}, p_{01}$ the prior distributions are assumed to come from the Dirichlet family, respectively, with parameters $(\alpha_1, \alpha_0)$, $(\alpha_{11}, \alpha_{10})$, and $(\alpha_{01}, \alpha_{00})$ conventionally chosen. It follows that

$$p(\lambda|\beta, \delta, p_{11}, p_{01}, y, x, w, I) \propto A \lambda^{n_{11}+\alpha_{11}-1} (1 - \lambda)^{n_{00}+\alpha_{00}-1}$$  \hspace{1cm} (8)$$

$$p(p_{11}|\beta, \delta, \lambda, p_{01}, y, x, w, I) \propto B p_{11}^{n_{11}+\alpha_{11}-1} (1 - p_{11})^{n_{10}+\alpha_{10}-1}$$  \hspace{1cm} (9)$$

$$p(p_{01}|\beta, \delta, \lambda, p_{11}, y, x, w, I) \propto C p_{01}^{n_{01}+\alpha_{01}-1} (1 - p_{01})^{n_{00}+\alpha_{00}-1}$$  \hspace{1cm} (10)$$

where $n_1$ and $n_0$ are, respectively, the number of observations with $I_i = 1$ and $I_i = 0$, $n_{11}$ and $n_{10}$ are, respectively, the number of observations with $(w_i = 1|I_i = 1)$ and $(w_i = 1|I_i = 0)$, and $n_{01}$ and $n_{00}$ are, respectively, the number of observations with $(w_i = 0|I_i = 1)$ and $(w_i = 0|I_i = 0)$. Moreover, $A$, $B$ and $C$ include all quantities not depending, respectively, on $\lambda, p_{11}$ and $p_{01}$. Therefore, the conditional posterior densities of $\lambda, p_{11}$ and $p_{01}$ are proportional to Beta probability functions, respectively, with parameters $(n_1 + \alpha_1, n_0 + \alpha_0)$, $(n_{11} + \alpha_{11}, n_{10} + \alpha_{10})$ and $(n_{01} + \alpha_{01}, n_{00} + \alpha_{00})$.

The likelihood functions of $\lambda, p_{11}, p_{01}$ are invariant under permutations of the latent risk aversion groups. Thus, if there are no prior restrictions on the parameter space, the model cannot be identified when the sample separation is imperfect (Lee and Porter, 1984). This occurs because the names of two values assumed by the latent classes can be mutually exchangeable, then equations cannot be distinguished. From an empirical point of view, this may result in switching labels of risk aversion classes during the simulation. The non-identification problem (label switching) can be avoided in different ways. One approach is suggested by Celeux et. al. (1999) and consists in reallocating the labels of the latent classes at each iteration of the Gibbs sampler. The approach followed here is to impose Assumpt. 2 that acts like a restriction on the parameter space.

In the case of misclassification dependent on explanatory variables the probability function $\Pr(w_j = 1|I = 1, x)$ is assumed to be a logit, thus I impose a normal prior centered at zero and with a sufficiently large variance on parameters $(\zeta_j, \eta_j)$.

Regarding parameters $(\beta, \delta) = \gamma$, since $\gamma$ are assumed to be independent from each other, their conditional posterior distributions are:
\[ p \left( \gamma_k | \gamma_{\neq k}, \lambda, p_{11}, p_{01}, y, x, w, I \right) \propto L \pi \left( \gamma_k \right) \] (11)

where \( \pi \left( \gamma_k \right) \) is a Gaussian density with mean 0 and variance \( \sigma \). Since \( L \) is a product of logits and \( \pi \left( \gamma_k \right) \) is the normal density, the conditional posterior distribution is not reducible to a standard distribution, so I can adopt the Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970)\(^2\).

Now I know how to draw from the posterior of each parameter conditional on the other parameters. I combine the procedures into a Gibbs sampler for the entire set of parameters. The Markov chain estimation scheme operates as follows:

a) Start with arbitrary initial values for the parameters.

b) Simulate the missing variable \( I_i \) from the conditional distribution specified in (6), given the current parameter values and the observed data.

c) Using the complete data set from the previous step, draw a new set of parameters from the conditional posterior distributions specified in (8)-(10) for parameters \( \lambda, p_{11}, p_{01} \).

d) Using a Metropolis-Hastings step, draw a value for parameters \( (\beta, \delta) = \gamma \) from (11).

e) Return to step (b) employing the updated parameters in place of the initial values.

4 Empirical analysis

4.1 The Data

The Survey of Household Income and Wealth (SHIW), conducted by the Bank of Italy every two years, is the main source of information on the financial behavior of Italian households at the micro level. The SHIW is a sample survey conducted by means of an interviewer-administered questionnaire which contains questions on the social and demographic characteristics of household members, their incomes and consumption expenditure. Since 1987 financial assets have been recorded on a regular basis.

The survey has a two-stage design (municipalities and households), with a stratification of the primary sampling units (municipalities) by region and demographic size. Within each stratum, the municipalities in which interviews would be conducted were selected to include all those with a population of more than 40,000 inhabitants (self-representing municipalities), while the smaller towns were selected on the basis of probability proportional to size. The individual households to be interviewed were selected randomly. In the 2006 survey 7,768 households were selected in this way. Further methodological details on the SHIW are given in Banca d’Italia (2008).

In this paper I use a subjective question on financial risk preferences from the 2006 Survey of Italian Households’ Income and Wealth. The answers to this question allow me to rank individuals with respect to their risk aversion without having to assume a particular functional form for the utility function.

Respondents were asked the following question\(^3\):

\(^2\)An alternative way to produce posterior inference for logistic models is to approximate the likelihood by a normal distribution in \( \beta \) and then apply the general computational methods for normal linear models (Gelman et al. 2004).

\(^3\)The question about risk attitude is put only to the head of a household, i.e., the person who is primarily responsible for the household budget.
When managing your financial investments, would you describe yourself as someone who looks for:

1. **VERY HIGH** returns, regardless of a **HIGH** risk of losing part of your capital;
2. a **GOOD** return, with **REASONABLE** security for your invested capital;
3. a **REASONABLE** return, with a **GOOD** degree of security for your invested capital;
4. **LOW** returns, **WITHOUT** any **RISK** of losing your capital.

The answer to the question allows us to identify four groups of investors ranked from most willing to least willing to take financial risk.

A great advantage of this survey question is that it offers a direct measure of individual attitudes, avoiding the need to recover behavioral parameters by making restrictive identifying assumptions. However, this measure has some theoretical and empirical problems. One of the most serious limitations concerns the fact that the measure comes from an absolutely subjective question and there is no way to reliably assess whether their actual behavior would mimic their answers. Many economists are sceptical about the use of information based on subjective survey questions in general. In recent studies, however, subjective information on various topics has been increasingly used (Dominitz and Manski, 1997). Another unfortunate aspect of the question is that people with no assets are expected to provide strongly unreliable answers on their propensity to risk.

The large number of zero holdings for different types of risky assets made it advisable to work with just one highly aggregated category of risky financial assets, including stocks, private bonds, mutual funds, foreign public bonds and foreign equities. The share of households owning at least one of these financial assets is about 15 per cent.

Table 1 shows the proportions of households owning risky financial assets according to different risk attitudes, as measured by the subjective question in the survey. As is clear, the number of owners of risky assets increases as risk aversion decreases, with the exception of the middle groups where the percentage of households with risky assets does not decrease with risk aversion. This pattern might change significantly when controlling for other household characteristics in a complete choice model of risky assets, but, as we will see later, it is mostly due to misclassification of investors.

### 4.2 The model estimate

As explanatory variables for the choice model of risky assets I include household income (euro value scaled by 10,000), net wealth (expressed in quartiles in order to attenuate the measurement error and to model nonlinearity), some household characteristics, such as household size, number of earners (dummy that equals 1 when there are two labour income earners or more), area of residence, ownership of mortgage debt, and variables related to the household head, such as age, gender, education (college education or high-school diploma) and occupation (two dummy variables for self-employed and employee).⁴

I choose a normal prior $N(0, \sigma)$ for the covariate parameters. To investigate the impact of the choice of prior variance, I carried out sensitivity analysis by running the chains with different values

---

⁴Occupation is defined in relation to the entire household instead of the head of the household. A household is defined as self-employed when there is at least one self-employed member and the number of self-employed members is higher than the number of employees. A household is defined as employee in a similar way.
for \( \sigma \) ranging between \( 10^3 \) to \( 10^6 \). The results do not change significantly, hence I report the summary statistics based on the runs with \( \sigma = 10^3 \).

For the Bayesian procedure, 20,000 iterations of the Gibbs sampling were performed. The first 5,000 iterations were considered burn-in and the last 15,000 draws after convergence were retained from the posterior. Since our inference of the posterior distribution of parameters is based on the simulation of the posterior distribution by combining different procedures into a Gibbs sampler, I need to monitor the convergence of the chain before results across simulations have been combined. Autocorrelation plots, not reported here for brevity, show that simulations have reached stationarity and I can retain parameters’ simulations spaced 50 after convergence, for a total of 300 draws for each parameter.

4.3 Results

4.3.1 The case of perfect measurement

Tables 2 and 3 display the parameter estimates and their corresponding marginal effects for the choice model described in sections 2.1 and 2.2. In particular in column [1] the choice model of risky assets (Assumpt. 1) is estimated under the hypothesis that the risk aversion variable is perfectly measured, i.e. it is not affected by misclassification.

Estimates show that parameters have the expected sign in most cases and are significantly almost everywhere. Among others, the coefficients of age, net wealth, area of residence and education are economically the most important.

In particular, the age pattern is hump shaped, with a maximum probability of holding risky assets between the ages of 45 and 65. This result should, however, be interpreted with caution since age effect cannot be separated from cohort effects in a single cross-section.

Net wealth has a positive sign, with richer investors being more likely to hold risky assets: when net wealth increases from the second to the third quartile the probability of holding risky assets increases by only 1 percentage point, but when passing from the third to the fourth quartile it increases by 14 percentage points, thus indicating that minimum investment requirements and monetary transaction costs are important sources of costs that limit participation in the stock market (Carroll, 2000; Vissing-Jorgensen, 2002).

The education dummies imply that the likelihood of investing in risky assets is correlated with education levels, thus supporting the idea that managing risky assets is information intensive and requires a degree of intellectual ability (King and Leape, 1987; Sims, 2001). A significant difference between college education and high school is discernible.

Residents in the North of Italy are much more likely to hold risky assets than residents in the Centre, with the latter being more likely to invest in risky assets than residents in the South. The importance of the area of residence might indicate strong differences in financial development between different areas of the country.

Self-employed people are less likely to hold risky assets than employees, in line with the view that self-employed people face undiversifiable income risks. No difference between employees and people who are not employed is detectable.
Male heads of households have a much higher probability of holding risky assets with respect to female investors. Also labour income has a positive but feeble impact on the ownership of risky assets.

Family size, number of income earners and the presence of mortgage debt do not seem to impact on the ownership of risky assets.

With regard to the risk preference model [1] identifies a strong effect of risk aversion on the ownership of risky assets, but, as anticipated in the previous descriptive analysis (section 4.1), the probability of holding risky assets increases when risk aversion decreases with the exception of the middle groups.

4.3.2 The case of independent misclassification

In column [2] of Tables 2 and 3 estimate results are shown for when the subjective question in the survey is considered imperfect and the framework outlined in section 2.2 is estimated under the assumptions (Assumpt. 1), (Assumpt. 2) and (Assumpt. 3). When information drawn from the subjective question in the survey is considered imperfect, the sign, significance and the marginal effects of the parameters remain almost the same, thus underlining the good performance of the subjective question when used as a control in a portfolio choice model.

Nevertheless a strong difference emerges concerning the risk aversion parameters. The coefficients of the latent risk attitude have the expected sign, indicating that the probability of holding risky assets increases when risk aversion decreases. Moreover, risk attitude now represents the most important element for explaining differences in risky financial behavior. In particular, when investors' risk aversion is "very low", the probability of holding risky assets increases by 31 percentage points with respect to investors with a "very high" aversion to risk; when the risk aversion is "low", the probability increases by 14 percentage points, and when the risk aversion is "high" the probability increases by about 8 percentage points.

Model [2] also allows me to analyze the misclassification of investors in terms of risk attitude. Table 4 reports the estimate of the transition probability matrix and clearly shows that the probability of misclassification decreases as latent risk aversion increases, that means that more risk tolerant investors tend to be classified erroneously more often than less risk tolerant investors. In other words, according to the model over-reporting of risk aversion is more penetrative than under-reporting of risk aversion. In particular, the probability of misclassifying a "very low" risk aversion investor (over-reporting) is about 8 times that of misclassifying a "very high" risk aversion investor (under-reporting).

One possible explanation for this result is that the subjective question on risk attitude is basically factual, so it seems to naturally lead respondents to over-report their risk aversion. For instance, a risk tolerant individual, who does not hold risky assets (because his/her level of education is low, for instance), may be led to say he/she is risk averse. In this way, individuals who are more risk tolerant and do not hold risky assets may be induced to declare they are more risk averse, while individuals who are less risk tolerant would provide more reliable answers on average.

Another reason for the differential misclassification may be that more risk tolerant investors might be affected by a kind of illusion of control (Langer, 1982) that leads them to underestimate the risk taken and thus to declare themselves as more risk averse than they effectively are.
When correcting for misclassification, the average level of willingness to take financial risk does not increase since misclassification of risk tolerant investors is compensated for by misclassification of risk averse investors who represent a large part of the population (see Table 5).

4.3.3 The case of misclassification dependent on explanatory variables

Results achieved in the previous sections are obtained by imposing three assumptions. In particular one of those implies that the probability of misclassification does not depend upon the other explanatory variables in the model (see Assumpt. 3). However, there is a strong feeling that misclassification error of risk attitude is correlated with the individual characteristics of investors. As outlined in section 2.3, in order to control for such dependence I need to use at least one more surrogate for the latent risk aversion and to impose the assumptions (Assumpt. 4), (Assumpt. 5) and (Assumpt. 6).

The second available surrogate for the true risk attitude comes from the 2004 Survey on Household Income and Wealth (Banca d’Italia, 2006). As part of the 2004 sample comprised households interviewed in the 2006 survey (panel households), it is possible to use the answers to the subjective question on risk preference for a subsample of households. Unfortunately not all panel households can be used since in the 2004 survey wave, the question on risk preference was put only to households with financial assets other than bank or post office current accounts. The final subsample comprises 1,253 households. In order to control for selection bias, the subsample of analysis is post-stratified on the basis of certain individual characteristics of the respondents in order to re-weight the various segments of the population within the sample (Kalton and Flores Cervantes, 2003). This is done by aligning the characteristics of the final sample to those of the entire 2006 sample in terms of gender, age group, geographical area, occupation and ownership of risky assets. However, after poststratification, the subsample is still a little different to the entire sample, thus the estimate results must be interpreted and compared to the previous ones with caution.

Table 6 is the 4 × 4 contingency table of investors according to the risk aversion variables observed in the 2004 and 2006 waves. The two categorical variables do not appear to be strongly associated: the Gamma index equals 0.3614, and only around 50 per cent of households declare exactly the same degree of risk aversion at the two dates. This result seems to confirm the presence of significant measurement errors in the two variables.

Since the number of observations in the subsample is very limited, the two variables of the observed risk attitude are recoded into two dummy variables that value 1 when the investor reports a low or very low risk aversion, and 0 when the investor reports a high or very high risk aversion. Thus, the latent risk attitude is now considered dichotomous. For the same reasons also the dummy variables for age are reduced to just two: one for investors aged between 45 and 65 and the other for those aged over 65.

Table 7 displays the parameter estimates: the first two columns report estimate results for the part concerning the misclassification model (7); the third column shows the marginal effects’ estimate for the part concerning the demand model.

---

5 The subsample size can be even smaller as some households are discarded when the household member who answered the risk aversion question in 2004 is different to the member who answered in 2006.
Estimates for the misclassification model show that age, net wealth, gender, education and area of residence represent the most important variables for explaining misclassification of risk attitude. From misclassification models it is possible to estimate the probability of over-reporting and under-reporting the degree of risk aversion as, respectively, \( \Pr (w_i = 2 | I_i = 1) \) and \( \Pr (w_i = 1 | I_i = 2) \). Results shown in Table 8 indicate that, as also outlined in the previous analysis, over-reporting of risk aversion is much more penetrative than under-reporting: 30 per cent of households who declare they are high risk averse are low risk averse, while 15 per cent of those who declare they are low risk averse are high risk averse. Over-reporting has differentiated patterns according to household characteristics: it is more prominent for older, male, residents in the south or islands, less educated and not employed people. By contrast, under-reporting of risk aversion does not vary much with household characteristics. Note also that over-reporting of risk aversion is higher for households owning risky assets, thus indicating that the hypothesis of investors’ illusion of control might be the main cause of misclassification.

Even though the hypothesis of non-differential misclassification behavior of investors cannot be discarded, estimates for the demand model are fundamentally in line with those of the previous analysis. Some important differences are, however, discernible. In particular, coefficient of second quartile of net wealth is not significant now and, in general, the effect of net wealth is strongly attenuated, thus further strengthening the hypothesis of the presence of minimum investment requirements and monetary transaction costs that limit participation in financial markets.

Also the effect of age is reduced, and family size is negatively correlated with the probability of holding risky assets: this result might be explained by a reduction of savings coming from the increase of number of dependent household members. Households with mortgage debt have a higher probability of holding risky assets, supporting the idea that households consider mortgage debt a rather safe investment and use risky assets to diversify their assets.

The parameter estimate of risk aversion confirms the results outlined in the main analysis: risk attitude represents the most important element for explaining differences in risky financial behavior, also when controlling for the differential misclassification behavior of investors.

5 Concluding remarks

This paper analyzes the role of risk aversion in an empirical choice model of risky assets. The risk aversion measure comes from a subjective question in a survey and it is supposed to be an imperfect information about the true risk attitude of investors. Under a minimal set of distributional assumptions, the choice model of risky assets with misclassified risk aversion is point identified and a Bayesian procedure for incomplete-data problems can be developed in order to get robust estimates. The model is applied to data from the 2006 Survey on Household Income and Wealth by the Bank of Italy.

The main results of the paper is that when the misclassification of investors is taken into account, the economic content of the model remains the same in terms of household characteristics’ effect, thus underlining the good performance of the subjective question when used as a control in portfolio choice models. When correcting for misclassification the risk attitude represents the most important factor for explaining differences in risky financial behavior. A clear misclassification pattern emerges from
the data: probability of misclassification decreases as latent risk aversion increases, that means that less risk averse investors tend to be classified erroneously more often than more risk averse investors. The over-reporting of risk aversion might be mainly due to two factors: on the one hand risk tolerant individuals may be led to say they are risk averse since they do not hold risky assets because of their low level of wealth, education and so on; on the other hand more risk tolerant investors may be affected by a kind of illusion of control that leads them to underestimate the risk taken and thus to declare themselves as more risk averse than they effectively are. The analysis shows that the latter hypothesis seems to prevail over the former.
Table 1: Households owning risky financial assets\(^{(+)}\) by observed risk aversion

<table>
<thead>
<tr>
<th>Risk aversion:</th>
<th>Percentage of households</th>
<th>Percentage owning risky financial assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. very low</td>
<td>0.90</td>
<td>30.55</td>
</tr>
<tr>
<td>2. low</td>
<td>14.89</td>
<td>21.89</td>
</tr>
<tr>
<td>3. high</td>
<td>35.26</td>
<td>22.89</td>
</tr>
<tr>
<td>4. very high</td>
<td>48.95</td>
<td>7.46</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
<td>15.26</td>
</tr>
</tbody>
</table>


\(^{(+)}\): include stocks, private bonds, mutual funds, and foreign equities.
Table 2: Bayesian estimates of cross-sectional model of household demand for risky assets\(^{(+)}\)

<table>
<thead>
<tr>
<th>Variable</th>
<th>[1]</th>
<th>[2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-4.136* (0.1308)</td>
<td>-4.141* (0.1249)</td>
</tr>
<tr>
<td>Age: between 25 and 45</td>
<td>0.592* (0.1026)</td>
<td>0.589* (0.1176)</td>
</tr>
<tr>
<td>between 45 and 65</td>
<td>0.763* (0.1026)</td>
<td>0.761* (0.1133)</td>
</tr>
<tr>
<td>more than 65</td>
<td>0.641* (0.1174)</td>
<td>0.644* (0.1207)</td>
</tr>
<tr>
<td>Labour Income (scaled by 10,000)</td>
<td>0.048* (0.0076)</td>
<td>0.047* (0.0077)</td>
</tr>
<tr>
<td>Net wealth: second quartile</td>
<td>0.525* (0.0498)</td>
<td>0.524* (0.0505)</td>
</tr>
<tr>
<td>third quartile</td>
<td>0.603* (0.0556)</td>
<td>0.611* (0.0495)</td>
</tr>
<tr>
<td>fourth quartile</td>
<td>1.289* (0.0505)</td>
<td>1.300* (0.0501)</td>
</tr>
<tr>
<td>Male</td>
<td>0.192* (0.0322)</td>
<td>0.194* (0.0362)</td>
</tr>
<tr>
<td>Family size</td>
<td>0.0064 (0.0162)</td>
<td>0.0054 (0.0157)</td>
</tr>
<tr>
<td>Two income earners or more</td>
<td>-0.076 (0.0435)</td>
<td>-0.073 (0.0425)</td>
</tr>
<tr>
<td>North</td>
<td>1.042* (0.0412)</td>
<td>1.045* (0.0379)</td>
</tr>
<tr>
<td>Centre</td>
<td>0.274* (0.0443)</td>
<td>0.268* (0.0496)</td>
</tr>
<tr>
<td>College education</td>
<td>0.617* (0.0613)</td>
<td>0.614* (0.0551)</td>
</tr>
<tr>
<td>High school diploma</td>
<td>0.523* (0.0396)</td>
<td>0.526* (0.0378)</td>
</tr>
<tr>
<td>Self-employed</td>
<td>-0.349* (0.0688)</td>
<td>-0.347* (0.0608)</td>
</tr>
<tr>
<td>Employee</td>
<td>-0.061 (0.0523)</td>
<td>-0.056 (0.0428)</td>
</tr>
<tr>
<td>Mortgage debt</td>
<td>-0.009 (0.0654)</td>
<td>-0.003 (0.0564)</td>
</tr>
</tbody>
</table>

**Risk Aversion:**

<table>
<thead>
<tr>
<th>Observed</th>
<th>Latent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. very low</td>
<td>0.905* (0.1453)</td>
</tr>
<tr>
<td>2. low</td>
<td>0.506* (0.0478)</td>
</tr>
<tr>
<td>3. high</td>
<td>0.579* (0.0327)</td>
</tr>
</tbody>
</table>

| 4. very high (baseline)    |

No. of Obs.: 7768

Source: SHIW, 2006. 20,000 iterations; estimates on the last 5,000 iterations.

\(^{(+)}\): include stocks, private bonds, mutual funds, and foreign equities.

1. The choice model of risky assets is estimated under the hypothesis that risk aversion is not affected by misclassification.

2. The choice model of risky assets is estimated under the hypothesis that risk aversion is affected by misclassification, and the misclassification is independent on other variables.

*: significantly different from zero at 5 per cent. Standard error are reported in parentheses.
Table 3: *Bayesian estimates of cross-sectional model of household demand for risky assets*(+)

*Marginal effects*

<table>
<thead>
<tr>
<th>Variable:</th>
<th>[1]</th>
<th>[2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age: between 25 and 45</td>
<td>0.0901*</td>
<td>0.0900*</td>
</tr>
<tr>
<td>between 45 and 65</td>
<td>0.1147*</td>
<td>0.1149*</td>
</tr>
<tr>
<td>more than 65</td>
<td>0.1001*</td>
<td>0.1007*</td>
</tr>
<tr>
<td>Labour Income (scaled by 10,000)</td>
<td>0.0064*</td>
<td>0.0063*</td>
</tr>
<tr>
<td>Net wealth: second quartile</td>
<td>0.0816*</td>
<td>0.0816*</td>
</tr>
<tr>
<td>third quartile</td>
<td>0.0911*</td>
<td>0.0926*</td>
</tr>
<tr>
<td>fourth quartile</td>
<td>0.2289*</td>
<td>0.2318*</td>
</tr>
<tr>
<td>Male</td>
<td>0.0257*</td>
<td>0.0261*</td>
</tr>
<tr>
<td>Family size</td>
<td>0.0009</td>
<td>0.0007</td>
</tr>
<tr>
<td>Two income earners or more</td>
<td>-0.0104</td>
<td>-0.0101</td>
</tr>
<tr>
<td>North</td>
<td>0.1421*</td>
<td>0.1433*</td>
</tr>
<tr>
<td>Centre</td>
<td>0.0406*</td>
<td>0.0398*</td>
</tr>
<tr>
<td>College education</td>
<td>0.1039*</td>
<td>0.1036*</td>
</tr>
<tr>
<td>High school diploma</td>
<td>0.0793*</td>
<td>0.0802*</td>
</tr>
<tr>
<td>Self-employed</td>
<td>-0.0427*</td>
<td>-0.0428*</td>
</tr>
<tr>
<td>Employee</td>
<td>-0.0084</td>
<td>-0.0078</td>
</tr>
<tr>
<td>Ownership of mortgage debt</td>
<td>-0.0012</td>
<td>-0.0004</td>
</tr>
</tbody>
</table>

*Risk Aversion:*  
Observed  
Latent  
1. very low | 0.1727* | 0.3136* |
2. low | 0.0802* | 0.1435* |
3. high | 0.0848* | 0.0781* |
4. very high *(baseline)*

No. of Obs.: 7768

Source: SHIW, 2006. 20,000 iterations; estimates on the last 5,000 iterations.

*(+): include stocks, private bonds, mutual funds, and foreign equities.*

[1] The choice model of risky assets is estimated under the hypothesis that risk aversion is not affected by misclassification.

[2] The choice model of risky assets is estimated under the hypothesis that risk aversion is affected by misclassification, and the misclassification is independent on other variables.

*: significantly different from zero at 5 per cent. Standard error are reported in parentheses.
### Table 4: Bayesian estimates of transition probabilities, $p_{rs} = \Pr(w_i = s|I_i = r)$

<table>
<thead>
<tr>
<th>Latent risk aversion</th>
<th>$w_i = 1$</th>
<th>$w_i = 2$</th>
<th>$w_i = 3$</th>
<th>$w_i = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_i = 1$</td>
<td><strong>0.6383</strong></td>
<td>0.1305</td>
<td>0.1431</td>
<td>0.0882</td>
</tr>
<tr>
<td>$I_i = 2$</td>
<td>0.1061</td>
<td><strong>0.6812</strong></td>
<td>0.1218</td>
<td>0.0909</td>
</tr>
<tr>
<td>$I_i = 3$</td>
<td>0.0070</td>
<td>0.2575</td>
<td><strong>0.7250</strong></td>
<td>0.0105</td>
</tr>
<tr>
<td>$I_i = 4$</td>
<td>0.0025</td>
<td>0.0289</td>
<td>0.0125</td>
<td><strong>0.9560</strong></td>
</tr>
</tbody>
</table>


### Table 5: Fraction of households owning risky financial assets$^{(+)}$ by latent risk aversion

<table>
<thead>
<tr>
<th>Latent risk aversion:</th>
<th>Percentage of households</th>
<th>Percentage owning risky financial assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. very low</td>
<td>0.59</td>
<td>52.55</td>
</tr>
<tr>
<td>2. low</td>
<td>2.01</td>
<td>25.51</td>
</tr>
<tr>
<td>3. high</td>
<td>47.08</td>
<td>22.76</td>
</tr>
<tr>
<td>4. very high</td>
<td>50.32</td>
<td>7.44</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
<td>15.26</td>
</tr>
</tbody>
</table>


$^{(+)}$: include stocks, private bonds, mutual funds, and foreign equities.

### Table 6: Contingency table of risk aversion surrogates (percentage values)

<table>
<thead>
<tr>
<th>Risk aversion 2006</th>
<th>1. very low</th>
<th>2. low</th>
<th>3. high</th>
<th>4. very high</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. very low</td>
<td>0.08</td>
<td>0.80</td>
<td>0.15</td>
<td>0.03</td>
<td>1.06</td>
</tr>
<tr>
<td>2. low</td>
<td>0.17</td>
<td>2.89</td>
<td>4.92</td>
<td>2.42</td>
<td>10.40</td>
</tr>
<tr>
<td>3. high</td>
<td>0.54</td>
<td>4.69</td>
<td>16.81</td>
<td>14.11</td>
<td>36.15</td>
</tr>
<tr>
<td>4. very high</td>
<td>0.23</td>
<td>7.12</td>
<td>14.39</td>
<td>30.66</td>
<td>52.39</td>
</tr>
<tr>
<td>Total</td>
<td>1.01</td>
<td>15.51</td>
<td>36.27</td>
<td>47.21</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Gamma association index: 0.3614

Source: SHIW, 2004 and 2006
**Table 7:** Bayesian estimates of cross-sectional model of household demand for risky assets\(^{(+)}\) with misclassification model

<table>
<thead>
<tr>
<th>Variable:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Age: between 45 and 65</td>
</tr>
<tr>
<td>more than 65</td>
</tr>
<tr>
<td>Labour Income (scaled by 10,000)</td>
</tr>
<tr>
<td>Net wealth: second quartile</td>
</tr>
<tr>
<td>third quartile</td>
</tr>
<tr>
<td>fourth quartile</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Family size</td>
</tr>
<tr>
<td>Two income earners or more</td>
</tr>
<tr>
<td>North</td>
</tr>
<tr>
<td>Centre</td>
</tr>
<tr>
<td>College education</td>
</tr>
<tr>
<td>High school diploma</td>
</tr>
<tr>
<td>Self-employed</td>
</tr>
<tr>
<td>Employee</td>
</tr>
<tr>
<td>Ownership of mortgage debt</td>
</tr>
</tbody>
</table>

**Latent Risk Aversion:**

|                       |                       |                       |
| low                   | 3.765*                | 2.287*                | 0.7370*                       |
| high (baseline)       |                       |                       |

No. of Obs.: 1485

Source: SHIW, 2004 and 2006. 20,000 iterations; estimates on the last 5,000 iterations.

\(^{(+)}\): include stocks, private bonds, mutual funds, and foreign equities.

*: significantly different from zero at 5 per cent.
Table 8: *Probability of misclassification by household characteristic (percentage values)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Risk aversion</th>
<th>Risk aversion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>over-reporting</td>
<td>under-reporting</td>
</tr>
<tr>
<td>Ownership of risky assets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>9.39</td>
<td>14.71</td>
</tr>
<tr>
<td>Yes</td>
<td>37.44</td>
<td>16.10</td>
</tr>
<tr>
<td>Age:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>less than 45</td>
<td>1.43</td>
<td>22.44</td>
</tr>
<tr>
<td>between 45 and 65</td>
<td>29.65</td>
<td>12.28</td>
</tr>
<tr>
<td>more than 65</td>
<td>53.79</td>
<td>7.74</td>
</tr>
<tr>
<td>Net wealth:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>first quartile</td>
<td>18.06</td>
<td>14.86</td>
</tr>
<tr>
<td>second quartile</td>
<td>50.34</td>
<td>16.20</td>
</tr>
<tr>
<td>third quartile</td>
<td>40.52</td>
<td>13.22</td>
</tr>
<tr>
<td>fourth quartile</td>
<td>18.23</td>
<td>15.32</td>
</tr>
<tr>
<td>Gender:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>28.07</td>
<td>9.67</td>
</tr>
<tr>
<td>male</td>
<td>32.73</td>
<td>17.17</td>
</tr>
<tr>
<td>Area of residence:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>north</td>
<td>17.80</td>
<td>13.55</td>
</tr>
<tr>
<td>centre</td>
<td>51.76</td>
<td>18.79</td>
</tr>
<tr>
<td>south or islands</td>
<td>56.44</td>
<td>14.41</td>
</tr>
<tr>
<td>Education:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>college education</td>
<td>3.48</td>
<td>20.71</td>
</tr>
<tr>
<td>high school diploma</td>
<td>14.69</td>
<td>22.89</td>
</tr>
<tr>
<td>middle school or less</td>
<td>42.46</td>
<td>8.97</td>
</tr>
<tr>
<td>Occupation:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>self-employed</td>
<td>31.35</td>
<td>17.29</td>
</tr>
<tr>
<td>employee</td>
<td>20.53</td>
<td>16.42</td>
</tr>
<tr>
<td>not employed</td>
<td>44.08</td>
<td>12.53</td>
</tr>
<tr>
<td>Total</td>
<td>30.61</td>
<td>14.89</td>
</tr>
</tbody>
</table>

References


RECENTLY PUBLISHED “TEMI” (*)

N. 664 – Portfolio selection with monotonotone mean-variance preferences, by Fabio Maccheroni, Massimo Marinacci, Aldo Rustichini and Marco Taboga (April 2008).
N. 670 – Credit risk and business cycle over different regimes, by Juri Marcucci and Mario Quagliariello (June 2008).
N. 671 – Cyclical asymmetry in fiscal variables, by Fabrizio Balassone, Maura Francese and Stefania Zotteri (June 2008).
N. 672 – Labour market for teachers: Demographic characteristics and allocative mechanisms, by Gianna Barbieri, Piero Cipollone and Paolo Sestito (June 2008).
N. 673 – Output growth volatility and remittances, by Matteo Bugamelli and Francesco Paternò (June 2008).
N. 674 – Agglomeration within and between regions: Two econometric based indicators, by Valter Di Giacinto and Marcello Pagnini (June 2008).
N. 677 – Forecasting inflation and tracking monetary policy in the euro area: Does national information help?, by Riccardo Cristadoro, Fabrizio Venditti and Giuseppe Saporito (June 2008).
N. 678 – Monetary policy effects: New evidence from the Italian flow of funds, by Riccardo Bonci and Francesco Columba (June 2008).
N. 679 – Does the expansion of higher education increase the equality of educational opportunities? Evidence from Italy, by Massimiliano Bratti, Daniele Checchi and Guido de Blasio (June 2008).
N. 680 – Family succession and firm performance: Evidence from Italian family firms, by Marco Cuculelli and Giacinto Micucci (June 2008).
N. 681 – Short-term interest rate futures as monetary policy forecasts, by Giuseppe Ferrero and Andrea Nobili (June 2008).
N. 682 – Vertical specialisation in Europe: Evidence from the import content of exports, by Emanuele Breda, Rita Capparelli and Roberta Zizza (August 2008).
N. 683 – A likelihood-based analysis for relaxing the exclusion restriction in randomized experiments with imperfect compliance, by Andrea Mercatanti (August 2008).
N. 684 – Balancing work and family in Italy: New mothers’ employment decisions after childbirth, by Piero Casadio, Martina Lo Conte and Andrea Neri (August 2008).

(*) Requests for copies should be sent to: Banca d’Italia – Servizio Studi di struttura economica e finanziaria – Divisone Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.


P. Del Giovane and R. Sabbatini, Perceived and measured inflation after the launch of the Euro: Explaining the gap in Italy, Giornale degli economisti e annali di economia, Vol. 65, 2 , pp. 155-192, TD No. 532 (December 2004).

M. Caruso, Monetary policy impulses, local output and the transmission mechanism, Giornale degli economisti e annali di economia, Vol. 65, 1, pp. 1-30, TD No. 537 (December 2004).


L. MONTEFORTE, *Aggregation bias in macro models: Does it matter for the euro area?*, Economic Modelling, 24, pp. 236-261, TD No. 534 (December 2004).


**FORTHCOMING**


R. Felici and M. Pagnini, *Distance, bank heterogeneity and entry in local banking markets*, The Journal of Industrial Economics, TD No. 557 (June 2005).

M. Bugamelli and A. Rosolia, *Produttività e concorrenza estera*, Rivista di politica economica, TD No. 578 (February 2006).


S. FEDERICO and G. A. MINERVA, *Outward FDI and local employment growth in Italy*, Review of World Economics, Journal of Money, Credit and Banking, **TD No. 613** (February 2007).

M. BUGAMELLI, *Prezzi delle esportazioni, qualità dei prodotti e caratteristiche di impresa: analisi su un campione di imprese italiane*, Economia e Politica Industriale, **TD No. 634** (June 2007).


S. MAGRI, *The financing of small innovative firms: The Italian case*, Economics of Innovation and New Technology, **TD No. 640** (September 2007).
