

Subjective Beliefs and Schooling Decisions*

Christian Belzil
Centre de National de Recherche Scientifique,
Institute for the Study of Labor (IZA)
and CIRANO

May 25, 2007

*I thank Christian Gourieroux, Robert Sauer, Robin Mason, Jean Marc Robin, David Huffman, David Margolies, Marco Leonardi and seminar participants at Southampton, Milano, GATE (Lyon) and Paris School of Economics (Pantheon-Sorbonne) for helpful comments and discussions.

Abstract

This paper considers the estimation of sequential schooling decisions made by agents who are endowed with subjective beliefs about their own ability. I use unique Italian panel data which provide information on i) the curvature of the per-period utility function, ii) schooling decisions, iii) post-schooling earnings, in order to estimate the future component of the differences in intertemporal utilities of school and work independently from the present component, (as in Geweke and Keane, 1995, 2001), and evaluate the importance of “present bias”. Under certain conditions, which include imposing equality between the modal belief and true ability, I recover individual specific subjective probability distributions. I estimate both the degree of confidence (a measure of spread) and the incidence of over (and under) estimation. I find that the future component of intertemporal utilities dominates schooling decisions. I find a strong incidence of under-estimation among the more able and a much smaller incidence of over-estimation among the low ability group. At the medium ability spectrum, there is evidence of some over-estimation. The degree of confidence is high and imply that agents have a substantial amount of inside information (36% of the population act on a degenerate subjective distribution). Overall, the variance of the objective ability heterogeneity distribution is 4 times as large the variance of the distribution characterizing subjective beliefs.

JEL Classification: J24.

Key Words: Subjective distributions, Expectation Parameterization, Rational Expectation, Schooling, Dynamic programming, Present bias, Over-Confidence.

1 Introduction

This paper considers the estimation of optimal (sequential) schooling decisions made within an imperfect information framework. Imperfections arise for two main reasons. First, agents have imperfect information about their abilities in the labor market and are endowed with a subjective probability distribution that characterizes their beliefs. This distribution is taken as exogenous. A second form of imperfection arises because agents may have subjective beliefs about future earnings random shocks that diverge from the objective distribution of post-schooling earnings¹

The approach to estimation proposed herein is in the spirit of Geweke and Keane (1995, 2001). I use unique Italian panel data which provide information on i) the curvature of the per-period utility function, ii) schooling decisions, iii) post-schooling earnings, in order to estimate the future component of the differences in intertemporal utilities of school and work independently from the present component. Because the form of the per-period utility is known (the risk aversion parameter), the information about the curvature of the per-period utility function enters both the current utility and the future utility components separately. The separation between present and future components implies that I can estimate the degree of “present bias” in schooling decisions.²

Under certain conditions, which include imposing equality between the modal belief and true ability, I can recover individual specific subjective ability distributions. This has two immediate implications. First, I can estimate the location parameter of the subjective distribution and investigate the accuracy of subjective beliefs. This leads to the measurement of the incidence of over (or under) estimation. Second, the dispersion of the subjective ability distribution allows me to evaluate the degree of confidence of the agents (as measured by the variance of the subjective beliefs). In turn, the ratio of the variance of the objective heterogeneity distribution (from the perspective of the econometrician) over the variance of subjective beliefs may be informative about the degree of inside information held by agents (over the

¹Because I do not model how subjective distributions are inferred, the model is not necessarily inconsistent with a standard rational expectation framework.

²This also implies that I can determine if young individuals perceive schooling as a form of insurance or as a risky asset. However, this issue (and other surrounding questions) are investigated in a companion paper (Belzil and Leonardi, 2007). For this reason, I ignore the link between risk and education in this paper.

econometrician). The analysis proposed herein may therefore be viewed as a generalization of the notion of “Overconfidence” that has recently gained in popularity. Finally, note that the degree of confidence is also a measure of ex-ante risk caused by the ignorance that agents may have about their own abilities.³

The empirical results display 6 major findings.

1. At all grade levels, both the average values and the degree of dispersion of the future component of the intertemporal utilities exceed the equivalent measure for the present component. The predominance of future components in schooling decisions, seem to be supportive of the standard time consistent model. However, at the same time, idiosyncratic utility shocks appear to be an important determinant of schooling decisions.
2. There is a high degree of heterogeneity in subjective beliefs, but it is explained mostly by the importance of heterogeneity in the objective ability distribution (identified from panel data). Subjective beliefs are found to be quite accurate on average (within-type dispersion is small).
3. Precisely, 36% of the population is found to take decisions based on a degenerate ability distribution.
4. I find a strong incidence of under-estimation among the more able but a small incidence of over-estimation among the low ability group. At the medium ability spectrum, there is evidence of some over-estimation.
5. Defining the degree of confidence as the ratio of the variance of the objective distribution (as perceived by the econometrician) to the variance of the subjective distribution, would lead to a degree approximately equal to 4. In other words, only 25% of unobserved ability heterogeneity is actually perceived as ex-ante risk. The degree of confidence is high and imply that agents have a substantial amount of inside information.
6. Both the level of ex-ante risk (the degree of non-confidence) and the degree of inaccuracy (the incidence of either under-estimation or over-estimation) are increasing with market ability.

³The importance of distinguishing between ex-ante risk and heterogeneity (or inequality) is made forcefully in Cunha, Heckman and Navarro (2005).

The remaining sections of the paper are structured according to the following format. In Section 2, I discuss the motivation and some background literature. Section 3 describes the model. The Bank of Italy Survey of Income and Wealth (SHIW) is presented in Section 4. All assumptions required to implement the model are laid out in Section 5. In Section 6, I discuss how risk aversion is measured. Section 7 is devoted to issues surrounding identification and estimation. In Section 8, I briefly outline the presentation of the structural parameters. In Section 9, I discuss the relative importance of present and future components. In Section 10, I present the results characterizing the objective ability distribution. In Section 11, I discuss a set of issues (over and under estimation, confidence) related to subjective beliefs. A brief summary of results is reported in Section 12. The Conclusion is in Section 13.

2 Motivation and Background literature

The structural literature concerned with dynamic models of human capital accumulation has expanded rapidly in the past 10 years. There are good reasons for that. Structural models provide a transparent illustration of several dynamic self-selection issues which are central to the analysis of education and skill formation policies. Perhaps more importantly, the structural approach provides the only econometric framework which enables the researcher to distinguish the determinants of ex ante decisions from ex-post outcomes. To achieve this, those who have estimated dynamic models have relied invariably on various forms of the “Rational Expectation” (RE) hypothesis.⁴

Dynamic rational expectation models are based on the assumption that agents use available information efficiently. At a practical level, the RE hypothesis implies a coincidence between subjective probabilities used by the agent and the objective distribution that generates observable outcomes such as endogenous state variables (the law of motion) or random shocks. The RE hypothesis is therefore a powerful identifying restriction.⁵ Because

⁴In the structural literature on schooling decisions, Keane and Wolpin, 1997, Eckstein and Wolpin, 1999, Belzil and Hansen, 2002, and Heckman and Navarro (2006) are examples of dynamic rational expectations models. The literature is surveyed in Belzil (2007).

⁵There is indeed a relatively wide literature on identification of dynamic discrete choices. See Rust (1994), Magnac and Thesmar (2002), and more recently, Heckman and Navarro (2006).

the agent knows the objective distribution of interest, the econometrician may infer the distribution of stochastic shocks (and the law of motion) from data, and use it in order to solve the dynamic program faced by the agent.⁶

In many applications, the RE hypothesis involves more than the distribution of idiosyncratic random shocks. If there are time invariant state variables unobserved to the econometrician (what is usually referred to as unobserved heterogeneity), it is customary to maintain the assumption that individuals know their values from the beginning of the horizon, even though some of these variables may turn out to be realized at a later stage in the life cycle. In the schooling literature, this implies that labor market ability is known before entering the labor market.⁷

While the RE hypothesis is convenient, its empirical justification is rarely discussed. In his insightful criticism of the rational expectation hypothesis in micro-econometric models, Manski (2004) argues against the capacity of agents to form subjective beliefs that are consistent with actual outcomes based on the lack of agreement among micro-econometricians on a proper inference method.⁸ Manski's criticisms may appear particularly relevant in the context where risk averse agents make optimal schooling decisions. Measuring the relationship between attitudes toward risk and educational choices is a long standing problem in labor economics. This is partly due to the fact that quantifying the marginal risk which characterizes the transition from one level of schooling to the next is much harder than evaluating the relative riskiness of various financial assets.

Not surprisingly, the controversial nature of the RE hypothesis has pushed many econometricians to suggest obtaining direct measures of subjective probability distributions, from various forms of elicitation. This literature is examined in Manski (2004).⁹ While using actual data on beliefs would always be a dominant strategy, such data are rarely available. For this rea-

⁶In more extreme cases, such as in Rust's 1987 seminal piece on engine replacement, the random shocks is assumed to follow an extreme value distribution (with no free parameter) and the econometrician does not even need to use data on outcomes.

⁷In a certain sense, perfect knowledge of one's abilities in the labor market requires a form of perfect foresight.

⁸Obviously, the RE hypothesis is also difficult to maintain outside equilibrium, or in the presence of potential technological changes. In the presence of search friction, persistent individual effects identified in earnings panel data would also be affected by firm heterogeneity. By construction, it would be impossible for any agent to know these factors ex-ante.

⁹See Bellemare, Kröger and van Soest (2005) for a recent example.

son, it is particularly important to design an estimation method that may work in a context where agents have imperfect knowledge about their abilities and where econometricians do not have access to data on beliefs and expectations¹⁰. This is indeed the main objective of this paper.

However, the contributions are not solely located to the econometric literature on the estimation of dynamic discrete choices. The method that I propose in this paper is quite different from the usual structural strategy. It builds on a method proposed by Geweke and Keane (1995, 2001).¹¹ Because the estimation method decomposes sequential choices into three separate components (present utility, future utility, and utility shocks) which are freely estimated, I can evaluate the relative importance of the future in schooling decisions. This is a second objective, and it is an important issue. Standard structural (dynamic) estimation usually requires to posit a parametric family for the utility function (or other primitive objects) as well as all the probability distributions that characterize the subjective beliefs of the agents. The future component of the intertemporal utility of a specific choice is a by-product of the model structure. However, in recent years, the validity of intertemporal expected utility models has been the object of debates between Behavioral economists and proponents of the expected utility theory. More precisely, the existence of “present biased” behavior has been put forward as a challenge to conventional time consistent intertemporal models.¹² A second objective is to contribute to the debate between behavioral economists and proponents of the classical intertemporal models. I can do so by evaluating the relative importance of the future in schooling decisions.

A third objective is to evaluate both the degree of accuracy and the degree of confidence (dispersion) that characterize subjective ability evaluations. This allows me to link another segment of the behavioral literature with the literature on dynamic discrete choices. In recent years, the notion of “over-confidence” has attracted much attention. It has been used in order

¹⁰In the behavioral economics literature, it is common to elicit beliefs within laboratory experiments. However, the extrapolation of laboratory outcomes to actual lifecycle decisions may sometimes be viewed as controversial.

¹¹As far as I know, the Geweke-Keane approach (a form of expectation parameterization) has not been used widely by microeconometricians. However, it has been applied in an experimental framework by Houser, Keane and McCabe (2004).

¹²The literature on Hyperbolic discounting, which has emerged in the past 10 years, is perhaps the best example. Rubinstein (2005) present a good exposition of the major theoretical issues that are at stake in the current debate between behavioral economists and proponents of the more classical approach.

to explain financial market behavior, and has also been analyzed in various laboratory experiments. As of now, over-confidence has been used mostly as a relative notion (using individual's own subjective ranking within a benchmark distribution in comparison with the individuals' actual ranking), but it has not been analyzed in a formal structural context.¹³ Interestingly, most applications seem to consider over-confidence in a degenerate context, in which ex-ante risk (about ability) plays no role.¹⁴ As the approach suggested herein relies on estimation a full distribution of beliefs, I can distinguish between the degree of accuracy (under-estimation and over-estimation) and the degree of confidence (the level of ex-ante dispersion in the individual specific parameter). Ultimately, the approach I suggest generalizes the popular notion of over-confidence.

Finally, because the model distinguishes ex-ante beliefs from ex-post distributions (identified from panel data), a fourth objective is to contribute to the recent literature on the measurement of the degree of ex-ante risk and its comparison with cross-sectional heterogeneity which appears to be the central empirical measure motivating the literature on income or wage inequality.¹⁵

3 The Model

In this section, I present a simple theoretical model of schooling choices. The control variable, d_t , is equal to S when the individual chooses to continue in school, and to W when entering the labor market is chosen. The level of schooling completed by at the beginning of each period, denoted j , may take $j = 1, 2, \dots, J$ values. While it would be possible to equate time (t) with grade level (j), I keep a separate time index. Finally, it is important to note that in the presentation we distinguish between subjective probability distributions, denoted by $\mu(\cdot)$, and objective distributions, denoted by $F(\cdot)$.

¹³Camerer and Lovello (1999) is a classical reference.

¹⁴Obviously, this does not mean that other aspects of the experiments do not involve risk.

¹⁵See Cunha, Heckman and Navarro (2005).

3.1 The Utility of Attending School

The per-period utility of attending school for an individual i , at time t , and who has already completed grade level j , $U_{it}^S(j)$, is

$$U_{it}^S(j) = \phi\{I_i^S(j); \alpha_i\} + \varepsilon_{it}^S(j) \quad (1)$$

where

$$I_i^S(j) = \theta_i^S + \bar{I}_i^S(j) \quad (2)$$

and where

- $\phi(\cdot; \alpha_i)$ is the deterministic part of the utility function, α_i is an individual specific parameter that indicates the degree of curvature.
- $I_i(j)$ is the net individual income while school (in grade j). $\bar{I}(j)$ may capture the effect of institutional features on the cost (or utility) of being in grade j .
- θ_i^S is individual specific taste/ability affecting net income¹⁶ It is known by the agent.
- $\varepsilon_{it}^S(j)$ is a random utility shock affecting the utility of attending grade level j .¹⁷

3.2 Labor market income

Labor market income is stochastic. It is expressed as:

$$I_{it}^W(j) = \bar{I}^W(j) + \theta_i^W + \varepsilon_{it}^W(j) \quad (3)$$

where $\varepsilon_{it}^W(j)$ is a random shock, which distribution may depend on grade level. It is discussed in details below. The term θ_i^W represents individual specific market ability (it is unobserved to the econometrician). The per-period utility of work will be introduced below, after having defined clearly the information set of the agent.

¹⁶It is common to refer to θ_i as the true psychic costs of attending school.

¹⁷Another way to proceed would be to assume income while income in school as stochastic.

3.3 Individual Persistent Abilities

In order to solve the problem, young individuals must form a subjective probability distribution of θ_i^W . I am agnostic about its determination. It could be understood as a belief conditional on θ_i^S or/and on an individual specific vector of family background variables (X_i). It is denoted $\mu_{\theta^w}(\cdot)$.¹⁸ It may correspond to a formal inference process that use wages and abilities (school and labor market) of individuals who belong to a different (older) cohort.

3.4 Present and Future Earnings Random shocks

I assume young individuals do not know the stochastic shock affecting labor market income when the decision between school and work is exercised. It may be justified if, for instance, new entrants have to invest in search activities. Because $\varepsilon_{it}^W(j)$, reflects the status of the labor market at time t (for an individual with grade level j), it is reasonable to assume that agents know the actual (objective) distribution $F(\varepsilon_{it}^w)$. The subjective probability distributions over future outcomes is denoted $\mu_t^F(j)$. It represents the subjective beliefs about all (relevant) future random shocks affecting earnings, but also random components affecting job search, career changes, training (or re-training) opportunities or the incidence of technological change, which are formed at time t . It is not necessarily equal to its corresponding objective distribution. Obviously, it may (and should) depend on grade level (j) since agents may believe that the higher moments of the random shocks are affected by accumulated schooling. At this stage, $\mu_t^F(\cdot)$ is left unspecified.

3.5 Information Set and Timing of Decisions

I now define the information set of individual i , at time t , which is denoted Ω_{it} . Because I focus on schooling decisions, I consider periods that precede labor market entrance. The information set, Ω_{it} , contains the following elements:

$$\Omega_{it} = \{j, \varepsilon_{it}^S(j), \bar{I}_i^S(j), \theta_i^S, X_i\} \quad (4)$$

where X_i is an individual specific vector of family attributes. The timing of the decision is therefore the following. Upon entering period t with grade

¹⁸Formally, it should be denoted $\mu_{\theta}(\theta_i^W | \theta_i^S, X_i)$ but I use $\mu_{\theta^w}(\cdot)$ to avoid cumbersome notation.

level j already completed, individuals observe the utility shock (ε_{it}^S). Then, they decide on entering the labor market or completing an additional grade level, based on θ_i^S , $F(\cdot)$, $\mu_t^F(\cdot)$ and $\mu_{\theta^w}(\cdot)$.

3.6 The Utility of work

From the expected utility paradigm, the expected utility of entering the labor market (with completed grade level j) is:

$$U_{it}^W(j) = E\phi[I_{i,t}^W(j); F(\varepsilon_{it}^w), \mu_{\theta^w}(\cdot)] = \int \int \phi[I_{i,t}^W(j)] dF(\varepsilon_{it}^w) d\mu_{\theta^w}(\cdot) \quad (5)$$

where the double integral reflects uncertainty about both the present earnings shock and labor market ability.

3.7 Value Functions

Using grade level (j) as the only conditioning state variable, the intertemporal utility of entering grade $j + 1$ (with grade j completed), $V_t^S(j)$, is

$$V_t^S(j) = \phi\{I_i^S(j)\} + \varepsilon_{it}^S(j) + \beta E(\text{Max} V_{t+1}^s(j+1), V_{t+1}^w(j+1); \mu_t^F(\cdot), \mu_{\theta^w}(\cdot)) \quad (6)$$

which entails taking expectation over future random shocks ($\mu_t^F(\cdot)$) and over subjective ability beliefs ($\mu_{\theta^w}(\cdot)$). To reduce notational burden, I re-express $V_t^S(j)$ as

$$V_t^S(j) = U_{it}^S(j) + \beta EV_{t+1}(j+1) \quad (7)$$

Assuming that work is an absorbing state, the value function of work, $V_t^W(j)$, is

$$V_t^W(j) = E\phi[I_{it}^W(j); F(\varepsilon_{it}^w(j)), \mu_{\theta^w}(\cdot)] + \beta EV_{t+1}^W(j; \mu_t^F(\cdot), \mu_{\theta^w}(\cdot)) \quad (8)$$

which I also re-express as

$$V_t^W(j) = U_{it}^W(j) + \beta EV_{t+1}^W(j) \quad (9)$$

The future component of the expected utility of work may include the option of taking a wide variety of actions (job search, career changes, training (or re-training) opportunities), which are not modeled explicitly.

4 The Bank of Italy Survey of Income and Wealth (SHIW)

The data come from the 1995 wave of the Bank of Italy Survey of Income and Wealth (SHIW). The survey collects information on consumption, income and wealth in addition to several household characteristics for a representative sample of 8,135 Italian households. More importantly, the 1995 survey contains a question on household willingness to pay for a lottery which can be used to build a measure of individual risk attitudes.¹⁹

Apart from the lottery question, I use information on the level of education attained by the head of household, as well as variables such as age, gender, region of birth, parental education and parental occupation. This set of variables is comparable to those which are used in US studies based on the National Longitudinal Survey (NLS). I select the sample of all heads with a valid answer to the lottery question (3,458) and eliminate those who report a missing value in any of the following variables: education, age, gender, region of birth, education and occupation of the head's father and mother. This selection process leaves us with a final sample of 3,288 heads of household.

The schooling variable takes values for 1 to 6 corresponding to no education, elementary school (typically attained at 11 years of age), junior high school (attained at 14), high school (attained at 18), university degree (attained at 23-24) and post-university degree. The data include schooling attainment (no qualification, elementary school, lower high school, upper high school and higher education) and panel data on post schooling labor market earnings.

The vector of family background variables (X_i) contains a set of education binary indicators equal to 1 the father and the mother have reached higher education ($FED = 1$, $MED = 1$) and a set of occupation indicators equal to 1 when they are classified as white collar worker ($FWC = 1$, $MWC = 1$). The data also provide information on gender and geographic region of birth (North, Center and South). These are summarized in a vector denoted Z_i . As will be clear later, In order to identify subjective beliefs, post-schooling

¹⁹Guiso and Paiella (2005), use the question on risk aversion to analyze occupation choice, portfolio selection, insurance demand, investment in education (in the linear OLS case) and migration decisions. They find substantial effects of this measure of risk aversion in ways that are consistent with the theory i.e. that more risk averse individuals choose lower returns in exchange for lower risk.

panel data on earnings are crucial. In the SHIW, I observe a maximum of 7 post-schooling earnings (1989, 1992, 1995, 1997, 1999, 2000, 2002). Earnings are measured on a yearly basis, and are set in euros. Sample statistics are found in Appendix.

5 Empirical Implementation

In this section, I document all explicit assumptions that are needed in order to estimate the model. There are four main four components; unobserved heterogeneity, the distribution of the random shocks, the specification of the future component, and the subjective and objective probability distributions.

5.1 Preferences

I assume that

$$\phi(I_i(j); \alpha_i) = -\exp(-\alpha_i \cdot I_i(j)) \quad (10)$$

where α_i is an individual specific parameter that measures the degree of absolute risk aversion. Risk aversion (loving) is obtained when α_i exceeds (is below) 0.

5.2 The Earnings Equation

There are 5 levels of education attainable by the individual. These levels are no qualification (level 1), elementary school (level 2), junior high school (level 3), Senior high school (level 4) and higher education (level 5).²⁰ I allow for non-linearities in the returns to schooling by using level specific dummy variables. I assume that

$$I_{i,j}^W = \theta_i^W + \zeta_j^W \cdot Level(j) + \zeta_6 \cdot \exp + \zeta_7 \cdot \exp^2 + \varepsilon_{it}^W(j) \text{ for } j = 1, 2..5 \quad (11)$$

and

$$I_{i,j}^S = \theta_i^S + \zeta_j^S \cdot Level(j) \text{ for } j = 1, 2..5 \quad (12)$$

²⁰I group those who report having completed graduate studies with those who have a university degree because their number is too small to form a distinct class.

where the ζ_j^W 's and the ζ_j^S 's are parameters to be estimated. The binary variables, $Level(j)$, are equal to 1 when individual i has reached level j and 0 if not. More details about θ_i^W and θ_i^S are found below.

5.3 The Distribution of Random Shocks

In order to minimize the impact of distributional assumptions, I approximate the density of the earnings shocks with a mixture of M unrestricted normal densities and the distribution the utility shocks with a mixture of M normal distribution functions. Precisely,

$$F(\varepsilon_{it}^S) = \sum_{m=1}^M P_m^S \cdot \Phi(\mu_m^S, \sigma_m^S). \quad (13)$$

where P_m^S is the mixing probability and $\Phi(\mu_m, \sigma_m)$ denotes the normal cumulative distribution function, and

$$F(\varepsilon_{it}^W) = \sum_{m=1}^M P_m^W \cdot \phi(\mu_m^W, \sigma_m^W). \quad (14)$$

where P_m^W is the mixing probability and $\phi(\mu_m, \sigma_m)$ denotes the normal density. More details concerning are provided in Appendix 4.

5.4 Ability Heterogeneity, Parents Background and Separability

Both θ_i^W and θ_i^S are decomposed into a separable regression component that depends on parents background variables and an orthogonal component. That is

$$\theta_i^S = \theta_1^S \cdot FED + \theta_2^S \cdot MED + \theta_3^S \cdot FWC + \theta_4^S \cdot MWC + \theta_Z^S \cdot Z_i + \tilde{\theta}_i^S \quad (15)$$

$$\theta_i^W = \theta_1^W \cdot FED + \theta_2^W \cdot MED + \theta_3^W \cdot FWC + \theta_4^W \cdot MWC + \theta_Z^W \cdot Z_i + \tilde{\theta}_i^W \quad (16)$$

where θ_Z^S and θ_Z^W are parameters associated to the socio economic variables.

I assume that agents are capable to infer the effects of parents' background on θ_i^W . In order to solve the dynamic programming problem, they must solely generate a subjective belief about $\tilde{\theta}_i^W$. This separability assumption will play a key role in the specification of the subjective beliefs.

5.5 The Objective Probability Distribution of Abilities

The population is composed of 6 types of individuals. Each type is endowed with a vector of relevant parameters $(\tilde{\theta}_i^W, \tilde{\theta}_i^S)$. The market ability term takes three values; $\tilde{\theta}_H^W$ (the highest value), $\tilde{\theta}_M^W$ (the medium value) and $\tilde{\theta}_L^W$ (the lowest value). The school ability term takes 2 values: $\tilde{\theta}_H^S$ and $\tilde{\theta}_L^S$. The six types are generated by all combinations possible of $\tilde{\theta}_i^W$ and $\tilde{\theta}_i^S$. The typology of the objective ability distribution is as follows:

Type	Market ($\tilde{\theta}_i^W$)	School ($\tilde{\theta}_i^S$)
$n = 1$	High	High
$n = 2$	High	low
$n = 3$	Medium	High
$n = 4$	Medium	Low
$n = 5$	Low	high
$n = 6$	low	Low

The (objective) type probabilities, $\Pr(\text{type } n)$, are estimated using logistic transform. That is

$$\Pr(\text{type } n) = \frac{\exp(\omega^n)}{1 + \exp(\omega^1) + \dots \exp(\omega^5)} \quad (17)$$

where the ω^n s are parameters to be estimated.

5.6 The Subjective Beliefs about Market Ability

In order to estimate the model, I must first specify the subjective probabilities of the agent. Each type (objective) of individual is endowed with a different set of subjective type probabilities. By this, I mean probabilities of being of type H , type M and type L . This heterogeneity may arise for several reasons. Individuals may use their inside information about θ_i^S to predict θ_i^W . An alternative hypothesis is that individuals with different abilities may use available information differently, and therefore be endowed with different subjective probabilities..

For an individual of type n , the subjective beliefs (or the set of subjective probabilities) are

$$\mu_{\theta^w}(\text{type } n) =$$

$$\begin{aligned} \{\text{Pr}(\tilde{\theta}_i^W &= \tilde{\theta}_H^W \mid \text{type } n), \text{Pr}(\tilde{\theta}_i^W = \tilde{\theta}_M^W \mid \text{type } n), \text{Pr}(\tilde{\theta}_i^W = \tilde{\theta}_L^W \mid \text{type } n)\} \\ &= \{p_{Hn}, p_{Mn}, p_{Ln}\} \end{aligned} \quad (18)$$

where each probability is specified as a logistic distribution;

$$p_{ln} = \frac{\exp(\mu_l^n)}{1 + \exp(\mu_l^n)} \text{ for } l = H, M, L \text{ and } j = 1, 2, \dots, 6 \quad (19)$$

where μ_l^n is a parameter to be estimated. Identification is addressed below.

5.7 The Present and Future Components of the Intertemporal Utility

5.7.1 The Present Component

The type-specific per-period utility (expected) of entering the labor market with schooling level j , $E\phi[I_{it}^W(j); F(\varepsilon_{it}^w), \mu_{\theta^w}(\cdot) \mid \text{type } n]$ is:

$$\sum_{l=H,M,L} p_{ln} \cdot \left\{ - \sum_{m=1}^M p_m^* \cdot \exp(-\alpha_i \cdot (\delta_j + \tilde{\theta}_l^W) + \frac{\alpha_i^2 \cdot \sigma_m^2}{2}) \right\} = h^n(j) \quad (20)$$

and the type specific difference in present utilities (the present component of the intertemporal utility), $P^n(j)$, is

$$P^n(j) = h^n(j) - \phi(I(j)) \quad (21)$$

5.7.2 Future component

To estimate the model, I must posit the form of the future component of the difference between the utility of work and the utility of attending school. The future component is to be understood as a weighted average of conditional future component (for each possible pair $(\tilde{\theta}_i^W, \tilde{\theta}_i^S)$). It represents the future component of the intertemporal utility of an individual with a given school

ability type and a given (perhaps counterfactual) market ability type. However, it is defined under the maintained assumption that all individual (all types) share a common belief about post-schooling random shocks $\mu_t^F(\cdot)$. I denote the conditional future component by $F_{kl}(j)$, where the index k identifies the school ability type (H or L), the index l identifies labor market ability type (H, M, L), and where j is again grade level. It is given by the following expression

$$F_{kl}(j) = F_{0klj} + F_{1klj} \cdot \alpha_i + F_{2klj} \cdot FED + F_{3klj} \cdot MED + F_{4klj} \cdot FWC + F_{5klj} \cdot MWC + F_{6klj} \cdot Z_i \quad (22)$$

where all the F'_{klj} s are parameters to be estimated. The parameter F_{0klj} plays the role of a heterogeneity term for each school and market ability type and grade level combinations.²¹ In words, $F_{kl}(j)$ is the difference in future components of value functions for a given level of school ability and a potential level of market ability. As opposed to the conditional future component, the type specific future component depends on the actual type simply because each type is endowed with its own subjective beliefs.

The type specific future component, $\bar{F}_{nj}(\cdot)$, is therefore equal to

$$\bar{F}_n(j) = p_{Hn} \cdot F_{kH}(j) + p_{Mn} \cdot F_{kM}(j) + p_{Ln} \cdot F_{kL}(j) \quad (23)$$

Note that the separate estimation of the future component from the present component implies that the marginal effect of the curvature of the per-period utility function (the degree of risk aversion α_i) on future expected utility of attending school is freely estimated.

6 Measuring Risk Aversion

In the survey, each head of household is asked to report the maximum price he/she is willing to pay to participate to an hypothetical lottery. The question is worded as follows:

“We would now like to ask you a hypothetical question that we would like you to answer as if the situation was a real one. You are offered the opportunity of acquiring a security permitting you, with the same probability,

²¹I impose an additive (as opposed to multiplicative) heterogeneity term in order to minimize the number of parameters.

either to gain a net amount of 10 million lire (roughly 5,000 dollars) or to lose all the capital invested. What is the most you are prepared to pay for this security?"²² The respondent can answer in three possible ways: 1) give the maximum price he/she is willing to pay, which we denote as *bet*; 2) don't know; 3) don't want to participate. In total, 3,458 answered they were willing to participate and reported a positive maximum price they were willing to bet (prices equal to zero are not considered a valid response). The valid responses to the question - *bet* - range from 1,000 lire to 100 million lire and constitute our measure of individual risk aversion. Of the 3,288 heads for whom I have access to parents background variables, 3,131 reported a maximum price *bet* less than 10 million lire which implies that they are risk averse individuals, 117 reported *bet* exactly equal to 10 million lire (i.e. they are risk neutral) and 40 reported *bet* more than 10 million indicating that they are risk lovers. The empirical distribution of *bet* is reported in Appendix 1B. Although the majority of the respondents are risk averse and only 5% of the sample is either risk-neutral or risk-loving, there is a large heterogeneity in the degree of risk aversion within the risk averse individuals which shows that preferences are very heterogenous with respect to risk.

In general, the optimal bet depends on $U_i(\cdot)$ and on consumer endowment (w_i). To obtain a measure of risk aversion, I use the one-to-one correspondence between the value attached to the lottery, and the degree of risk aversion (given wealth). Given the static (single period) nature of the lottery, I assume that the per-period utility function, along with the value of the bet, is sufficient to uncover the degree of risk aversion.

In the applied literature, it is common to rely on a Taylor series expansion of the expected utility equation of a lottery for a small risk.²³ In such a case, it is easy to express the Arrow Pratt measure of risk aversion (the degree of absolute risk aversion) in terms of the parameters of the lottery. Such a measure is indeed available in the SHIW. However, in this paper, I use a more direct approach. As is usually the case in structural analysis, I assume a particular parametric form for the per-period utility. This implies that I may re-interpret the maximum bet offered by a given individual as the solution

²²In other words, the expected value of entering the lottery is $0.5 \cdot (10,000,000 - bet)$. Guiso and Paiella (2004) write that the interviews were conducted by professional interviewers at the respondents' homes and to help the respondent to understand the question the interviewers showed them an illustrative card and were ready to provide explanations.

²³For more discussion, see Gollier (2001).

to the following equation;

$$\frac{1}{2}\phi_i(w_i + g; \alpha_i) + \frac{1}{2}\phi_i(w_i - bet_i, \alpha_i) = \phi_i(w_i, \alpha_i) \quad (24)$$

Using the restriction that $U(x) = -\exp(-\alpha_i \cdot x)$, it is easy to solve for the degree of risk aversion; that is to find an individual specific parameter α_i which is uniquely determined by the value of the bet and the gain

$$\alpha_i = \alpha(bet_i, w_i) \quad (25)$$

To do so, I use a simple Gauss-Newton algorithm.²⁴ The results reported in Table 2 indicate that, on average, Italian heads of households are risk averse (the average value of α_i is 0.14). The degree of risk aversion ranges between 0.18 and -0.13. The data indicate that 98% of the sample is characterized by risk aversion.

7 Identification and Estimation

In the structural econometric literature, identification is typically considered within a particular parametric structure.²⁵ The structural literature on schooling and human capital is no exception. In most applications, preference heterogeneity is allowed through heterogeneity in the costs (or the monetary equivalent utility) of schooling and the joint distribution of school and market abilities is inferred from schooling attainments and post-schooling earnings panel data. Differences in the curvature of the utility function are typically ignored²⁶ This is not the case in the present model. While I have observa-

²⁴In order to check the validity of the absolute risk aversion measure used by most of the SHIW users, I computed the correlation between the exact solution to the non-linear equations and the Taylor-series expansion approximation. As the correlation was found to be only 0.10, I disregarded the use of the measure obtained from a Taylor series expansion.

²⁵The common view is that structural models require to specify preferences and technology and therefore imply functional forms and parametric assumptions. The degree of under-identification (non-parametric) is analyzed precisely in Rust (1994) and Magnac and Thesmar (2002) although all of them consider the identification of generic models in which data on outcomes are ignored. Heckman and Navarro (2006) show that non-parametric identification may be obtained in certain type of dynamic structural model, with optimal stopping properties.

²⁶This is the case, for instance, in Keane and Wolpin (1997), Eckstein and Wolpin (1999) and Belzil and Hansen (2002). The typical identification procedure characterizing the structural literature on schooling is surveyed in Belzil (2007).

tion on the individual specific degree of curvature of the utility function, I do not have data on the earnings while in school (the argument of the utility function of attending school) and, furthermore, schooling decisions are affected by a supplementary source of heterogeneity caused by individual specific subjective probability distributions. This is one aspect of identification that has to be tackled. A second one arises because I assume that parents' background variables do not only affect observable components (like labor market earnings), but also affect two distinct unobservable components; namely the present and the future components of the schooling choice determinants. To get around this fundamental degree of under-identification (and obtain parametric) identification, some restrictions have to be imposed.

7.1 Identifying Restrictions

- **Assumption 1- Parameter Space restrictions.** Consistent with the high level of intergenerational education correlation reported in the literature, I assume that i) the difference between the future component of the intertemporal utility of work and the intertemporal utility of school must be decreasing with family education and that ii) both school and market ability increase with parents' education. This ensures that those coming from more educated background must be more likely to obtain higher education. This restricts the parameter space of $F_{1klj}, F_{2klj}, F_{3klj}$ and F_{4klj} to negative values and the parameter space of $\theta_1^W, \theta_2^W, \theta_3^W$ and θ_4^W as well as $\theta_1^S, \theta_2^S, \theta_3^S$ and θ_4^S to positive values.
- **Assumption 2-Common Support points:** I assume that the support points $(\tilde{\theta}_H^W, \tilde{\theta}_M^W, \tilde{\theta}_L^W)$ of the subjective market ability distribution, $\mu_{\theta^w}(\cdot)$, coincide with the support points of the objective probability distribution .
- **Assumption 3- Modal belief and actual type:** I impose that individuals have sufficient information about their actual market ability so to construct subjective beliefs that assign a higher probability of being their actual type than each probability of being of being a counterfactual type²⁷ This entails forcing equality between the actual market ability type and the mode of the subjective distribution. For the more

²⁷This is not the only identifying restriction which is feasible. An alternative strategy would be to restrict subjective probabilities across types.

(less) able, the subjective probabilities are monotonically decreasing (increasing) from $\tilde{\theta}_H^W$ to $\tilde{\theta}_L^W$. Precisely, I assume that

$$\begin{aligned}
p_{H1} &> p_{M1} > p_{L1} \\
p_{H2} &> p_{M2} > p_{L2} \\
p_{M3} &> \max\{p_{H3}, p_{L3}\} \\
p_{M4} &> \max\{p_{H4}, p_{L4}\} \\
p_{H5} &< p_{M5} < p_{L5} \\
p_{H6} &< p_{M6} < p_{L6}
\end{aligned} \tag{26}$$

In order to implement these inequality restrictions, it is sufficient to restrict the parameter space of the elements inside the logistic transform of the subjective probabilities. Details are found in the Appendix

- **Assumption 4- Common Belief:** I assume that all individual (all types) share a common belief about post-schooling random shocks $\mu_t^F(\cdot)$. This allows me to focus solely on beliefs regarding ability.

7.2 Estimation Method

The probability of continuing from one grade level (j) to the next ($j + 1$), $\Pr(j, t)$ is

$$\Pr(j, t) = \Pr[\varepsilon_{it}^S(j) > h^n(j) + \bar{F}_n(j)] \tag{27}$$

It is straightforward to form the likelihood function for an individual who has reached level j^* , and has experienced a series of post-schooling earnings (..) The contribution to the likelihood for the schooling attainment j^* is denoted L_1 , and is

$$L_1 = \prod_{j=1}^{j^*-1} (\Pr(j, t)) \cdot (1 - \Pr(j^*, t)) \tag{28}$$

Given an observed earnings in post-schooling year t (denoted I_t), the density of post-schooling earnings, L_{2t} , is

$$L_{2t} = \sum_{m=1}^M P_m^* \cdot \phi\left(\frac{I_t^W - \mu_m}{\sigma_m}\right) \tag{29}$$

The contribution to the likelihood of the full earnings vector $\{I_{89}^W, I_{89}^W \dots I_{02}^W\}$, L_2 , is

$$L_2 = L_{2,89} \cdot L_{2,92} \cdot \dots \cdot L_{2,02}$$

while the total likelihood is given by

$$L(.) = \sum_{n=1}^6 [L_1(. | type n) \cdot L_2(. | type n)] \cdot \Pr(type n) \quad (30)$$

The parameters are obtained by maximizing the logarithm of the product of each individual specific likelihood functions.

8 Results: Parameter Estimates

In order to estimate the model, I have separated the individuals in two cohorts. The youngest cohort is made of all those individuals born in or after 1956. The second cohort is comprised of those born between 1945 and 1955. The model has been estimated on both cohorts, although I only report the results obtained for Italian Head of households born after 1956.²⁸ In total, the model requires the estimation of around 140 free parameters. As is usually the case in structural dynamic models, not all parameter estimates raise a direct interest. For this reason, I report them in Appendix 4. They are presented in the following blocks; objective type probabilities, earnings shocks distribution (mixture of normals), utility shocks (mixture of normals), market earnings, earnings in school, subjective beliefs, future component heterogeneity, and future component (risk aversion and parents' background).

In the presentation, I shall focus my attention on the parameter estimates of the subjective and the objective ability distribution, but as I go along, I will occasionally refer to particular parameter estimates reported in Appendix 4. Most tables that will be discussed below are devoted to particular quantities implied by the parameters of the model.

²⁸I do this because mostly because the results are very similar, and partly because the presentation of a different set of results would be cumbersome.

9 Present and the Future Components of the Intertemporal Utilities

In Table 1A, I report the average present and future utilities (in differences) for each type. Schooling decisions are partly based upon the actual taste for schooling but also on the subjective distribution of market abilities. As we move from one type to the next, actual (objective) abilities change at the same time as do subjective ability distributions. For this reason, the link between schooling and specific type identity is difficult to evaluate, and is not as limp as it would be in a more standard case.²⁹

However, the type specific utilities may turn out to be very useful in order to comprehend the inherent importance that agents attach to the future (as opposed to the present). In standard structural dynamic models, the difference between the present and the future components of the intertemporal utilities is largely conditioned by the value of the discount factor. In many applications, it is set to a fixed value (usually between 0.9 and 0.95). In other cases, it may be estimated. For a given value of the discount factor, the parametric form of the utilities and their related random shocks translate into corresponding value functions. In practice, this means that the incidence of “present biased” behavior is hardly detectable.

In the present case, the explained part of the model is composed of three independent components of the intertemporal utilities. The present component of the choice determinant is the difference between per-period utilities of school and work. However, the future component is freely estimated and there is also an unexplained part that is represented by the stochastic utility shock. Although the stochastic term is linked to the present component in the model presentation, it could as well be interpreted as an element of the future component. After all, the random shock represents an i.i.d. innovation realized at the time of the decision. Attaching a particular meaning to it is therefore an ad-hoc decision.³⁰

The future components (reported for the differences in the utility of at-

²⁹In the structural dynamic discrete choice literature (Keane and Wolpin, 1997), it is common to illustrate the role of heterogeneity by presenting type specific schooling decisions. Because agents are assumed to know their ability endowments, it is easy to associate schooling/career profiles, to individual persistent taste for schooling.

³⁰In Houser, Keane and McCabe (2004), the unexplained part of the model is interpreted as an optimization error which occurs when participants to a laboratory experiment attend to solve a Bellman equation.

tending junior high-school, senior high-school and higher education) appears to be, by far, the dominant factor generating schooling decisions. More precisely, for all types, the difference in current utilities are close to 0 while the difference in future utilities moves from large negative numbers at lower schooling levels (between -14 to -23) to a large positive number. While the type specific average values indicate a higher importance of the future component, it is also important to vary the degree of cross-sectional dispersion. Recall that a higher value for the present or the future components indicates a higher propensity to leave school since the component measures the difference between the utility of work and the utility of school.

In Table 1B, I report the standard deviation of the present and future components. The standard deviations reflect variations across types and also take into account heterogeneity parents background. The numbers indicate that the present component is not more dispersed than the future component. At all grade levels, the degree of heterogeneity in the future component exceeds the one corresponding to the present component. It also exceeds the standard deviation of the random shocks (equal to 7.7 at all levels).

In recent years, several economists have questioned the foundation of the standard time-consistent intertemporal model and, in particular, have pointed out particular economic behaviors that are may be characterized by “present bias”. To the extent that the results reported in Table 1A and 1B indicate a predominance of future components in schooling decisions, they seem to be supportive of the standard time consistent model.³¹

10 The Objective Distribution of Skill Heterogeneity

The objective distribution of ability heterogeneity is summarized in Table 3A. The population is composed of 17% of high ability individuals (for whom θ^W is 15.56), of 49% of medium ability individuals (for whom θ^W is 9.65) and of 34% of low ability individuals (for whom θ^W is 5.12). The objective distribution is characterized by a variance equal to 12.50. At a later stage, I will compare the importance of heterogeneity to the degree of ex-ante risk perceived by individuals. At the same time, the population is composed of

³¹For a theoretical discussion of the foundations of Behavioral Economics, see Rubinstein (2005).

74% of individuals with high taste for schooling (for whom θ^S is 13.72) and 26% of individuals with low taste (-1.62).

The effect of parents background variables on both school and market abilities (the $\theta^{S'}$ s and the $\theta^{W'}$ s) are also found in appendix. Consistent with what has been reported in the structural literature, the effect of parents' background on the utility of attending school is much stronger than on the earnings equation's intercept. This is a standard result in the structural literature on schooling and it does not require further discussion.

11 The Subjective Beliefs

In this section, I summarize the results that allow me to characterize the subjective beliefs. I first analyze the relative incidence of over-estimation and under-estimation (the measure of location of the subjective distribution). Then, I consider the notion of confidence (the measure of spread) and quantify the degree of ex-ante risk.

11.1 Over or Under Estimation? Measuring Accuracy

The subjective distribution of labor market skills is reported in Table 3B. In Table 3C, I report three important summary statistics that characterize the subjective beliefs; the mean belief, the bias (as defined by the difference between the mean belief and the true (objective) value), and the degree of ex-ante risk as evaluated by the variance of the subjective distribution. Note that a negative (positive) bias indicates the incidence of under-estimation (over-estimation). To a certain extent, this allows me to investigate the notion of "over-confidence", which has appeared recently in the behavioral economics. However, it is important to note that in the literature, it is typically defined in a context where the variable (or the parameter) that characterizes over-confidence is non-stochastic.³² As my approach is more general because I distinguish between location and spread, I prefer to use the term "over-estimation" (or under-estimation) when referring to the central location parameter of the distribution of subjective beliefs. As will be clear in the next section, when analyzing the spread of the distribution, I use the

³²Indeed, over-confidence is usually defined in general terms without any specific detail.

term “confidence”.³³

As may be inferred from Table 3B and Table 3C, a substantial level of heterogeneity characterizes subjective beliefs. For both the most able (type 1 and type 2) and the least able (type 5 and type 6) individuals, there exists one type that are endowed with degenerate subjective beliefs that coincide with their actual type. These are type 2 and type 5 individuals. For these types of individuals, both the bias and the ex-ante risk are equal to 0.(see Table 3C).³⁴

Among the more able, under-estimation appears to be important. Individuals of type 1 attribute only a probability equal to 0.46 to their actual type. The bias that characterize their beliefs (-3.60) is the highest among all types. This high level of dispersion translates into a high degree of ex-ante risk, which is equal to 12.57. Overall, type 1 individuals represent 65% of the able individuals.

At the lower end, over-estimation appears to be minor when compared to the degree of under-estimation for the more able, as type 6 individuals appear to over-estimate their ability only by a slight margin. Their subjective probability of being their actual type, 0.87, is reasonably close to 1 and drives the bias at a value of 0.59 and the degree of ex-ante risk at 2.32.

Finally, in the medium ability range (type 3 and type 4 individuals), both subjective distributions are characterized by a positive bias and therefore imply over-estimation. In the case of type 3 individuals, the subjective probability assigned to the actual type, 0.87, implies only a modest bias (0.04) but the almost equal probability assigned to the high ability and the low ability values (0.06 and .07) implies a relatively important level of risk (3.53). Both over-estimation and ex-ante risk are more important among type 4 individuals. The bias (1.89) is the second highest in absolute value, while the variance of the subjective distribution (7.60) is also the second highest.

Overall, when considering types as separate sub-populations, over-estimation (or over-confidence) appears to me more frequent than under-estimation (under-confidence). However, the degree of under-estimation found among the more able is much more severe (in absolute value) than the degree of

³³Arguably, the term “over-confidence” used in the behavioral economics literature is somewhat misleading since it conveys a notion of ex-ante risk.

³⁴The reader should note that the restriction on the mode of the subjective distribution implies that, in a case of degenerate beliefs, there cannot be any bias. In words, this rules out the case where an agent is certain of his type, but is actually wrong about it.

over-estimation found among the medium and the low ability spectra.

11.2 Ex-Ante Risk: Measuring the Degree of Confidence

As reported in Table 3A, the variance of the objective ability distribution (as perceived by the econometrician) is relatively high (12.50). When averaged over all types, the level of ex-ante risk (as measured by the average variance over each subjective ability distribution) is equal to 3.81. Defining the degree of confidence of the agent as the ratio of the variance of the objective distribution (as perceived by the econometrician) to the variance of the subjective distribution, would lead to a degree approximately equal to 4. In other words, only 25% of unobserved ability heterogeneity is actually perceived as ex-ante risk. Given the relatively accurate characterization of their own ability (as indicated by the low average bias), these number therefore imply that agents have a substantial amount of inside information, when compared to the econometrician (they are four times more informed). Indeed, as seen earlier, the type specific analysis revealed that around 36% of the population act on the basis of a degenerate subjective ability distribution.

11.3 Accuracy, Confidence and Objective Market Ability.

Given the importance of heterogeneity that characterizes the subjective beliefs, it would be important to relate both a measure of accuracy (the absolute value of the bias or the bias squared) and a measure of ex-ante risk (the variance of the type specific subjective distributions) to actual market ability. This may be achieved informally. A quick inspection of Table 3C reveals that both the incidence of a large bias and a high degree of dispersion are confined to high ability individuals, and to a lesser extent, the medium ability types. Those endowed with low ability appear to have an accurate depiction of their true ability.³⁵

³⁵This may be verified formally. To do so, I performed an ordinary regression of both the bias squared and the ex-ante risk on the actual heterogeneity component (school and market abilities). The regressions indicate that both the level ex-ante risk and the degree of inaccuracy is increasing with market ability.

Obviously, this does not necessarily indicate that the more able are less capable to form accurate beliefs than the less able. In a rigid labor market, characterized by the presence of minimum wages and income support policies for the disadvantaged, the wage distribution is more likely truncated from below. It is therefore reasonable to expect low wage workers to use a relatively accurate distribution of their potential wages. For the more able individuals, wages are intrinsically more variable.

12 A Summary of the Major Findings

In this section, I summarize 6 major findings.

1. At all grade levels, both the average values and the degree of dispersion of the future component of the intertemporal utilities exceed the equivalent measure for the present component. The predominance of future components in schooling decisions, seem to be supportive of the standard time consistent model. However, at the same time, idiosyncratic utility shocks appear to be an important determinant of schooling decisions.
2. There is a high degree of heterogeneity in subjective beliefs, but it is explained mostly by the importance of heterogeneity in the objective ability distribution (identified from panel data). Subjective beliefs are found to be quite accurate on average (within-type dispersion is small).
3. Precisely, 36% of the population is found to take decisions based on a degenerate ability distribution.
4. I find a strong incidence of under-estimation among the more able but a small incidence of over-estimation among the low ability group. At the medium ability spectrum, there is evidence of some over-estimation.
5. Defining the degree of confidence as the ratio of the variance of the objective distribution (as perceived by the econometrician) to the variance of the subjective distribution, would lead to a degree approximately equal to 4. In other words, only 25% of unobserved ability heterogeneity is actually perceived as ex-ante risk. The degree of confidence is high and imply that agents have a substantial amount of inside information.

6. Both the level of ex-ante risk (the degree of non-confidence) and the degree of inaccuracy (the incidence of either under-estimation or over-estimation) are increasing with market ability.

13 Conclusion Remarks

As far as I know, this paper is the first attempt to obtain micro-econometric estimates of subjective beliefs in a context where the econometrician has no access to such direct measures. While using elicited beliefs would always be the preferred strategy, it is important to note that such data are rarely available. For this reason, I believe that this approach is promising.

In line with the recent creation of data sets that incorporate various measurements on psychometric and/or risk attitude questions, it would be interesting to use factor analysis in order to investigate deeper issues regarding subjective risk evaluation and risk aversion.³⁶

Evidently, the structure of the models could also be modified in various ways. As I have focussed on schooling decisions at the detriment of post schooling choices, and used data on post schooling earnings, the model implicitly disallows individuals to take actions to correct their mistakes. It would therefore be interesting to endogenize post schooling earnings growth.

Finally, this approach could also be applied in a treatment effect heterogeneity framework so to model the incidence of over-education (or under-education). It could equally be applied to model the duration of unemployment, or job search behavior.

References

- [1] Bellemare, Charles, Kröger, Sabine and Arthur van Soest (2005), “Actions and Beliefs: Estimating Distribution-Based Preferences Using a Large Scale Experiment with Probability Questions on Expectations” IZA Working Paper 1666.
- [2] Belzil, Christian and Marco Leonardi (2007) “Risk Aversion and Schooling Decisions” IZA Working Paper.

³⁶Such measures are now available in the German Socio Economic Panel.

- [3] Belzil, Christian (2007) "The Return to Schooling in the Structural Dynamic Models: A Survey " IZA Working paper, forthcoming in *The European Economic Review*.
- [4] Belzil, Christian and Hansen, Jörgen (2007) "A Structural Analysis of the Correlated Random Coefficient Wage Regression Model", forthcoming in *Journal of Econometrics*
- [5] Belzil, Christian. and Jörgen Hansen (2002) "Unobserved Ability and the Return to Schooling" *Econometrica*, 70(5), 2075-2091.
- [6] Camerer, Colin and Dan Lovallo (1999) "OverConfidence and Excess Entry: An experimental Approach" *The American Economic Review*, vol 89, no 1 (March), 302-318
- [7] Cunha, F., J. Heckman and Salvador Navarro (2005) "Separating Uncertainty from Heterogeneity in Life Cycle Earnings" NBER Working Paper 11024.
- [8] Eckstein, Zvi and Kenneth Wolpin (1999) "Why Youths Drop Out of High School: the Impact of Preferences, Opportunities and Abilities" *Econometrica*, 67(6), 1295-1339.
- [9] Flinn, Christopher (2002) "Labour Market Structure and Inequality: A Comparison of Italy and the U.S." *Review of Economic Studies*, 69(3), 611-645.
- [10] Geweke, J. and M. Keane (2000). Bayesian Inference for Dynamic Discrete Choice Models without the Need for Dynamic Programming. Simulation Based Inference and Econometrics: Methods and Applications, Mariano, Schuermann and Weeks (eds.), Cambridge University Press, 100-131. (Also available as Federal Reserve Bank of Minneapolis WP # 564, January 1996).
- [11] Gollier, Christian (2001), *The Economics of Risk and Time*, The MIT Press.
- [12] Guiso, L. and Monica Paiella (2001) "Risk Aversion, Wealth and Background Risk" CEPR Discussion Paper 2728.

- [13] Guiso, L. and Monica Paiella (2005) "The Role of Risk Aversion in Predicting Individual Behaviour" Banca d'Italia Temi di Discussione 546.
- [14] Guiso, L., T. Jappelli and L. Pistaferri (2002) "An Empirical Analysis of Earnings and Employment Risk" *Journal of Business and Economic Statistics*, 20, 1-13.
- [15] Heckman, J., Lochner, L. and Petra Todd (2005) "Earnings Functions, Rates of Return, and Treatment Effects: The Mincer Equation and Beyond" NBER Working Paper 11544.
- [16] Heckman, J. and Salvador Navarro (2007) "Dynamic Discrete Choices and Dynamic Treatment Effects" *Journal of Econometrics (April)*.
- [17] Houser, Daniel, Michael Keane and Kevin McCabe (2004) "Behavior in a dynamic decision problem: An analysis of experimental evidence using a Bayesian type classification algorithm," *Econometrica*, 72:3 (May), 781-822.
- [18] Rubinstein, Ariel (2005) "Discussion of Behavioral Economics" forthcoming in *Advances in Economics and Econometrics Theory and Applications, Ninth World Congress*

Table 1A
The average Present and Future components
of the Intertemporal utilities

	Average Components					
	<i>j</i> =Elementary		<i>j</i> =Junior HS		<i>j</i> =Senior HS	
	Present	Future	Present	Future	Present	Future
	$P(j)$	$F(j)$	$P(j)$	$F(j)$	$P(j)$	$F(j)$
type 1 (θ_H^S, θ_H^W)	-0.01	-13.3	0.06	-3.62	0.15	9.43
type 2 (θ_L^S, θ_H^W)	0.2	-14.7	1.0	16.8	0.74	-7.05
type 3 (θ_H^S, θ_M^W)	-0.4	-14.4	-0.3	3.91	-0.2	9.24
type 4 (θ_L^S, θ_M^W)	0.1	-15.4	0.9	-6.8	0.8	4.1
type 5 (θ_H^S, θ_L^W)	0.1	-23.5	0.2	-3.6	0.3	11.2
type 6 (θ_L^S, θ_L^W)	0.6	-19.4	1.4	-10.1	1.6	11.9
average over types	0.1	-17.4	0.3	-0.7	0.5	8.2

Note: The present ($P(j)$) and future ($F(j)$) are defined as the difference in per-period utilities of school and work and the difference in future utilities of school and work for those who have already completed level j . The components reported are for the transition from elementary to junior high school, junior high school to senior high school, and senior high school to higher education.

Table 1B
The Relative Importance of the
Present and the Future Components

	Standard Deviations		
	<i>j</i> =elementary	<i>j</i> =Junior HS	<i>j</i> =Senior HS
Present			
$P(j)$	1.7	1.9	5.3
Future			
$F(j)$	16.4	8.4	6.2
utility shocks			
$\varepsilon_{it}^S(j)$	7.7	7.7	7.7

Note: The standard deviations represent the dispersion in the present ($P(j)$) and future ($F(j)$) components across all types.

Table 1C
The Type specific grade termination rates

Fraction Leaving School

$j = \mathbf{Elementary}$ $j = \mathbf{Junior HS}$ $j = \mathbf{Senior HS}$

type 1 (θ_H^S, θ_H^W)	0.42	0.67	0.97
type 2 (θ_L^S, θ_H^W)	0.24	0.98	0.32
type 3 (θ_H^S, θ_M^W)	0.27	0.87	0.97
type 4 (θ_L^S, θ_M^W)	0.21	0.47	0.92
type 5 (θ_H^S, θ_L^W)	0.00	0.70	0.98
type 6 (θ_L^S, θ_L^W)	0.08	0.33	0.98
average	0.18	0.72	0.92
Actual	0.23	0.79	0.91

Note: The grade termination (hazard) rates should be understood as the conditional probability of leaving school with a particular grade level already completed

Table 2
Risk Aversion: Summary Statistics

	Risk Aversion
Average	0.14
Min	-0.13
Max	0.18
St. dev	0.02
fraction risk averse	0.98

Note: the estimates of the risk aversion parameter are obtained from the solution of the expected utility equation corresponding to the lottery.

Table 3A
Skill Heterogeneity
The Objective Distribution of Market Ability

	θ^S	θ^W	type Prob.
type 1	13.7248	15.5604	0.11
type 2	-1.6169	15.5604	0.06
type 3	13.7248	9.6504	0.33
type 4	-1.6169	9.6504	0.16
type 5	13.7248	5.1204	0.30
type 6	-1.6169	5.1204	0.04
Mean	10.42	9.25	-
Variance	63.24	12.50	-

Table 3B
Subjective Beliefs:
The type specific subjective probability distributions

	Actual Type		Potential Type		
	School ab	Market. ab	High $\tilde{\theta}_i^W = 15.56$	Medium $\tilde{\theta}_i^W = 9.65$	Low $\tilde{\theta}_i^W = 5.12$
type 1	High	High	0.46	0.45	0.09
type 2	Low	High	1.00	0.00	0.00
type 3	High	Medium	0.06	0.87	0.07
type 4	Low	Medium	0.32	0.68	0.00
type 5	High	Low	0.00	0.00	1.00
type 6	Low	Low	0.00	0.13	0.87

Table 3C
Mean, Bias and Ex-ante risk characterizing the Subjective Beliefs

	Mean	Actual value	Bias	Risk (variance)
type 1 (θ_H^S, θ_H^W)	11.96	15.5604	-3.60	12.57
type 2 (θ_L^S, θ_H^W)	15.56	15.5604	0.00	0.00
type 3 (θ_H^S, θ_M^W)	9.69	9.6504	0.04	3.53
type 4 (θ_L^S, θ_M^W)	11.54	9.6504	1.89	7.60
type 5 (θ_H^S, θ_L^W)	5.12	5.1204	0.00	0.00
type 6 (θ_L^S, θ_L^W)	5.71	5.1204	0.59	2.32
Average over types	9.08	9.25	-	3.81

Note: The type specific bias is the difference between the type specific mean belief and the true (objective) value). A negative (positive) bias indicates the incidence of under-estimation (over-estimation).

Appendix 1
Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
bet (in 000 liras)	2513.083	4798.066		1	100000
edu1 (graduate)	0.005	-		0	1
edu2 (universirty)	0.099	-		0	1
edu3 (senior high school)	0.461	-		0	1
edu4 (junior high school)	0.392	-		0	1
edu5 (elementary)	0.038	-		0	1
edu6 (no educ)	0.005	-		0	1
edu_father	0.127	-		0	1
edu_mother	0.088	-		0	1
north	0.406	-		0	1
south	0.423	-		0	1
female	0.179	-		0	1
wc_father	0.250	-		0	1
wc_mother	0.101	-		0	1
age (in 1995)	32.45	3.91		20	38

Appendix 2
Yearly Earnings Data

year	mean	St. Dev.	Min	Max
1989	16551	5577	2205	40847
1991	17015	8936	1807	216889
1993	17193	8094	592	72472
1995	16118	7631	421	78217
1998	16735	7811	507	84504
2000	17186	7795	272	68062
2002	17037	8361	1000	100000

Note: Yearly Earnings are measured in 2002 euros. The numbers include all cohorts.

Appendix 3 Further parameterization

Objective type probabilities:

$$\Pr(\text{type } n) = \frac{\exp(\omega^n)}{1 + \exp(\omega^1) + \dots + \exp(\omega^5)}$$

Earnings random shocks (normal mixtures):

$$P_m^W = \frac{\exp(p_m^{W*})}{1 + \exp(p_1^{W*}) + \exp(p_2^{W*})} \text{ for } m = 1, 2 \text{ and } p_3^{W*} = 0$$

$$\mu_1^W = -\exp(\mu_1^{S*}), \mu_2^W = 0, \mu_3^W = \exp(\mu_3^{S*})$$

$$\sigma_m^S = \exp(\sigma_m^{W*})$$

Utility Shocks (normal mixtures):

$$P_m^S = \frac{\exp(p_m^{S*})}{1 + \exp(p_1^{S*}) + \exp(p_2^{S*})} \text{ for } m = 1, 2 \text{ and } p_3^{S*} = 0$$

$$\mu_1^S = -\exp(\mu_1^{S*}), \mu_2^S = 0, \mu_3^S = \exp(\mu_3^{S*})$$

$$\sigma_m^S = 1 \text{ for } m = 1, 2, 3$$

Parents' background/market abilities:

$$\theta_1^W = \exp(\theta_1^{W*}), \theta_2^W = \exp(\theta_2^{W*}), \theta_3^W = \exp(\theta_3^{W*}), \theta_4^W = \exp(\theta_4^{W*})$$

parents' background/school abilities:

$$\theta_1^S = \exp(\theta_1^{S*}), \theta_2^S = \exp(\theta_2^{S*}), \theta_3^S = \exp(\theta_3^{S*}), \theta_4^S = \exp(\theta_4^{S*})$$

Future component:

$$F_{2j} = -\exp(F_{2j}^*), F_{3j} = -\exp(F_{3j}^*), F_{4j} = -\exp(F_{4j}^*), F_{5j} = -\exp(F_{5j}^*) \forall j$$

Subjective Beliefs:

type 1/type 2

$$\mu_H^{1,2} = \exp(\mu_1^{1,2}) + \exp(\mu_2^{1,2}), \mu_M^{1,2} = \exp(\mu_2^{1,2}), \mu_L^{1,2} = 0$$

Subjective Beliefs:

type 3/type 4

$$\mu_H^{3,4} = \exp(\mu_1^{3,4}), \mu_M^{3,4} = \exp(\mu_1^{3,4}) + \exp(\mu_2^{3,4}), \mu_L^{3,4} = 0$$

OR

$$\mu_H^{3,4} = 0, \mu_M^{3,4} = \exp(\mu_2^{3,4}) + \exp(\mu_3^{3,4}), \mu_L^{3,4} = \exp(\mu_3^{3,4})$$

Subjective Beliefs:

type 5/type 6

$$\mu_H^{5,6} = 0, \mu_M^{5,6} = \exp(\mu_2^{5,6}), \mu_L^{5,6} = \exp(\mu_2^{5,6}) + \exp(\mu_3^{5,6})$$

Appendix 4
Structural Estimates

	estimate	st-error
Type prob.		
ω^1	0.5411	1.52
ω^2	0.3259	1.30
ω^3	2.1311	8.35
ω^4	0.5598	1.91
ω^5	1.8588	6.18
ω^6	-	-
earnings shock		
p_1^{W*}	3.4441	0.13
p_2^{W*}	2.6870	0.16
p_3^{W*}	0.00	-
μ_1^{W*}	-9.5008	0.21
μ_2^{W*}	-	-
μ_3^{W*}	0.3710	0.47
σ_m^{W*}	-0.1073	0.20
σ_m^{W*}	0.9387	0.21
σ_m^{W*}	-15.5809	0.65
Utility shocks		
p_1^{W*}	-0.9025	0.38
p_1^{W*}	2.2243	0.12
p_1^{W*}	-	-
μ_1^{S*}	1.7120	0.15
μ_1^{S*}	-	-
μ_1^{S*}	3.0763	0.08
σ_1^{S*}	-	-
σ_2^{S*}	-	-
σ_3^{S*}	-	-

Appendix 4 (continued)

	estimate	st-error
Market Earnings		
$\tilde{\theta}_H^W$	15.4604	0.31
$\tilde{\theta}_M^W$	9.6504	0.26
$\tilde{\theta}_L^W$	5.1204	0.29
ζ_{univ}	8.3444	0.38
$\zeta_{u.H.S.}$	6.5436	0.25
$\zeta_{u.L.S.}$	1.3004	0.22
ζ_{elem}	0.00	-
experience	11.7200	0.69
experience ²	-3.9448	0.22
$\theta_1^W(FED)$	-5.9525	0.21
$\theta_2^W(MED)$	-10.8193	0.22
$\theta_3^W(FWC)$	-7.8566	0.20
$\theta_4^W(MWC)$	-7.4735	0.19
$\theta_5^W(female)$	-3.6078	0.28
$\theta_6^W(North)$	0.3950	0.18
$\theta_7^W(South)$	-1.3824	0.19
earnings in school		
$\tilde{\theta}_H^S$	13.7248	0.25
$\tilde{\theta}_L^S$	-1.6169	0.22
ζ_{univ}^S	-0.0889	0.20
$\zeta_{U.H.S.}^S$	-1.1805	0.30
$\zeta_{L.H.S.}^S$	6.2336	1.41
ζ_{elem}^S	-	-
$\theta_1^S(FED)$	-1.6710	0.22
$\theta_1^S(MED)$	1.7110	0.19
$\theta_1^S(FWC)$	-4.9025	0.22
$\theta_1^S(MWC)$	-3.4981	0.22
$\theta_1^S(female)$	-5.4962	0.54
$\theta_1^S(North)$	8.6653	0.53
$\theta_1^S(South)$	-1.2457	0.72

Appendix 4 (continued)

	estimate	st.error
Subj beliefs		
type 1		
μ_1^1	-2.6024	0.81
μ_2^1	-5.2051	1.31
μ_3^1	0.0000	-
type 2		
μ_1^2	1.5132	0.31
μ_2^2	-1.0272	0.27
μ_3^2	-0.1073	-
type 3		
μ_1^3	1.3763	0.22
μ_2^3	1.8312	0.22
μ_3^3	0	-
type 4		
μ_1^4	0.2371	0.24
μ_2^4	1.4390	0.39
μ_3^4	0	-
type 5		
μ_1^5	0	-
μ_2^5	-0.1526	0.24
μ_3^5	2.7275	0.50
type 6		
μ_1^6	0	-
μ_2^6	3.5198	0.92
μ_3^6	-3.0832	0.80

Appendix 4 (continued)

	estimate	st. error
Future comp. heterogeneity (F_{oklj})		
F_{0HH1}	-14.4515	0.53
F_{0LH1}	-10.2853	0.46
F_{0HM1}	-9.8577	0.58
F_{0LM1}	-14.8952	0.87
F_{0HL1}	-7.4294	0.91
F_{0LL1}	-9.1069	0.57
F_{0HH2}	-3.8859	0.47
F_{0LH2}	-3.8189	0.37
F_{0HM2}	-16.0533	0.35
F_{0LM2}	-4.6045	0.38
F_{0HL2}	-9.3889	0.59
F_{0LL2}	-10.8507	0.72
F_{0HH3}	-9.4381	0.93
F_{0LH3}	7.1000	0.86
F_{0HM3}	-1.6309	0.64
F_{0LM3}	21.0022	0.43
F_{0HL3}	-12.6764	0.29
F_{0LL3}	-5.2139	0.21
F_{0HH4}	6.7245	0.22
F_{0LH4}	5.4495	0.48
F_{0HM4}	12.4105	0.36
F_{0LM4}	-5.2858	0.37
F_{0HL4}	10.5117	0.74
F_{0LL4}	11.1104	0.67
F_{0HH5}	11.9737	0.66
F_{0LH5}	23.3494	0.62
F_{0HM5}	22.7861	0.37
F_{0LM5}	20.8798	0.34
F_{0HL5}	23.0335	0.38
F_{0LL5}	17.4454	0.40

Appendix 4 (continued)

	estimate	st. error
future component (risk aversion & parents' background)		
$F_{11}(\alpha_i)$	-1.8112	0.22
$F_{21}(FED)$	3.2393	0.32
$F_{31}(MED)$	0.7484	0.13
$F_{41}(FWC)$	2.7521	0.20
$F_{51}(MWC)$	0.8819	0.18
$F_{12}(\alpha_i)$	-1.8146	0.66
$F_{22}(FED)$	2.4683	0.25
$F_{32}(MED)$	2.6411	0.34
$F_{42}(FWC)$	2.8356	0.33
$F_{52}(FWC)$	2.7586	0.25
$F_{13}(\alpha_i)$	-4.7483	0.66
$F_{23}(FED)$	2.2273	0.43
$F_{33}(MED)$	1.2871	0.33
$F_{43}(FWC)$	1.1604	0.38
$F_{53}(MWC)$	1.8164	0.29
$F_{14}(\alpha_i)$	-1.1640	0.15
$F_{24}(FED)$	1.6738	0.27
$F_{34}(MED)$	1.1331	0.34
$F_{44}(FWC)$	1.2897	0.24
$F_{54}(MWC)$	-3.8955	0.22
$F_{15}(\alpha_i)$	0.0353	0.20
$F_{25}(FED)$	-4.0968	0.22
$F_{35}(MED)$	-4.3264	0.26
$F_{45}(FWC)$	-3.9703	0.30
$F_{55}(MWC)$	-3.4445	0.30