NATURE, NURTURE AND MARKET CONDITIONS: ABILITY AND EDUCATION IN THE POLICY EVALUATION APPROACH*

Eduard Gracia and Bernarda Zamora**

WP-AD 2007-29

Correspondence: Departament d'Economia. Universitat Jaume I, Campus Riu Sec, E-12071 Castellón (Spain). Tf.: +34964387166. Fax: +34964728591. e-mail: zamoram@eco.uji.es.

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.

Primera Edición Diciembre 2007

Depósito Legal: V-214-2007

Ivie working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication.

---

* Special thanks to Gema Zamarro for useful discussion. Bernarda Zamora gratefully acknowledges financial support by the Spanish Ministry of Education (project SEJ2005-02829/ECON). The usual disclaimer applies.

** E. Gracia: Booz Allen Hamilton (consulting firm); B. Zamora: Universitat Jaume I.

NATURE, NURTURE AND MARKET CONDITIONS:
ABILITY AND EDUCATION IN THE POLICY EVALUATION APPROACH

Eduard Gracia and Bernarda Zamora

ABSTRACT

The present paper follows the rationality of the Human Capital Theory to explain the heterogeneity of returns to schooling in a policy evaluation model with the purpose of testing whether people are blocked in any way (credit constraints, uncertainty or other market environment conditions) when they make their schooling choices. The minimal assumption that abler people face lower costs of schooling guarantees the possibility of making the right choice in this framework. The empirical implications of the model are extended further from the properties of ordinary least squares and instrumental variable estimators and centred on predictions about the sign of different policy evaluation parameters (sorting gains and selection biases) and on the shape and variability of marginal returns to education. Within this framework, the paper revises the modern empirical literature on returns to schooling in combination with the theoretical literature on human capital. Empirical evidence for the U.S., shown by a binary choice model, supports the assumption. Evidence obtained from Spanish data in a sequential choice setting does not support the assumption.

JEL Classification: C10, D84, I21, J31

Keywords: Ability gap, Schooling, Selection Models, Heterogeneity

"Another set of paired opposites which tend to merge in real life are nature and nurture. For what we come into the world with -our innate qualities, our "constitutional givens"- interacts with the nurture we receive. We cannot view development in terms of either environment or heredity. We must include both"

- Judith Viorst, *Necessary Losses*

1. INTRODUCTION

In 1975, in the second edition of his landmark work, *Human Capital*, Gary Becker set the foundations for an explanation of wage differentials as a return on an investment in education[1]. Arguably, however, wage differentials cannot be interpreted solely as a result of different investments in education, for multiple other factors, such as health, luck, discrimination, nepotism or even trade union power all play a crucial role as well. Therefore, in the case there is a possibility of choice, the implications of the theory on the determination of the amount of education invested depend on the expectation of economic returns that such an investment provides based upon the maximization of the expected value of lifetime utility. For compulsory education, there is no possibility of choice of the amount of education invested but there is a possibility of choice of school quality, that is, of the amount of skills learned during this level of education. For non-compulsory education, either the individual or the family decides on both the amount and the quality of the investment in education.

For policy makers, as Taber (2001) acknowledges, the distinction between types of education which prompt learning and then economic productivity is crucial to the design of policies aimed at subsidizing the skills learned in each level of education. But the value added of acquired skills, which influences the individual choices, is determined in the market and therefore depends on the market environment and regulations that characterize a specific market (sector, country, etc.). For instance, in a socialist market such as the one prevailing in the ex-soviet republics, the skills acquired during college education, of any kind and quality, had the same value from a private point of view, such that the individual did not have economic incentives upon which to base her choice. In this kind of environment, where returns and costs of schooling are independent of the level and type of schooling, optimal aggregate allocations of persons to schooling are determined by an arbitrage condition which equalizes returns across choices. At the other extreme, in a market system that assigns economic rewards based on individual merit, economic incentives at the individual level play the main role in the choice of the amount and quality of education.

[1] The theoretical model on education and the distribution of earnings had been previously developed by Becker and Chiswick (1966) and Becker (1967).

If the market environment provided with the correct economic incentives to education and rational individuals had perfect information for their decisions, there would be no scope for *ex-post* regret, that is, most people would not change their previous schooling decisions if they could make the choice with hindsight. The correct economic incentives are provided if the economic system allows individuals to have a sorting gain. In other words, the economic system allows sorting of individuals among different levels of education based on their idiosyncratic gains and costs of studying. Credit constraints of any type block the right decision based upon the sorting gain. Carneiro and Heckman (2002) describe two types of credit constraints: (i) short term credit constraints which are of a financial type, and (ii) long term credit constraints which are those which block the development of cognitive and non-cognitive skills derived from the parental environment. At the same time, the level of uncertainty plays a role in the chances of making the right choice.

Another form of distortion of the individuals' incentives is the signaling role of education i.e. the fact that academic credentials "signal" to the market that the individual possesses certain superior capabilities (say IQ, or drive, or a strong work ethic) required to complete those studies, even though the education per se did not create those capabilities and, therefore, they could theoretically also be found in non-educated people. In this case, the premium that the market would be willing to pay for educated people would not reflect their superior capabilities as much as the lower risk that hiring them entails, since the presence of those desired abilities has already been tested in them. Whether this mechanism has a large or a small impact on the labor market, however, cannot be determined in abstract, for different countries and industries will respond differently to signaling. Thus, for example, in countries with a very heavy public-sector component and/or a very strict labor regulation, signaling is likely to play a much stronger role than in more "liberal" or dynamic economies. In the former economies, we should expect education to have a smaller positive impact on productivity, since it is not used to increase the productivity of the educated individuals, but only as a means to filter the most productive ones (furthermore, to the extent going through many years of education can be a fairly expensive process in terms of both direct and opportunity costs, the overall aggregate impact on productivity could even be negative).At the same time, one would also expect the income variability for those with higher academic credentials to be smaller than for the population as a whole because their offering is perceived, after their educational filtering, as a more standardized offering and, therefore, if this signaling aspect is predominant, their premium should also be expected to be fairly homogeneous within the same educational level.

This paper puts forward a new methodology to test whether individuals, having the right information, are blocked in any way, i.e., by credit constraints, uncertainty, or the economic environment itself, when they make their schooling choice based on wealth maximization. Within this framework, the paper revises the modern empirical literature on returns to schooling in combination with the theoretical literature on human capital.

People make the right choice in this sense when they make a gain from self-selecting or sorting into a specific schooling level by comparison with the gain they would have obtained in the other level (the counterfactual state). Our methodology consists of deriving the empirical consequences on the sign of several parameters (sorting gains and selection bias parameters), on the shape of the distribution of returns to education, and on its variability, as obtained in policy evaluation models. In this sense, the present paper follows the rationality made by Card (1999, 2001) to explain the heterogeneity of returns to schooling and extends its empirical implications further from the properties of ordinary least squares and instrumental variable estimators.

The paper is structured as follows. In the next two sections, we reinterpret the concept of “ability” through the methodology of policy evaluation. To this end we define the term “ability gap” so that it captures the individual’s latent learning capabilities. We show that, under the assumption that abler people face lower marginal costs of schooling, we can derive empirical properties on several parameters measuring sorting gains and selection bias, as well as on the relationship between the distribution of marginal returns and the probability of going to school. We derive our empirical implications as the necessary conditions or properties we should observe, in a binary and in a sequential choice framework, such that empirical evidence supports the above assumption. Then, in Section four, we proceed to analyze empirical evidence for two countries: the U.S. and Spain. In the U.S. case, it seems that most of people have well founded incentives on which to base their schooling choices, such that empirical evidence supports the assumption that abler people face lower costs. On the contrary, in Spain, it seems that there exists a big burden of psychic costs on the person-specific returns to education, such that empirical evidence does not support the assumption of a negative correlation between “ability” and costs of schooling.

2. THE RELATIONSHIP AMONG UNOBSERVABLES IN THE DEMAND OF EDUCATION: THE ABILITY GAP.

In structural models of the demand of education, the schooling decision is the result of an optimization problem in which the value function is the net discounted value of the expected gains derived from choosing a particular schooling level (see for example the pioneering model of Willis and Rosen 1979). In every one of these structural models, the gains from education, measured by earnings gains, can be decomposed into a first observable component related to observed characteristics (observed by both the econometrician and the agent who makes the choice) plus an unobservable component which contains intrinsic characteristics of person *i*, such as talents, as well as non-observed environmental characteristics and `luck'. The agent who makes the choice can have private information *ex ante* on part of these unobservables although she can also face some uncertainty on them.

Let's Y_k^i denote the realized earnings or *ex post* present value of earnings which person *i* obtains if she achieves the level *k* of education. The observed component of such earnings is denoted by $X^i\beta_k$, being β_k the market retribution of the observed characteristic X^i. The unobserved component at the moment of the schooling decision, denoted by U_k^i, contains the wage premium associated with individual characteristics for which the market has not *ex ante* information but which can be partially contained in the agent's private information set, such as her talents, as well as the unobserved environmental and market characteristics that affect the realized earnings . The direct financial costs associated to schooling, C_k^i, may arise in part because of observable pecuniary costs, e.g., tuition. Observable variables affecting costs can include X^i as well as additional variables Z'^i which shift costs without affecting earnings, such that $Z^i = (X^i, Z'^i)$. Let´s $Z^i\gamma_k$ denote the index that measures such costs. Other part of costs are unobserved non-pecuniary costs, U_{Ck}^i, which are related to idiosyncratic tastes and effort which can be partially known by person *i* when she makes her schooling choice. It is important to remark here that these idiosyncratic costs can be considered psychic costs and they can include expectational errors and individual attitudes toward risk, as well as specific credit constraints faced by the individual. Carneiro and Heckman (2002) examine the type of credit constraints which affect the schooling choice.

We now go on to model the decision making of the individual when she faces the schooling choice and to analyze the economic incentives which govern her choice. Naturally, the schooling choice is based upon the information that the individual has at the decision moment. Let I^i denote this information set (*ex ante* information). If all the X^i variables that determine earnings and costs are included in I^i, the estimated gross earnings gains derived from schooling considered in the decision making problem are *ex ante* gross gains such that the decision in based on them. Therefore, to analyze the schooling decision based on the proper incentives is important to correctly specify the *ex ante* information contained in X^i and Z^i. If these sets of variables include information accumulated after the schooling decision, then the estimated return is an *ex post* return which includes an element of uncertainty upon which the decision is not based (see the survey by Cunha and Heckman, 2006, to differentiate and identify *ex ante* from *ex post* returns). Economic incentives to schooling are then determined by variables (observables and unobservables) contained in the information set I^i. Once the decision is made, the realized gains will be revealed subsequently.

The schooling decision is made based on the net expected gain derived from enrolling at level *k*. In the simplest model, we ignore option values of continuing studying after the level k and we assume complete markets in which all risks are diversifiable. In this model a separation theorem applies so that the wealth maximization problem which governs the schooling choice is independent of consumption decisions, including of the consumption component that may affect the schooling choice. In this setting, the next expecting gain which governs the schooling choice can be summarized by the index function $S_k^{*i} = E\left(Y_k^i - Y_{k-1}^i - C_k^i \mid I^i\right)$, such that the observed rule of binary choice is captured by the Bernoulli random variable $S_k^i = 1$ if $S_k^{*i} > 0$, and $S_k^i = 0$. By expressing earnings and costs as functions of the observed and unobserved variables, we can write the index function as:

$$S_k^{*i} = E\left[X^i(\beta_k - \beta_{k-1}) - Z^i\gamma_k + \left(U_k^i - U_{k-1}^i - U_{Ck}^i\right) \mid I^i\right] = E\left[Z^i\mu_k - U_{Sk}^i \mid I^i\right] \qquad (1)$$

Let's assume that all elements of Z^i and U_{Sk}^i are included in the agent's information set, so that the reduced form of the schooling choice function can be represented by

$$S_k^i = 1\left(Z^i\mu_k - U_{Sk}^i \geq 0\right) \qquad (2)$$

where $1(A)$ is the indicator function which takes value 1 if A is true and zero otherwise.

In terms of both a wage equation model which includes ability, A^i, (Griliches 1977) and a policy evaluation model, our schooling decision model can be described by the following equations (assume now that earnings are given in logarithms, although this assumption implies non-linear preferences in the schooling choice function and so the consideration of risk aversion or partial insurance):

$$Y^{i} = \alpha + \delta_{k}^{i} S_{k}^{i} + \gamma A^{i} + \varepsilon^{i} \tag{3}$$

$$Y_{k}^{i} = X^{i} \beta_{k} + U_{k}^{i} \tag{4}$$

$$Y_{k-1}^{i} = X^{i} \beta_{k-1} + U_{k-1}^{i} \tag{5}$$

Since, $Y^{i} = S_{k}^{i} Y_{k}^{i} + \left(1 - S_{k}^{i}\right) Y_{k-1}^{i}$, the relationship among (3), (4) and (5) gives the following identities:

$$\delta_{k}^{i} = Y_{k}^{i} - Y_{k-1}^{i} = X^{i}\left(\beta_{k} - \beta_{k-1}\right) + \left(U_{k}^{i} - U_{k-1}^{i}\right) \tag{6}$$

$$\alpha = X^{i} \beta_{k-1} \tag{7}$$

$$\gamma A^{i} + \varepsilon^{i} = U_{k-1}^{i} \tag{8}$$

$$S_{k}^{i} = \begin{cases} 1, \; if \;\; Z^{i} \mu_{k} - U_{Sk}^{i} \geq 0 \\ \quad 0, \; \text{otherwise.} \end{cases} \tag{9}$$

Constructing the gross earnings gain for each individual, $Y_{k}^{i} - Y_{k-1}^{i}$, entails making a counterfactual comparison which faces the fundamental missing data identification problem in policy evaluation: only earnings obtained for the schooling level actually achieved are observed but not counterfactual earnings. There are different methods to solve this identification problem and to recover several moments or the joint distribution of realized and counterfactual earnings.

What is usually known as the rate of return to schooling, δ_{k}^{i}, is in general the percentage growth rate in earnings with the level k of schooling or gross earnings gain. It coincides with the rate of return when the only costs of schooling are earnings foregone and markets are perfect. Cunha and Heckman (2007) discuss the role of δ_{k}^{i} in an equation of the type of equation (3) with omitted ability as a correlated random coefficient according to Card’s (2001) model. In terms of the policy evaluation model, the correlation of δ_{k}^{i} with S_{k}^{i} is derived from the relationship between $\left(U_{k}^{i} - U_{k-1}^{i}\right)$ and U_{Sk}^{i}.

In Willis and Rosen’s (1979) model, as well as in more recent models based on policy evaluation techniques, the relationship among the unobservables U_{k}^{i} , U_{k-1}^{i} and U_{Sk}^{i} is given by a joint distribution function whose functional form can be assumed, but covariances between unobservables are always unrestricted. In this sense and to the best of our knowledge, Card (1999,

2001) is the only author who assumes "that the benefits of schooling are no higher for people with higher marginal costs". Also in this sense, Carneiro and Heckman (2002) highlight the importance of the correlation between costs and returns when considering the difference between marginal and average returns. In Card's model, this assumption implies that the covariance between the intercepts of the demand of education and of the supply of education is negative. The Beckerian model for the optimal investment in human capital interprets such intercepts as capturing individual capabilities and opportunities, respectively. Therefore, Card's assumption can be translated in our terms into a negative correlation between opportunities or financial constraints included in U^i_{Ck} and talents or capabilities included in U^i_k for every level *k*. For the correct specification of Card's assumption, we need to define how the demand of education captures capabilities or talents in our decision model, that is, how it captures the person-specific growth rate of earnings. Our assumption will be based on the definition of the following concept: "*the ability gap*".

Definition 1 *The ability gap is the difference expressed by* $\left(U^i_k - U^i_{k-1}\right)$.

This definition relies on the assumption that investments in human capital develop latent abilities, either observables or not, so that a person i can derive an increase in her returns by investing in human capital due to the development of unobservable talents. It can also happen that latent abilities depreciate when no investments in human capital are done. School quality also affects the development of latent abilities. For example, Berhman and Birdsall (1983) model returns to schooling as a function of the school quality. A part of the ability gap may be known and acted on by the individual when she makes her choice. Therefore, private information about own abilities as well as about school quality affect the quantity and quality of the schooling choice.

In the schooling choice model specified above, the reduced form expression of the schooling choice is $S^{*i}_k = Z^i\gamma_k - U^i_{Sk}$, so that we can express the unobservable heterogeneity related to the schooling decision, U^i_{Sk}, in terms of the ability gap and of nonpecuniary costs: $U^i_{Sk} = \left[U^i_{Ck} - \left(U^i{}_k - U^i{}_{k-1}\right)\right]$. In latent index models, U^i_{Sk} may depend on U^i_k and U^i_{k-1} in a general way. We show here that we can restrict this relationship among unobservables by using the above mentioned Card's assumption and our concept of ability gap. As a result, we show that the interpretation of the original Roy's model (with $U^i_{Ck} = 0$) in terms of comparative advantage and hierarchical sorting can be applied to this generalized Roy model.

Card's assumption can be expressed in our terms as:

Assumption 1. $Cov\left(U_{Ck}^{i}, U_{k}^{i}-U_{k-1}^{i}\right) \leq 0$, *that is, the development of latent abilities throughout the investment in human capital is negatively related to the unobservable costs of schooling.*

Note that this assumption does not impose any sign on the covariances between U_{Ck}^{i}, U_{k}^{i} and U_{k-1}^{i}, which, in any case, are not empirically observable. And yet, our assumption allows us to restrict the empirically observable relationship between U_{Sk}^{i} and the ability gap as we proof in the following result.

Proposition 1. $Cov\left(U_{Sk}^{i}, U_{k}^{i}-U_{k-1}^{i}\right) \leq 0$, *that is, the development of latent abilities throughout the investment in human capital is negatively related to the unobservable net discounted costs and, for that matter, it is positively related to the probability of going to school.*

Proof:

$$Cov\left(U_{Sk}^{i}, U_{k}^{i}-U_{k-1}^{i}\right)=Cov\left(U_{Ck}^{i}-\left(U_{k}^{i}-U_{k-1}^{i}\right), U_{k}^{i}-U_{k-1}^{i}\right)=Cov\left(U_{Ck}^{i}, U_{k}^{i}-U_{k-1}^{i}\right)-Var\left(U_{k}^{i}-U_{k-1}^{i}\right)$$

From this expression, it is immediate to show that under assumption 1 the whole expression of the covariance is negative. In the case of homogeneous returns $\left(U_{k}^{i}=U_{k-1}^{i}\right)$ the covariance is zero and the mean growth rate of earnings with schooling can be estimated by least squares. Furthermore, the higher the variance of the ability gap, and the higher the benefits obtained for people with lower marginal costs, the higher will be the absolute value of this covariance. We expect to observe the highest absolute values of this covariance for the highest levels of education and the highest quality of schooling.

Introducing some notation, we express Proposition 1 in terms of a difference between covariances:

$$Cov\left(U_{Sk}^{i}, U_{k}^{i}-U_{k-1}^{i}\right)=Cov\left(U_{Sk}^{i}, U_{k}^{i}\right)-Cov\left(U_{Sk}^{i}, U_{k-1}^{i}\right)=\rho_{k,S}-\rho_{k-1,S}.$$

3. THE ABILITY GAP AND THE DISTRIBUTION OF RETURNS TO SCHOOLING

In Becker's original model as well as in Card's, the demand of education is the decreasing function which represents gross marginal returns with respect to the amount invested and the supply of education is the increasing function which represents marginal costs with respect to the amount invested. In equilibrium, the equality of marginal costs with marginal returns implies that the level of enrollment (usually known as the demand of education) is an increasing function with respect to the net idiosyncratic gain which results from the difference between capabilities and opportunities or tastes. In our terms, this idiosyncratic gain is the term $E\left(U_k^i - U_{k-1}^i - U_{Ck}^i \mid I^i\right)$, that is, $E\left(-U_{Sk}^i \mid I^i\right)$. Following Becker's model, our policy evaluation model should predict a distribution of the amounts invested in education across the population that is decreasing with respect to the random variable U_{Sk}^i. For this optimal or chosen distribution of enrollment, Card's model predicts an increasing distribution of marginal returns to schooling with respect to both capabilities (that is, with respect to the ability gap) and opportunities (that is, with respect to U_{Ck}^i). However, we do not observe the distribution of either capabilities or opportunities. What we observe is the distribution of schooling choices across the population which parallels the distribution of U_{Sk}^i. That is, the distribution of U_{Sk}^i informs us of the probability of going to school or, for that matter, of the level of enrollment at a particular level. So far, in Becker's model as expressed by Card nothing can be said about the relationship between marginal returns and U_{Sk}^i. In contrast, our model has several empirical implications on the sign of the differences of average returns between different population groups and on the relationship between marginal returns to schooling and the values taken by the idiosyncratic gain as expressed by the random variable U_{Sk}^i.

Next, we derive the empirical implications of our model in two different settings. First, when the schooling choice is binary. Second, when the schooling choice is sequential, i.e., subsequent levels of education are chosen at different moments in time.

3.1 BINARY CHOICES

Considering an independent binary choice (k=1) in a parametric control function approach and using results from the mean of truncated distributions, we can compare the expected earnings for those who choose a particular sector with the expected earnings of a person chosen at random from the population of the sector.

Person *i*, who made ex ante her decision before entering the level of education 1, obtains ex post the following person-specific wage premiums with respect to the average person in each sector (*1* and *0*, respectively*)* if she chose to enter the respective sectors :

$$E\left(U_1^i \mid S=1\right) = -\rho_{1,S}\,\frac{f\left(\bar{\gamma}\right)}{F\left(\bar{\gamma}\right)} \quad \text{and} \quad E\left(U_0^i \mid S=0\right) = \rho_{0,S}\,\frac{f\left(\bar{\gamma}\right)}{1-F\left(\bar{\gamma}\right)} \tag{10}$$

where $f(.)$ and $F(.)$ are the density and cumulative distribution functions of U_S^i, respectively, and $\bar{\gamma}$ denote the mean of the index $Z^i\gamma$ in the index function S^{*i}.

Comparative advantage implies that earnings of people who choose each sector are greater than those of the average person in each sector. That is, $E\left(U_1^i \mid S=1\right)>0$ and $E\left(U_0^i \mid S=0\right)>0$. On the other hand, hierarchical sorting implies that those choosing the level *1* perform better in both groups that the average person of each group, but they are better at level *1* than at level *0*, that is, $E\left(U_1^i \mid S=1\right)>0$, $E\left(U_0^i \mid S=1\right)>0$ and $E\left(U_1^i-U_0^i \mid S=1\right)>0$. For that matter, comparative advantage implies that $\rho_{1,S}<0$ and $\rho_{0,S}>0$. Hierarchical sorting holds when $\rho_{1,S}<0$ and $\rho_{0,S}<0$.

Hence, Proposition 1, which states that $\left(\rho_{1,S}-\rho_{0,S}\right)<0$, is compatible with both comparative advantage and hierarchical sorting, including with inverse hierarchical sorting ($\rho_{1,S}>0$ and $\rho_{0,S}>0$) in which the less educated perform better than average in both sectors but better in sector *0* than in sector *1*. However, "comparative disadvantage" in which persons who choose each group perform worse than the average person of each group is not allowed.

This result has several implications on the typical treatment parameters of policy evaluation, at least in a parametric framework. Consider the following definitions:

The average return (ex post return based upon an ex ante decision) for a person randomly drawn from the population (Average Treatment Effect: ATE) is:

$$E\left(Y_1^i-Y_0^i \mid X\right)=X\left(\beta_1-\beta_0\right) \tag{11}$$

The average return for those who achieved the upper level (Treatment on the Treated: TT) is:

$$E\left(Y_1^i-Y_0^i \mid X, S=1\right)=X\left(\beta_1-\beta_0\right)+E\left(U_1^i-U_0^i \mid S=1\right)=X\left(\beta_1-\beta_0\right)-\left(\rho_{1,S}-\rho_{0,S}\right)\frac{f\left(\bar{\gamma}\right)}{F\left(\bar{\gamma}\right)} \tag{12}$$

The average return for those who left education at the lower level (Treatment on the Untreated: TUT) is:

$$E\left(Y_1^i-Y_0^i \mid X, S=0\right)=X\left(\beta_1-\beta_0\right)+E\left(U_1^i-U_0^i \mid S_k=0\right)=X\left(\beta_1-\beta_0\right)+\left(\rho_{1,S}-\rho_{0,S}\right)\frac{f\left(\bar{\gamma}\right)}{\left(1-F\left(\bar{\gamma}\right)\right)} \tag{13}$$

The OLS bias or selection bias which can be measured as the difference between the OLS and TT parameters (OLS-TT) is:

$$E\left(U_0 \mid S=1\right)-E\left(U_0 \mid S=0\right)=-\rho_{0,S}\left(\frac{f\left(\bar{\gamma}\right)}{F\left(\bar{\gamma}\right)}+\frac{f\left(\bar{\gamma}\right)}{\left(1-F\left(\bar{\gamma}\right)\right)}\right) \tag{14}$$

Another interesting parameter which has not been studied in the literature is the difference between the OLS parameter and the TUT parameter (OLS-TUT) which results in:

$$E(U_1 \mid S=1) - E(U_1 \mid S=0) = -\rho_{1,S}\left(\frac{f(\bar{\gamma})}{F(\bar{\gamma})} + \frac{f(\bar{\gamma})}{(1-F(\bar{\gamma}))}\right) \tag{15}$$

We designate this parameter by the name of "college premium bias", considering the upper level of education as college education.

Therefore, under Proposition 1 and for any distribution of the unobservables, TT>ATE>TUT. That is, the sorting gain for upper-achievers (TT-ATE) is always positive and the sorting gain for underachievers (TUT-ATE) is always negative. This is another way of interpreting comparative advantage that has been done in most of empirical work based on policy evaluation models. However, we have seen that this result, guaranteed by Assumption 1, is also compatible with hierarchical sorting of any type. For the OLS selection bias, it is negative either under comparative advantage or inverse hierarchical sorting, and positive under (normal) hierarchical sorting. The last parameter, "college premium bias", is positive under both comparative advantage and (normal) hierarchical sorting.

Since correct decisions in which there is no scope for ex post regret have to be based on positive sorting gains for each sector[2], the accomplishment of Assumption 1 guarantees such correct decisions without the necessity of having comparative advantage. The decrease of psychic costs for those with larger ability for capturing the person-specific gain in earnings is a sufficient condition that guarantees a correct decision under full information.

Carneiro, Heckman and Vytlacil (2005) estimate the above measures of average returns by instrumental variables methods. They point out the problems of identification of these average returns or treatment effects since different instruments identify different parameters and, as a result, these measures do not answer well-posed economic questions. In the attempt to construct policy relevant parameters, these authors show that every average return measure can be constructed as a weighting average of marginal returns, being these marginal returns a measure of the willingness to pay for schooling for people at the margin of indifference between two schooling levels.

---

[2] That happens when the person-specific gain in the chosen sector is larger than its contrafactual. The choice of sector 1 should be based on $E(U_1^i - U_0^i \mid S=1) > 0$ and the choice of sector 0 is correct when $E(U_1^i - U_0^i \mid S=0) < 0$.

In policy evaluation models, the marginal returns to education are measured by the Marginal Treatment Effect, MTE, is defined as the mean gain to schooling for individuals of given characteristics, $X^i = x$ and a level of unobservables $U_S^i = u_S$. Once the distribution of unobservables is normalized, MTE measures the average return for a given level of the probability of continuing studying. Then, given the model of education choice represented by (1) and (2), MTE measures the average return at the margin of indifference. The MTE corresponds to the following definition:

$$MTE^i(x,u_S) = E\left(Y_1^i - Y_0^i \mid X^i = x, U^i{}_S = u_S\right) = \left[x(\beta_1 - \beta_0)\right] + E\left(U_1^i - U_0^i \mid U^i{}_S = u_S\right) \qquad (16)$$

The term $E\left(U_1^i - U_0^i \mid U_S^i = u_S\right)$ is the regression function of the ability gap on U_S^i. It is commonly assumed that this function is linear: $E\left(U_1^i - U_0^i \mid U^i{}_S = u_S\right) = \eta\, u_S$ (Kagan, Linnik and Rao 1973 show that under a variety of joint densities of (X, Y), and in particular the bivariate normal distribution function, the conditional expectation $E(Y \mid X)$ yields a linear function). Given that $\eta = Cov\left(U_1^i - U_0^i, U_S^i\right)/Var\left(U_S^i\right)$, Proposition 1 implies that MTE is a decreasing function in u_S.

Without loss of generality, U_S^i is assumed to be distributed as a Uniform[0,1] so that it is possible to reparameterize the model for any random variable V^i with $p(z) = \Pr\left(V^i < \bar{\gamma} \mid Z = z\right) = F_V(\bar{\gamma})$, $U_S^i = F_V\left(V^i\right)$, $u_S = \Pr(V < v)$ and $v = F_V^{-1}(u_S)$. Therefore, MTE is decreasing in both u_S and v. Commonly, the original random variable V^i is a standard normal; then, MTE is a decreasing linear function with respect to v and a decreasing non-linear function with respect to u_S. For particular joint densities or in a nonparametric framework, however, MTE can be a nonlinear function with respect to both u_S and v. Since the schooling selection rule can also be expressed as the indicator function $S = 1\left[P(Z) - U_S^i \geq 0\right]$, it is possible to express the MTE as a function of the propensity score $p(z)$ so that, in this case, it is an increasing function under Proposition 1[3].

---

[3] In latent index models it is assumed without loss of generality that U_S is uniformly distributed between 0 and 1. If the instrument Z' is externally set such that $p(z) = u_S$, where $p(z) = \Pr(S = 1 \mid Z = z)$ is the propensity score, the parameter MTE can be expressed as a function of $p(z)$. It is assumed that $E(Y_1 - Y_0 \mid Z) = E(Y_1 - Y_0 \mid \Pr(S = 1 \mid Z))$ according to the index sufficiency restriction assumption frequently used in the selection models' literature.

In a parametric framework, the distribution of $\left(U_1^i - U_0^i\right)$ given U_S^i can be degenerate so that we do not observe variability of marginal returns for a given value of U_S^i. However, from the estimation of η in a simple linear model without intercept, we know that $Var(\hat{\eta}) = \left[Var\left(U_1^i - U_0^i\right) \Big/ Var\left(U_S^i\right) \right]$. To identify parametric models, $Var\left(U_S^i\right)$ is assumed to take a given value, for example 1. Therefore, the variance of estimated marginal returns across the population equals (or is proportional to) the variance of the ability gap (taken fixed values of observable characteristics).

This result implies that, if education acts as a catalysts of latent abilities, that is, if investments in education have a large effect on the development of latent abilities so that they develop more the abilities of the best endowed, the variability of marginal returns will be higher the higher the variability in innate capabilities. Conversely, in a “pure” human capital view, if education only teaches general observable skills but does not develop latent abilities we should observe the same variability of marginal returns across the population than the variability of innate abilities. Besides, assuming again that education does not act as a catalyst of abilities and if the market only pays innate abilities but not educational investments, as in a pure signalling model, we should observe a decrease in the variability of returns to education with the level of schooling since there is a progressive homogeneity of abilities for those who chose higher levels of education through a screening process. Thus, the old quest for an empirical distinction between the human capital and signalling theories can be solved in terms of the observed variability of returns to schooling for subsequent schooling levels.

3.2 SEQUENTIAL CHOICES

The framework can be generalized to a sequential model where the individual chooses different levels of schooling at different moments in time (see for example, Zamarro 2007). This generalization is important if option values of continuing studying have an important weight in returns to schooling. As well, the sequential framework allows drawing conclusions to the whole population, for if we only consider he binary choice between high school and college we are taking a much selected part of the population than if we also include previous levels. In a sequential schooling model, Heckman and Navarro (2006) show that estimated option values for attending college are relatively small (at most 2% of the total return for U.S. data).

A sequential setting with three levels of education can be parametrically expressed by a joint distribution function of the form

$$\begin{pmatrix} U_k \\ U_{S1} \\ U_{S2} \end{pmatrix} \sim F\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_k{}^2 & \rho_{k,S1} & \rho_{k,S2} \\ \rho_{k,S1} & 1 & 0 \\ \rho_{k,S2} & 0 & 1 \end{pmatrix} \right) \quad k = 0,1,2$$

where $S_1 = 1(\overline{\gamma_1} - U_{S1}) = 1$ if the individual left the school at level 1, and similarly for S_2.

Assumption 1 applied to both levels entails:

Assumption 2: $Cov(U^i_{C1}, U^i_1 - U^i_0) \leq 0$ and $Cov(U^i_{C2}, U^i_2 - U^i_1) \leq 0$

In these terms, Proposition 1 applied in this sequential setting implies

Proposition 2: $(\rho_{1,S1} - \rho_{0,S1}) < 0$ and $(\rho_{2,S2} - \rho_{1,S2}) < 0$

In a sequential setting as the one presented in Zamarro (2007) for *k*=0,1,2, if we consider for example the expected earnings of those who chose to complete the level *k*=2, these are:

$$E(Y^i_2 \mid S_1 = 0, S_2 = 1) = X\beta_2 + E(U^i_2 \mid S_1 = 0, S_2 = 1) \tag{17}$$

To avoid truncation in the empty set given that S_1 and S_2 always take opposite values, it is necessary to define the variable $\tilde{S} = S_1 + S_2$ which is the indicator function for the choice of continuing education after the level 0, $\tilde{S} = 1(\overline{\gamma^*} - U_{\tilde{S}})$, where $\overline{\gamma^*} = S_1\overline{\gamma_1} + S_2\overline{\gamma_2}$ and $U_{\tilde{S}} = S_1 U_{S1} + S_2 U_{S2}$.

This way, the above expression (17) gives:

$$E(Y^i_2 \mid \tilde{S} = 1, S_2 = 1) = X\beta_2 + E(U_2 \mid \tilde{S} = 1, S_2 = 1) = X\beta_2 - \left[\rho_{2,\tilde{S}} \frac{f(\overline{\gamma^*})}{F(\overline{\gamma^*})} + \rho_{2,S_2} \frac{f(\overline{\gamma_2})}{F(\overline{\gamma_2})} \right] \tag{18}$$

We see in expression (18) that, because of the dependence between subsequent choices, self-selection in terms of comparative advantage or hierarchical sorting of those who chose to achieve the level of education *k* cannot be assessed by the covariance ρ_{2,S_2} solely. Their idiosyncratic gain or loss is also related to the choice they did at a first moment (they decided to continue studying after the level *0*).

Therefore, Proposition 2 is not testable in its terms but it has to be expressed in terms of the random variable $\tilde{S}$. To find the testable result, we define the unobservable costs that will be faced by a person who overcomes the level 0 by $U_{\tilde{C}} = S_1 U_{C1} + S_2 U_{C2}$.

The expressions of the covariance between these costs and the ability gap for each level of education gives:

$$Cov\left(U_{\tilde{C}},U_1-U_0\right)=Cov\left(S_1U_{C1},U_1-U_0\right)+Cov\left(S_2U_{C2},U_1-U_0\right) \tag{19}$$

$$Cov\left(U_{\tilde{C}},U_2-U_1\right)=Cov\left(S_1U_{C1},U_2-U_1\right)+Cov\left(S_2U_{C2},U_2-U_1\right) \tag{20}$$

$$Cov\left(U_{\tilde{C}},U_2-U_0\right)=Cov\left(S_1U_{C1},U_2-U_0\right)+Cov\left(S_2U_{C2},U_2-U_0\right) \tag{21}$$

We cannot assume any sign on expressions (19) and (20) without assuming that a person with lower costs in an education level has higher returns in other education level. However, we should expect a negative sign for expression (21) if a person with higher benefits from the entire educational investment faces lower costs on both levels of education. In these terms, we derive the following result:

Assumption 3: $Cov\left(U^i_{\tilde{C}},U^i_2-U^i_0\right)\leq 0$

Assumption 4: $Cov\left(U^i_1-U^i_0,U^i_2-U^i_0\right)\geq 0$

Assumption 5: $Cov\left(U^i_2-U^i_1,U^i_2-U^i_0\right)\geq 0$

Proposition 3: If Assumptions 3, 4 and 5 hold, then $Cov\left(U^i_{\tilde{S}},U^i_2-U^i_0\right)\leq 0$.

Proof: It can be shown by developing the expression:

$$\begin{aligned}
&Cov\left(U^i_{\tilde{S}},U^i_2-U^i_0\right)=Cov\left(U^i_{\tilde{C}}-\left(S_1\left(U^i_1-U^i_0\right)+S_2\left(U^i_2-U^i_1\right)\right),U^i_2-U^i_0\right)=\\
&Cov\left(U^i_{\tilde{C}},U^i_2-U^i_0\right)-Cov\left(S_1\left(U^i_1-U^i_0\right),U^i_2-U^i_0\right)-Cov\left(S_2\left(U^i_2-U^i_1\right),U^i_2-U^i_0\right)=\\
&=Cov\left(U^i_{\tilde{C}},U^i_2-U^i_0\right)-Cov\left(U^i_1-U^i_0,U^i_2-U^i_0\right)\Pr(S_1=1)-Cov\left(U^i_2-U^i_1,U^i_2-U^i_0\right)\Pr(S_2=1)
\end{aligned}$$

Proposition 3 is testable and its meaning is as follows. Assuming that (i) the development of latent abilities through schooling is negatively related to all unobservable costs incurred to obtain the corresponding schooling level and (ii) the total development of latent abilities through schooling is positively related to the development in each level, then we will observe a negative relationship between unobservable returns and the net unobservable costs of the early schooling decision ($U^i_{\tilde{S}}$), or a positive relationship between unobservable returns and the probability of continuing studying after the low level . Furthermore, this relationship will be stronger for large correlations between the ability gaps obtained at each level of education. Consequently, if the effect of education on the development of latent abilities is accumulative, these correlations are larger the larger the catalyst effect of education and the larger the variability of talents are.

Just as in the binary case, we can define different sorting gains for those who chose a particular schooling level as well as different selection bias (see Zamarro, 2007).

$$E(U_1 - U_0 \mid S_1 = 1) = -\left(\rho_{1,\tilde{S}} - \rho_{0,\tilde{S}}\right)\frac{f\left(\overline{\gamma^*}\right)}{F\left(\overline{\gamma^*}\right)} + \left(\rho_{1,S2} - \rho_{0,S2}\right)\frac{f\left(\overline{\gamma_2}\right)}{\left(1 - F\left(\overline{\gamma_2}\right)\right)} \tag{22}$$

$$E(U_2 - U_0 \mid S_2 = 1) = -\left(\rho_{2,\tilde{S}} - \rho_{0,\tilde{S}}\right)\frac{f\left(\overline{\gamma^*}\right)}{F\left(\overline{\gamma^*}\right)} - \left(\rho_{2,S2} - \rho_{0,S2}\right)\frac{f\left(\overline{\gamma_2}\right)}{F\left(\overline{\gamma_2}\right)} \tag{23}$$

where the terms in parenthesis are identifiable in a parametric framework. Zamarro (2007) presents parametric estimations for them that we discuss in the next section.

Zamarro also obtains in a semiparametric framework the distributions of the MTE for each schooling level which helps us to analyze the variability of marginal returns in addition to the sorting gains.

4. EMPIRICAL EVIDENCE FOR THE U.S. AND SPAIN

Empirical evidence on the testable implications of Proposition 1 and 3 allows us to test the associated assumptions. This test is done to check whether abler people face lower marginal costs so that they have the correct incentives in their schooling choice. It can happen that the burden of costs on returns is very high due to different reasons so that people do not have the correct incentives to choose the level and type of schooling from a wealth maximization point of view. Among the reasons impeding people to make their choice we can consider the three following: (i) the market dos not remunerate person-specific skills, that is, there is not skill premium so that any kind of psychic costs of schooling entails a big burden, (ii) people face specific credit constraints of any type (short terms or long terms constraints), and (iii) expectational errors and risk attitudes associated with uncertainty costs amount for a large part of the person-specific gains.

The empirical properties derived from Proposition 1 and 3 are necessary conditions of the assumptions under which such properties are obtained, whose meaning is basically that abler people face lower marginal costs. Obviously, if we do not find evidence against our results this does not guarantee that the underlying assumptions are true. On the contrary, if empirical evidence rejects our predictions we should reject our assumptions.

We obtain empirical evidence for U.S. data on the predictions done for the binary choice case in Carneiro, Heckman and Vytlacil (2005) whose discussion will be complemented and compared with empirical results found in several factor models by Cunha, Heckman and Navarro (2005), Cunha and Heckman (2006), and Carneiro and Heckman (2002). For Spanish data, we use the work by Zamarro (2007) which has set the framework of our sequential analysis. Some alternative evidence from more traditional studies about earnings inequality (Simón, 2007) and returns to education (López-Bazo and Moreno, 2007) can complement Zamarro's evidence for the Spanish case.

4.1 BINARY CHOICES IN THE U.S.

Carneiro, Heckman and Vytlacil (2005) obtain their estimations from a sample drawn from the U.S. National Longitudinal Survey of Youth 1979 (NLSY79). The sample consists of white males, in 1995, with either a high school degree (611 individuals for level 0) or above (715 individuals for level 1). Wages are reported in 1994 and 1996. First, they show estimates based on a simulation of a Generalized Roy model under a trivariate normal distribution of (U_0, U_1, U_S). Simultaneously, they estimate a logit model for the schooling choice and a semiparametric model for the wage equation. Based upon these estimates, they obtain estimates of the MTE parameter over the support of the propensity score. They obtain estimates for several parameters measuring average returns to one year college by integrating the weighted MTE over different supports and with different weights. In particular, they estimate those parameters which are relevant to test our predictions (ATE, TT, TUT, OLS).

Wage equations are conditioned on experience, schooling-adjusted ability test scores (AFQT) and local unemployment rates. Family background variables are also included in an alternative model. All of these X variables are included in the schooling choice model as well as variables affecting schooling costs (tuition and distance to college). In this setting, we can assume that all observable X variables are known ex ante so that the estimated MTE by instrumental variables are estimates of ex ante gross gains and, therefore, these gains are taken into account at the moment of the schooling choice.

In the simulated Roy model, the sorting gain of graduated is about a 4% (TT-ATE=0.042) and the sorting gain of non-graduated is a 4% loss (TUT-ATE=-0.0433). Therefore, both sorting gains are fully compatible with Proposition 1. The selection bias (OLS-TT) is negative indicating that $\rho_{0,S}$ is positive. The estimate for the parameter (OLS-TUT) gives a slightly positive value (0.0165) which implies that $\rho_{1,S}$ is negative. Then, both the OLS bias and "the college premium bias" confirm the hypothesis of comparative advantage.

Several estimates of the sorting gains and selection bias in the semiparametric model give values around the following ranges:

Sorting gain for colleges: TT-ATE= 6 to10%

Sorting gain for high school: TUT-ATE= -7 to -10%

OLS selection bias: OLS-TT= -15 to -23%

"College premium bias": OLS-TUT= -1 to -2 %

As it happens in the simulated Roy model, the semiparametric estimates of the sorting gains are consistent with Proposition 1. But the "college premium bias" becomes slightly negative. This result joint with the negative OLS selection bias result, does not disallow the possibility of existence of some inverse hierarchical sorting, at least for a small part of the sample such that some college graduates are performing worse than the average person in both sectors, although it could be possible that their worse performance in the counterfactual sector 0 incentives them to choose the college sector if they can finance it. The following discussion on the semiparametrical estimates of the shape of the MTE for those with the lowest probability of choosing the college sector helps explain this result in terms of credit constraints.

With respect to the shape of the MTE, we observe that MTE is a decreasing function of u_s in the simulated Generalized Roy model. This particular nonlinear shape of the function gives a decreasing linear function of MTE with respect to v where $v = \Phi^{-1}(u_s)$ and Φ is the cumulative normal distribution function. This decreasing pattern is fully consistent with Proposition 1.

Figure 5 of the Appendix (extracted from Carneiro, Heckman and Vytlacil, 2005) presents semiparametric estimates of the MTE function as a decreasing function with respect to u_s, although the function increases slightly for the highest values of u_s. The increasing pattern of the MTE for those with the lowest probabilities of attending college appears to be contradictory with Proposition 1, case in which the choice can be informed in terms of comparative disadvantage. In this sense, those college graduates with the lowest probabilities of attending college are making the wrong decision by choosing the college sector. As the TUT parameter is a weighted average of the MTE parameter with larger weights for those with lower probabilities of attending college, this increasing pattern in the MTE could explain the relatively high value of the TUT parameter with respect to the OLS parameter and consequently the slightly negative value of the parameter we have called "college premium bias".

Further evidence explaining this fact is shown by Carneiro and Heckman (2002). They find that around an 8% of the U.S. population faces short-run liquidity constraints to finance college education so that some people able enough to have a sorting gain from college cannot have it. On the contrary, there may be a part of low able and rich people that choose college education. In both cases, the person-specific part of marginal costs of schooling, as reflected by psychic costs or personal credit constraints which impede these people either to finance college education or to inherit the genes and favorable parental environment, produces a positive correlation between these person-specific costs and the person-specific gain captured by the ability gap and, consequently, the hypothesis made in Assumption 1 breaks down.

The semiparametric estimation allows us to observe the variability of marginal returns for different levels of the propensity score. We observe that this variability is larger for extreme values of the propensity score. That is, those with the highest and the lowest probabilities of attending school face larger variability in returns. These individuals have larger differences between capabilities and non-pecuniary costs. The higher degree of noise at the extremes of the propensity score could be caused by a weaker identification at the extremes so that the cell of data is thinner. Although this is true, the larger variability of the MTE observed at the extremes of high probability (low u_s) is consistent with the expectation that the ability gap and its variability are larger for these groups under the hypothesis that education is acting as a catalyst. At the same time, the larger variability of the MTE at the extreme of low probability is consistent with the hypothesis that people facing credit constraints or any other impediment against their right choice present a high variability in their abilities.

Complementary evidence on the economic incentives to enrollment in college education in the U.S. is provided by studies which model and estimate the differences between heterogeneity (private information known by the agent when she makes the schooling choice) and uncertainty or `luck' which is revealed after the decision is made and affect ex post returns. In this sense, the works by Cunha, Heckman and Navarro (2005) and Cunha and Heckman (2006) find that agents have private information ex ante on two types of factors. The first one is related to their latent abilities and, although this factor does not allow agents to reduce uncertainty about their future earnings, this ability factor gives them the economic incentives relevant to the choice of the level of education. This evidence is consistent with the empirical findings by Carneiro, Heckman and Vytlacil (2005) discussed above. The second factor allows individuals to reduce their uncertainty about future earnings in a substantial amount, so that we interpret that this factor is capturing information about the labor market and the market environment in general. This kind of private knowledge about the market, however, does not help individuals to make their choice between levels of education. As Cunha and Heckman (2006) sum up, for a variety of market environments and assumptions about preferences, a robust empirical regularity is that over 50% of the ex post variance in the returns to schooling are forecastable at the time students make their college choices. So, uncertainty leaves a wide margin for ex post regret, as well as we saw before with credit constraints, with the difference that credit constrains operate ex ante and uncertainty does not.

4.1 SEQUENTIAL CHOICES IN SPAIN

Zamarro (2007) presents estimates of MTE from a parametric and a semiparametric framework by extending into a sequential framework the estimation method developed by Carneiro, Heckman and Vytlacil (2005) for binary choices. She obtains the estimates from a cross-sectional Spanish data set for the year 1991 (*"Encuesta de Conciencia y Biografia de Clase"*) which allows her to use different instruments for each decision stage to identify the model. She uses data from three levels of education: 1646 observations for the low level, 1185 observation for the high school level, and 414 observations for the college level.

Wage equations presented by Zamarro are conditioned on several X variables, some of them which we can assume known ex ante but some others are revealed ex post. Among the ex ante variables, there are geographical, age and gender variables. Among the ex post variables, there are industry, occupational, and tenure dummies. In the schooling choice, some of the instruments of the wage equation enter the set of Z variables as well as ex ante X variables: some variables describing family background by parents' education and job qualification, family income and unemployment rate of the province of residence at different ages, and distance to college as indicator of costs of schooling. Therefore, the estimates of MTE presented by Zamarro cannot be considered estimates of ex ante gross gains which are the relevant ones to inform the schooling choice. However, under market conditions similar to those prevailing in the studies realized for the U.S. by Cunha, Heckman and Navarro (2005), the uncertainty captured by the ex post gross gains estimated by Zamarro should be mainly related with a factor associated with mid-career wage developments (for example, tenure variable in Zamarro's case) whose distribution overlaps for different schooling levels, so that it would have not affected the schooling choice although it were known ex ante, and which affects positively returns. Therefore, we propose that the analysis of Zamarro estimates of MTE can be done in the same terms than the analysis done for the above estimates corresponding to the U.S. since uncertain ex post returns captured by Zamarro´s estimates are not affecting the ex ante choice.

The paper by Zamarro provides parametric estimates on several of covariances between the ability gap and the schooling decision heterogeneity at various levels of education, which we have specified in subsection 3.2. Specifically, she presents the following estimates from a parametric model under normality:

$$(\rho_{1,\tilde{S}} - \rho_{0,\tilde{S}}) = 0.32*$$
$$(\rho_{1,S2} - \rho_{0,S2}) = -0.55*$$
$$(\rho_{2,\tilde{S}} - \rho_{0,\tilde{S}}) = -0.053$$
$$(\rho_{2,S2} - \rho_{0,S2}) = 0.41$$

where the estimates not marked by the star are not statistically different from zero.

According to the parametric expressions of the sorting gains (22) and (23), these estimates indicate that the sorting gain for those who achieve the high level of education with respect to the low level, $E(U_1 - U_0 \mid S_1 = 1)$, is negative, whereas the overall sorting gain for college graduates with respect to the lowest level, $E(U_2 - U_0 \mid S_2 = 1)$, can be considered zero. Therefore, neither high school students nor college graduates are making a well informed choice based on their average sorting gains.

Proposition 2, which predicts that $(\rho_{2,\tilde{S}} - \rho_{0,\tilde{S}})$ should be non-positive, is not rejected by empirical evidence, although we cannot reject the hypothesis that is equal to zero. In this case, we could assume homogeneous returns for each level and independence among person-specific costs of schooling and person-specific gains from schooling for each level. So we do not find evidence supporting the hypothesis of the accumulative learning effect of education as well as of that abler people face lower costs of the entire schooling period.

Furthermore, the positive sign found for $Cov(U^i_{\tilde{S}}, U^i_1 - U^i_0)$ is not jointly consistent with assumptions (i) $Cov(U_{\tilde{C}}, U_1 - U_0) \leq 0$ and (ii) $Cov(U^i_2 - U^i_1, U^i_1 - U^i_0) \geq 0$. Therefore, we find evidence rejecting either the accumulative learning process of the schooling system or the fact that abler people in early stages faces lower costs of the overall schooling system.

On the contrary, the negative sign found for $Cov(U^i_{S2}, U^i_1 - U^i_0)$ is jointly compatible with assumptions (i) $Cov(U_{C2}, U_1 - U_0) \leq 0$ and (ii) $Cov(U^i_2 - U^i_1, U^i_1 - U^i_0) \geq 0$. Therefore, this evidence supports the assumption that abler people at early stages faces lower costs in subsequent stages of schooling as well as the assumption of the accumulative learning process of the schooling system.

On the whole, different cues of evidence based on the parametric model under normality appear to be contradictory and show the difficulty in the definition of the psychic costs that agents expect when deciding whether to continue studying after the low level. Moreover, Zamarro finds departures from normality, which could be causing in part some of the apparent contradictions, and presents semiparametric estimates of the distributions of marginal returns for the two subsequent binary choices which help explain the puzzle of parametric results.

Zamarro's estimates of the distribution of the MTE resulting from the two subsequent decisions are presented in Appendix Figures 8 and 9. At stage 0, MTEo estimates the gross earnings gain for those indifferent between leaving school at the low level and continuing studying. At stage 1, MTE_1 estimates the gross earnings gain for those indifferent between leaving school at the medium level (high school) and continuing the college level. First at all, Zamarro find negative marginal returns for those with low probability of continuing after the low level, so that these people experienced a loss when obliged by the 1970's reform to continue the low level up to the age of 14. For those with higher probabilities of continuing to high school we appreciate an increasing pattern in the distribution of marginal returns, although with more noise.

More striking is the fact that many people over all the support of the estimated probability of continuing studying after the medium level experience negative marginal returns. The pattern of the distribution of MTE_1 shows a flat pattern around zero with larger variance of returns for those with the lowest probabilities of going to college. Identification of the propensity score for values up to 0.4 is weak so that the cell of data becomes very thin.

The low values of marginal returns are consistent with the finding of negative and zero sorting gains for high school and college, respectively. Although, we have not derived concrete empirical properties on the shape and variance that MTE_0 and MTE_1 should present under different assumptions, their flat shape combined with negative gains give a hint on the difficulty that Spanish youngsters face to make their schooling choice based on earnings expectations. Indeed, the economic incentives which prompt enrollment in secondary and university education incentives do not operate clearly at the individual wealth maximization level as it happened in the U.S. It seems that, in Spain, the correlation between the psychic costs and the person-specific earnings gains, which should govern the decision of whether to continue schooling after the low or the medium level, is not negative.

This hint of a positive correlation between marginal costs and marginal returns is indicating a high burden of costs of schooling in terms of returns. We cannot disentangle which are the causes of this big burden among the three reasons we pointed out at the beginning of the section. However, given the characteristics of gratuity and ample geographical supply of the Spanish education system at secondary and tertiary levels, we can reject the hypothesis of the existence of short-term financial constraints. At the same time, we do not guess any reasons which lead us to assume the existence of long-term credit constrains or environmental constraints to the acquisition of cognitive and non-cognitive skills fostered by the parental environment in a different way that such acquisition operates in the U.S. This leaves us room to hypothesize in favor of the role of both uncertainty and a market environment which does not remunerate person-specific skills as the main factors explaining the big burden of psychic costs of schooling.

Some evidence from studies about earnings inequality (Simón, 2007) and returns to human capital (López-Bazo and Moreno, 2007) in Spain complements Zamarro's findings in this sense of the small role of the skill premium in the Spanish economy. The evidence shows a decrease in the variability or earnings inequality during the last years of expansion of college education as well as a decrease in the returns to education from both a private returns and an aggregate return to human capital point of views. Among the reasons that explain the decrease in the skill premium in Spain, some authors point out the phenomenon of overeducation experienced in Spain during the last decades (see, for example, Alba-Ramirez, 1993, and Hartog, 2000).

5. CONCLUSIONS

Our model of economic incentives in education is based on our regarding education as an investment whose purpose is to maximize lifetime wealth. We approach this Human Capital perspective from a policy evaluation model. This approach allows us to capture all the possible person-specific effects of education into a random variable that we denote by "ability gap". Policy evaluation models of the random-coefficient type explain the correlation between the amount of schooling and the returns to schooling by means of the correlation of this random term, the "ability gap", and the individual heterogeneity in the schooling choice. This correlation is the source of the heterogeneity of returns to schooling across the population. Our step forward consists of modeling in terms of the concept of "ability gap" an assumption already made (albeit in different terms) in the human capital model of Card (1999, 2001): namely that abler people face lower marginal costs. We show that this assumption guarantees that people can make the right schooling choice under full information from the human capital investment point of view: that is, people can self-select or sort among different levels of education based on "sorting gains". We show further that comparative advantage, which has been alluded as the source of this sorting gain, is not a necessary condition for the obtainment of this sorting gain because this gain can also be derived under hierarchical sorting. Moreover, the assumption has empirical implications on the shape and variability of marginal returns to schooling. We derive some empirical implications in a binary choice parametric framework and in a sequential choice framework.

Although the empirical implications of our model on the variability of marginal returns are developed in a parametric binary model solely, we outline the different implications of the human capital (acting as a catalyst of abilities) vs. the signaling hypothesis on such variability. We point out the possibility of further research on the empirical distinction between the two hypotheses upon the base of such variability.

Our analysis of empirical evidence found by different authors from U.S. and Spanish data lead us to conclude that the assumption that abler people faces lower marginal costs cannot be established as an assumption universally valid. While we can assume similar distributions of latent abilities for the populations of any country, empirical evidence demonstrates that different countries and market environments impose different relationships between “ability”, understood as the person-specific value added as paid by the market, and person-specific costs of education. The reasons under these differences come from differences in skills premiums as well as from differences in levels of market uncertainty and possibilities to finance investments in education and diversify risks.

The empirical findings for the two economies point out that the U.S. market environment and educational system allows sorting of individuals between high school and college based upon an individual wealth maximization decision. In this sense, empirical evidence for the U.S. supports the assumption that abler people faces lower marginal costs of schooling although there is some room for credit constraints and uncertainty. In contrast, in Spain, empirical evidence does not support the assumption which allows making well informed human capital investment decisions.

To highlight and complement the evidence based on estimates from policy evaluation models which present the differences between the U.S. and Spain, we point out different reasons which explain the different evolution of the skill premium in these countries. On the one hand, the increase in the skill premium in the U.S. during the eighties has been explained by an increase in the demand for higher ability (Taber, 2001 and Tobias, 2003) and, alternatively, by an increase in the capital’s flexibility of the economy (Mitchell, 2005). On the other hand, the decrease in the skill premium in Spain could be explained by an excess of supply of human capital or the overeducation phenomenon experienced in Spain during the last decades. Moreover, overeducation is found to impose costs on individuals, reducing earnings and lowering job satisfaction (Alba-Ramírez, 1993 and Hartog, 2000).

6. REFERENCES

Alba-Ramirez, A. (1993): "Mismatch in the Spanish labor market?" *Journal of Human Resources*, 28 (2), pp. 259–278.

Becker, G.S. (1993): *Human Capital. A Theoretical and Empirical Analysis with Special Reference to Education.* Third Edition. Chicago: The University of Chicago Press.

Becker, G.S. (1967): *Human Capital and the Personal Distribution of Income.* Ann Harbor Michigan: University of Michigan Press.

Becker, G.S and B.R. Chiswick (1966): "Education and the Distribution of Earnings" *American Economic Review*, 56(1), pp. 358-369.

Berhrman, J. and N. Birdsall (1983): "The Quality of Schooling: Quantity Alone is Misleading", *American Economic Review*, 73(5), pp. 928-946.

Card, D. (1999): "The Causal Effect of Education on Earnings" , in *Handbook of Labor Economics*, Volume 3A, ed. By Orley Ashenfelter and David Card. Amsterdam and New York: North Holland.

Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems", *Econometrica*, 69(5), pp.1127-1160

Carneiro, P. and J.J. Heckman (2002): "The Evidence on Credit Constraints in Post-Secondary Schooling", *Economic Journal*, 112(482), pp. 705-734.

Carneiro, P., J.J. Heckman and E. Vytlacil (2005): "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education", unpublished.

Cunha, F, J. Heckman and S. Navarro (2005): "Separating Uncertainty from Heterogeneity in Life Cycle Earnings", Oxford Economic Papers 57(2), pp. 191-261.

Cunha, F. and J. Heckman (2006): "A New Framework for the Analysis of Inequality", *Journal of Macroeconomics*, forthcoming.

Cunha, F. and J. Heckman (2007): "Identifying and Estimating the Distribution of *Ex Post* and *Ex Ante* Returns to Schooling: A Survey of Recent Developments", *Labour Economics,* In Press.

Griliches, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems", *Econometrica*, 45(1), pp. 1–22.

Hartog, J. (2000): "Over-education and Earnings: Where Are We, Where Should We Go?", *Economics of Education Review*, 19, pp. 131-147

Heckman, J. and S. Navarro (2006): "Dynamic Discrete Choice and Dynamic Treatment Effects", *Journal of Econometrics* 136, pp. 341-396.

Kagan, A., T. Linnik and C. R. Rao (1973): *Some Characterization Theorems in Mathematical Statistics, New York: Wiley*.

López-Bazo, E. and R. Moreno (2007): "Some Evidence on the Aggregate Profitability of Human Capital", Manuscript, AQR Research Group—IREA, University of Barcelona & Barcelona Science Park

Mitchell M. (2005): "Specialization and the Skill Premium in the 20th Century", International Economic Review, 46(3), pp. 935-955.

Simón, H. (2007): "La desigualdad salarial y su evolución en España (1995-2002): Una exploración con datos emparejados empresa-trabajador", Paper presented at the "X Encuentro de Economía Aplicada" Logroño, Spain.

Taber, C. (2001): "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability", *Review of Economic Studies* 68, pp. 665-691.

Tobias, J. (2003): "Are Returns to Schooling Concentrated Among the Most Able? A Semiparametric Analysis of the Ability-Earnings Relationships", Oxford Bulletin of Economic and Statistics, 65(1), pp. 1-29.

Willis, R. and S. Rosen (1979): "Education and Self-Selection", *Journal of Political Economy*, 87(5), pp. 507-536.

Zamarro, G. (2006): "Accounting for Heterogeneous Returns in Sequential Schooling Decisions", Discussion Paper No. 2006-21 CentER, Tilburg University.

7. APPENDIX

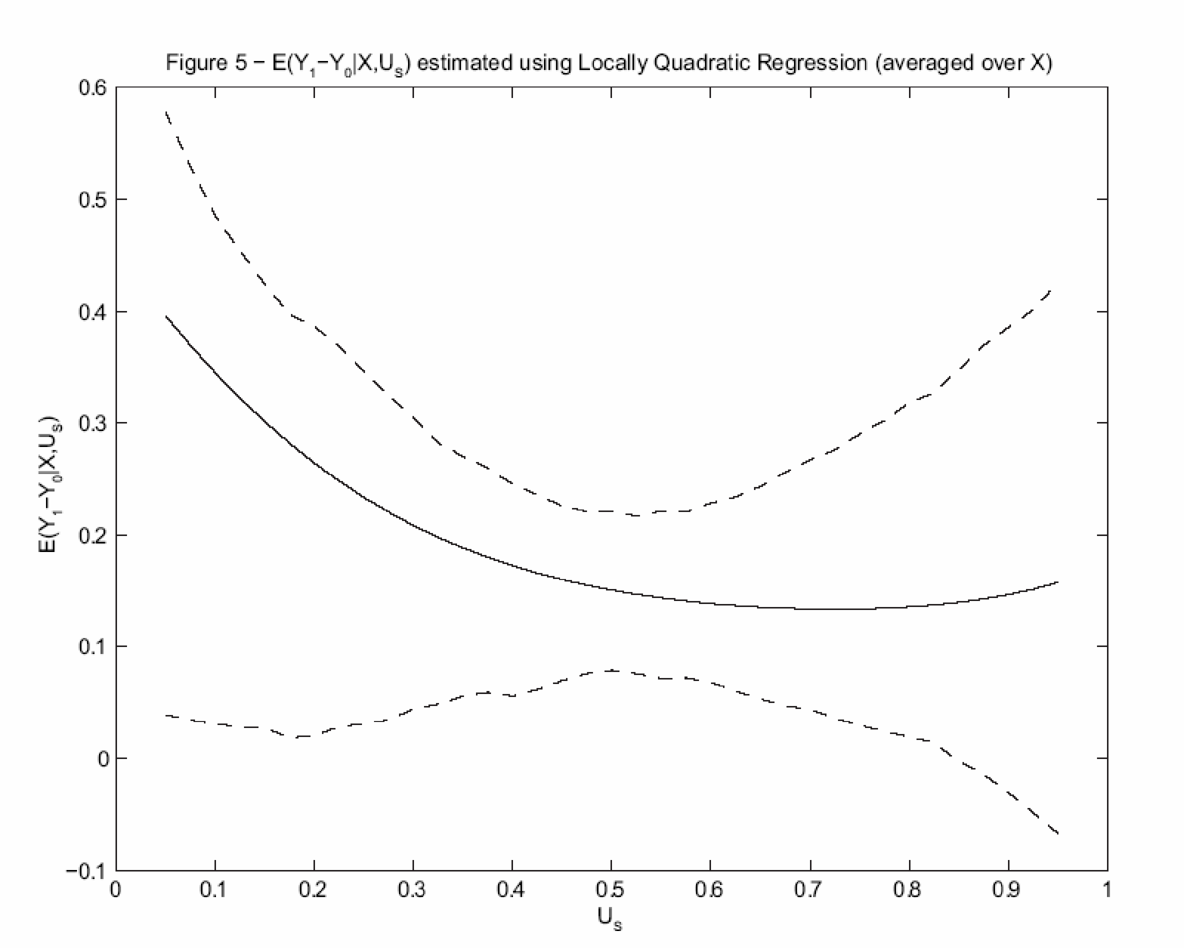
Figure 5 − E(Y1−Y0|X,Us) estimated using Locally Quadratic Regression (averaged over X)
E(Y1−Y0|X,Us)
Us
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

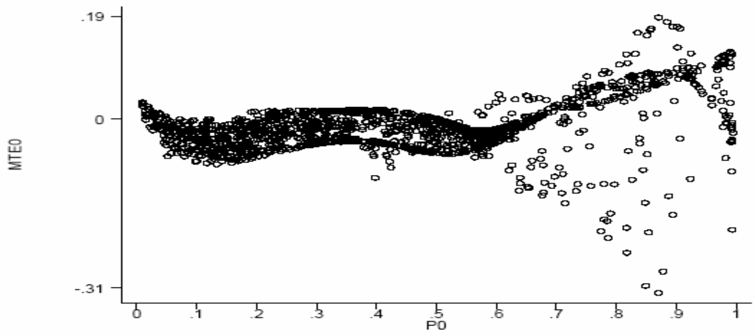

Figure 8. MTE_0 as a Function of the Probability of Continuing Studying after the Low Level (π_0)

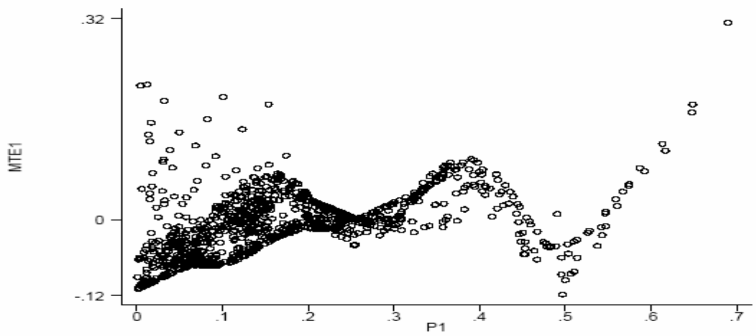

Figure 9. MTE_1 as a Function of the Probability of Continuing Studying after the Medium Level (π_1)