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. The empirical implications of the model are extended further from the properties of ordinary least squares and instrumental variable estimators and centered on predictions about the sign of different policy evaluation parameters and on the shape and variability of marginal returns to education. Empirical evidence shown by a binary choice model supports the assumptions that abler people face lower marginal costs of schooling and higher variability of returns. Evidence obtained from a sequential choice model also supports the former assumption.

JEL Classification: C10, D84, I21, J31
Keywords: Ability gap, Schooling, Selection Models, Heterogeneity

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

Correspondence address: Departament d’Economia. Universitat Jaume I. Campus Riu Sec s/n. 12071 Castelló (Spain).
Tel.: +34964387166. Fax: +34964728591. e-mail: zamoram@eco.uji.es
“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.

This paper will attempt to contribute to this debate by integrating and formalizing some of the new key evidence that has been put forward throughout the last few years. We contend that, despite the volumes of literature this topic has generated, the impact of the positive correlation between ability and education on wage disparity has not yet been fully clarified. Thus, the debate has traditionally been framed as a “nature vs. nurture” exercise to separate the contribution to individual wages of innate ability from that of educational investment, without always acknowledging that education does not just teach new skills and techniques but also allows innate abilities, which are not necessarily distributed evenly across the population, to manifest themselves and develop. In other words: against a “pure” ability premium, which would present no correlation to education, and a “pure” educational premium, which would provide a standard return proportional to the investment made, this ability-developing effect of education would result in returns on education displaying a large variability across individuals, as it would depend not just on the investment itself but also on the presence or not of those latent abilities in the individual. The heterogeneity in returns to education has been modeled and estimated in policy evaluation models since the pioneering work of Willis and Rosen (1979).

What has not been done so far is to derive the empirical consequences 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

\(^1\) The theoretical model on education and the distribution of earnings had been previously developed by Becker and Chiswick (1966) and Becker (1967).
heterogeneity of returns to schooling and extends its empirical implications further from the properties of ordinary least squares and instrumental variable estimators. We obtain empirical implications on the sign of different policy evaluation parameters and on the shape and variability of marginal returns to education.

This paper is structured as follows:

In the two following 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 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 the distribution of marginal returns in a binary and in a sequential choice framework.

Then we proceed in section four to analyze empirical evidence provided in two works which estimate returns to schooling for a binary and a sequential choice, respectively.

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 on part of these unobservables although she can also face some uncertainty on them.

Let’s \( Y^i_k \) denote the realized earnings which person \( i \) is to obtain if she achieves the level \( k \) of education. The ex ante observed component of such earnings is denoted by \( X^i \beta_k \), being \( \beta_k \) the
market retribution of the observed characteristic $X$. The unobserved component, denoted by $U^i_k$, contains the wage premium associated to 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. At the decision moment, person $i$ makes the choice of enrolling in a higher level $k$ of schooling depending on her perception of the potential earnings gains given her information set at this moment: $E(Y^i_k - Y^i_{k-1} | I^i)$, and on her perception of the costs associated to her schooling decision: $E(C^i_k | I^i)$. Similarly as with earnings, costs of schooling can be decomposed into observable pecuniary costs, $Z^i Y_k$, plus unobserved non-pecuniary costs, $U^i_C$, which are related to idiosyncratic tastes and effort partially known by person $i$. Observable variables affecting costs can include $X^i$ as well as additional variables $Z^i$ which shift costs without affecting earnings, such that $Z^i = (X^i, Z^i)$. All the relevant variables affecting ex ante the schooling decision are contained in the information set $I^i$. Once the decision is made, the realized gains will be revealed subsequently.

The net expected gain derived from enrolling at level $k$ of schooling can be written by the index function $S^i_k = E(Y^i_k - Y^i_{k-1} - C^i | I^i)$ such that the observed rule of binary choice is captured by the Bernoulli random variable $S^i_k = 1$ if $S^i_k > 0$, and $S^i_k = 0$ otherwise.

In the Willis and Rosen’s model, as well as in more recent models based on policy evaluation techniques, the relationship among the unobservables $U^i_k$, $U^i_{k-1}$ and $U^i_C$ 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”. 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 $U^i_C$ 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. 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 $\left( U^i_k - U^i_{k-1} \right)$. 

4
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. The part of the ability gap that is known and acted on by the individual when she makes her choice is \( E(U'_k - U'_{k-1} | I') \).

By adopting a typical formulation of the latent index model in policy evaluation (see for example Carneiro, Heckman and Vytlacil (2005)): \( S^i_k = 7_k - U^i_{Sk} \), we can express the unobservable heterogeneity related to the schooling decision, \( U^i_{Sk} \), in terms of the ability gap and of nonpecuniary costs such that \( U^i_{Sk} = [U^i_{ck} - (U^i_k - U^i_{k-1})] \). 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 \( U'_c = 0 \)) 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.** \( Cov(U^i_{ck}, U^i_k - U^i_{k-1}) \leq 0 \), 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 \( U^i_{ck} \), \( U^i_k \) and \( U^i_{k-1} \). And yet, our assumption allows us to restrict the relationship between \( U^i_S \) and the ability gap as we proof in the following result.

**Proposition 1.** \( Cov(U^i_S, U^i_k - U^i_{k-1}) < 0 \), 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:**

\[
Cov(U^i_S, U^i_k - U^i_{k-1}) = Cov(U^i_c, U^i_k - U^i_{k-1}, U^i_k - U^i_{k-1}) = Cov(U^i_c, U^i_k - U^i_{k-1}) - Var(U^i_k - U^i_{k-1})
\]

From this expression, it is immediate to show that under assumption 1 the whole expression of the covariance is negative. 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.

Introducing some notation, $\text{Cov}(U^i_s, U^i_k - U^i_{k-1}) = \text{Cov}(U^i_s, U^i_k) - \text{Cov}(U^i_s, U^i_{k-1}) = \rho_{k,S} - \rho_{k-1,S}$.

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 $E(U^i_k - U^i_{k-1} - U^i_C | I^i)$, that is, $E(-U^i_s | I^i)$. 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 $U^i_s$. 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 $U^i_C$). 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 $U^i_s$. That is, $U^i_s$ 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 $U^i_s$. In what follows, our task is to predict the relationship between marginal returns to schooling (for the marginal entrant who is indifferent between two subsequent levels of education) and the values taken by the idiosyncratic gain as expressed by the random variable $U^i_s$.

Next, we discuss the implications of assumption 1 on the shape of the distribution of marginal returns 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.

Assume that the idiosyncratic wage premiums, \(U_1^i\) and \(U_0^i\), are not fully contained in the agent’s information set at the decision moment or that they depend on random factors potentially contained in the information set. Person \(i\), who made a well informed decision before entering the level of education \(1\), obtains ex post the following wage premiums with respect to the average person in each sector (\(I\) and \(0\), respectively) if she chose to enter the respective sectors:

\[
E(U_1^i \mid S = 1) = -\rho_{1,s} \frac{f(\bar{y})}{F(\bar{y})} \quad \text{and} \quad E(U_0^i \mid S = 0) = \rho_{0,s} \frac{f(\bar{y})}{1 - F(\bar{y})}
\]

where \(f(\cdot)\) and \(F(\cdot)\) are the density and cumulative distribution functions of \(U_s\), respectively.

Comparative advantage implies that earnings of people who choose each sector are greater than those of the average person in each sector. That is, \(E(U_1^i \mid S = 1) > 0\) and \(E(U_0^i \mid S = 0) > 0\). On the other hand, hierarchical sorting implies that those choosing the level \(I\) perform better in both groups than the average person of each group, but they are better at level \(I\) than at level \(0\), that is, \(E(U_1^i \mid S = 1) > 0, E(U_0^i \mid S = 1) > 0\) and \(E(U_1^i - U_0^i \mid S = 1) > 0\). 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 \((\rho_{1,s} - \rho_{0,s}) < 0\), is compatible with 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 \(I\). 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:
\[ E(Y^i_i - Y^o_o \mid X) = X(\beta_i - \beta_o) \]

The average return for those who achieved the upper level (Treatment on the Treated: TT) is:
\[ E(Y^i_i - Y^o_o \mid X, S = 1) = X(\beta_k - \beta_{k-1}) + E(U^i_i - U^o_o \mid S = 1) = X(\beta_k - \beta_{k-1}) - (\rho_{1,s} - \rho_{0,s}) \frac{f(\gamma)}{F(\gamma)} \]

The average return for those who left education at the lower level (Treatment on the Untreated: TUT) is:
\[ E(Y^i_i - Y^o_o \mid X, S = 0) = X(\beta_k - \beta_{k-1}) + E(U^i_i - U^o_o \mid S_k = 0) = X(\beta_k - \beta_{k-1}) + (\rho_{1,s} - \rho_{0,s}) \frac{f(\gamma)}{1 - F(\gamma)} \]

The OLS bias or selection bias is:
\[ E(U^o_o \mid S = 1) - E(U^o_o \mid S = 0) = -\rho_{0,s} \left( \frac{f(\gamma)}{F(\gamma)} + \frac{f(\gamma)}{1 - F(\gamma)} \right) \]

Therefore, under proposition 1 and for any distribution of \( U_s \), 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 which is the interpretation normally done in policy evaluation models. However, we have seen that this result is also compatible with hierarchical sorting of any type. For the OLS selection bias, it is negative under comparative advantage and positive under hierarchical sorting.

In policy evaluation models, the marginal return to education is defined as the mean gain to schooling for individuals of given characteristics, \( X = x \), who are indifferent between taking schooling or not at the level of unobservables \( U_s = u_s \). This marginal return is measured by the parameter known as Marginal Treatment Effect (MTE) which, in our terms, corresponds to the following definition:
\[ MTE'(x, u_s) = E(Y^i_i - Y^o_o \mid X = x, U^i_i = u_s) = \left[ x(\beta_i - \beta_o) \right] + E(U^i_i - U^o_o \mid U^i_i = u_s) \]

Analogously, 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 \( p(z) \).²

² In latent index models it is assumed without loss of generality that \( U_s \) is uniformly distributed between 0 and 1. Also, it is assumed that \( E(Y_i - Y_o \mid Z) = E(Y_i - Y_o \mid \Pr(S = 1 \mid Z)) \) according to the index sufficiency restriction assumption frequently used in the selection models’ literature.
The term \( E(U_i - U_0 | U_s = u_s) \) is the regression function of the ability gap on \( U_s \). It is commonly assumed that this function is linear: \( E(U_i - U_0 | U_s = u_s) = \eta u_s \) (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 \( E(Y|X) \) yields a linear function). Given that \( \eta = \text{Cov}(U_i - U_0, U_s)/\text{Var}(U_s) \), proposition 1 implies that MTE is a decreasing function in \( u_s \) (or in the propensity score). In policy evaluation models, \( U_s \) is assumed to be distributed as a Uniform[0,1] so that it is possible reparameterize the model for any random variable \( V \) with \( p(z) = \Pr(V < \bar{\gamma} | Z = z) = F_{\bar{\gamma}}(z) \), \( U_s = F_{\gamma}(V) \), \( u_s = \Pr(V < v) \) and \( v = F_{\gamma}^{-1}(u_s) \). Therefore, MTE is decreasing in both \( u_s \) and \( v \). If the original random variable \( V \) 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 \( v \).

In a parametric framework, the distribution of \( (U_i - U_0) \) given \( U_s \) can be degenerate so that we do not observe variability of marginal returns for a given value of \( U_s \). In the estimation of \( \hat{\eta} \), however, we know that \( \text{Var}(\hat{\eta}) = \left[ \text{Var}(U_i - U_0)/\text{Var}(U_s) \right] \). To identify parametric models, \( \text{Var}(U_s) \) is assumed to take a given value, for example 1. Therefore, the variance of estimated marginal returns across the population equals the variance of the ability gap (taken fixed values of observable characteristics).

This result implies that 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, 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.

### 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). A sequential setting with three levels of education can be parametrically expressed by a joint distribution function of the form...
\[
\begin{bmatrix}
U_k \\
U_{s1} \\
U_{s2}
\end{bmatrix} 
\sim 
\mathcal{F}
\left(
\begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix},
\begin{bmatrix}
\sigma_k^2 & \rho_{k,s1} & \rho_{k,s2} \\
\rho_{k,s1} & 1 & 0 \\
\rho_{k,s2} & 0 & 1
\end{bmatrix}
\right) 
k = 0, 1, 2
\]

where \( S_1 = 1(\gamma_1 - U_{s1}) = 1 \) if the individual left the school at level 1 and \( 1(\cdot) \) is the indicator function, and similarly for \( S_2 \).

Assumption 1 applied to both levels entails
\[
\text{Cov}(U_{c1}^i, U_{1}^i - U_{0}^i) \leq 0 \quad \text{and} \quad \text{Cov}(U_{c2}^i, U_{2}^i - U_{1}^i) \leq 0
\]

In these terms, proposition 1 applied in this sequential setting implies
\[
(\rho_{1,s1} - \rho_{0,s1}) < 0 \quad \text{and} \quad (\rho_{2,s2} - \rho_{1,s2}) < 0 \quad \text{(2)}
\]

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 \( \rho_{k,sk} \) solely. Their idiosyncratic gain or loss is also related to the choice they did at a first moment (they decided to complete the level \( k-1 \)) and to the choice they did subsequently (they decided to leave education at the level \( k \)).

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:
\[
E(Y_2^i \mid S_1 = 0, S_2 = 1) = X\beta_2 + E(U_2 \mid S_1 = 0, S_2 = 1)
\]

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, \( \tilde{S} = 1(\gamma - U_{\tilde{s}}) \), where \( \gamma = S_1 \gamma_1 + S_2 \gamma_2 \) and \( U_{\tilde{s}} = S_1 U_{s1} + S_2 U_{s2} \).

This way, the above expression gives
\[
E(Y_2^i \mid \tilde{S} = 1, S_2 = 1) = X\beta_2 + E(U_2 \mid \tilde{S} = 1, S_2 = 1) = X\beta_2 - \left[ \rho_{2,\tilde{s}} f(\frac{\gamma}{\gamma}) \right] + \rho_{2,\gamma_2} \left[ \frac{f(\gamma_2)}{F(\gamma_2)} \right]
\]

Therefore, proposition 1 has to be expressed in terms of the random variable \( \tilde{S} \) instead of in terms of \( S_1 \).
We define the unobservable costs that will be faced by a person who overcomes the level 0 by 
\[ U_c = S_1 U_{c1} + S_2 U_{c2}. \]

Therefore, the expression of the covariance between these costs and each ability gap gives:

\[ \text{Cov}(U_c, U_0 - U_0) = \text{Cov}(S_1 U_{c1}, U_1 - U_0) + \text{Cov}(S_2 U_{c2}, U_1 - U_0) \] (4)

\[ \text{Cov}(U_c, U_2 - U_1) = \text{Cov}(S_1 U_{c1}, U_2 - U_1) + \text{Cov}(S_2 U_{c2}, U_2 - U_1) \] (5)

\[ \text{Cov}(U_c, U_0 - U_0) = \text{Cov}(S_1 U_{c1}, U_2 - U_0) + \text{Cov}(S_2 U_{c2}, U_2 - U_0) \] (6)

We cannot assume any sign on expressions (4) and (5) 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 (6) 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:

**Proposition 2.** If (i) \[ \text{Cov}(U_{c1}^i, U_{c2}^i - U_0^i) < 0, \]
(ii) \[ \text{Cov}(U_{c1}^i - U_0^i, U_{c2}^i - U_0^i) > 0 \text{ and} \]
(iii) \[ \text{Cov}(U_{c2}^i - U_0^i, U_{c2}^i - U_0^i) > 0, \text{ then} \]

\[ \text{Cov}(U_{c1}^i, U_{c2}^i - U_0^i) < 0. \]

**Proof:** It can be shown by developing the expression:

\[ \text{Cov}(U_{c1}^i, U_{c2}^i - U_0^i) = \text{Cov}(U_{c1}^i, U_{c2}^i - U_0^i) - \text{Cov}(U_{c1}^i - U_0^i, U_{c2}^i - U_0^i) \Pr(S_1 = 1) - \\
- \text{Cov}(U_{c2}^i - U_0^i, U_{c2}^i - U_0^i) \Pr(S_2 = 1) \]

Therefore, 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 unobservable heterogeneity of the early schooling decision as measured by its expected net unobservable costs. Furthermore, this negative 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 will be positive and they will be larger for larger nurture effects of education and for larger variability of nature.

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).
If proposition 2 holds, then expression (8) is positive. Therefore, those who choose to complete the highest level of education do so informed by a positive sorting gain with respect to the average person. In contrast, we cannot predict that the sorting gain defined in (7) for those who choose leaving education at the medium level is positive. Stronger assumptions would be necessary for doing such prediction.

Zamarro (2007) presents expressions of the marginal return of overcoming the low level of education (MTE₀) and of the marginal return of achieving the highest level of education (MTE₁) as functions of the propensity scores of the respective decisions for given values of the propensity scores of the subsequent and preceding decisions, respectively.

In a parametric framework under normality, it is shown that MTE₀ and MTE₁ are linear functions with respect to $\nu = F^{-1}(u_s)$ which slopes are respectively:

$$E(U_1 - U_0 | S_1 = 1) = -(\rho_{1,\delta} - \rho_{0,\delta}) \frac{f(y_1^*)}{F(y_1^*)} + (\rho_{1,s2} - \rho_{0,s2}) \frac{f(y_2^*)}{(1 - F(y_2^*))}$$ (7)

$$E(U_2 - U_0 | S_2 = 1) = -(\rho_{2,\delta} - \rho_{0,\delta}) \frac{f(y_2^*)}{F(y_2^*)} - (\rho_{2,s2} - \rho_{0,s2}) \frac{f(y_2^*)}{F(y_2^*)}$$ (8)

Without any further assumptions, we cannot predict from expressions (9) and (10) whether marginal returns to education are positively or negatively related to the probability of achieving such education given some observable characteristics.

4. SOME EMPIRICAL EVIDENCE

Empirical evidence on the implications of Proposition 1 and 2 allows us to test the associated assumptions. The empirical properties derived from both propositions are necessary conditions of the assumptions under which such properties are obtained. 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 on the predictions done for the binary choice case in Carneiro, Heckman and Vytlacil (2005). We will add later empirical evidence provided by factor models of binary choice (Cunha, Heckman and Navarro 2005, Cunha, and Heckman 2007) in which we can
estimate the values of $\rho_{1,S}$ and $\rho_{0,S}$ from the estimates of factor loadings provided in these works.

The work by Zamarro (2007) has set the framework of our sequential analysis. This work serves as well as empirical evidence on our predictions obtained upon it.

4.1 BINARY CHOICES

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 (level 0) or above (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).

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 linear function of MTE with respect to $v$ where $v = \Phi^{-1}(u_{s})$ and $\Phi$ is the cumulative normal distribution function. For this 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 and then pointing out comparative advantage. It is also possible to measure the bias $E(U_{1} | S=1, I^{1}) - E(U_{1} | S=0, I^{1}) = -\rho_{1,S} \left( \frac{f(y)}{F(y)} + \frac{f(y)}{1-F(y)} \right) = OLS-TUT=0.0165$ Therefore we obtain $\rho_{1,S} < 0$ which confirms further the hypothesis of comparative advantage.

In the semiparametric model, several estimation methods of the MTE function give decreasing functions with respect to the propensity score $u_{s}$, 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 seems contradictory with proposition 1, case in which the choice can be
informed in terms of comparative disadvantage. In this sense, those with the lowest probabilities of attending college could be performing worse than average in both sectors.

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 to 10%
- Sorting gain for high school: TUT-ATE= -7 to -10%
- OLS selection bias: OLS-TT= -15 to -23%
- Bias for college premium: OLS-TUT= -1 to -2%

These results are consistent with proposition 1 and indicate a slight hierarchical sorting effect.

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. Is the variability of the ability gap larger for these groups? Does education have a stronger nurture effect on these groups?

4.1 SEQUENTIAL CHOICES

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 ("Encuesta de Conciencia y Biografía de Clase") which allows her to use different instruments for each decision stage to identify the model.

The paper by Zamarro (2007) provides estimates on several covariances between the ability gap and the schooling decision heterogeneity at various levels of education. Specifically, she presents the following estimates from a parametric model under normality:

\[
\begin{align*}
(\rho_{1,5} - \rho_{0,5}) &= 0.32^* \\
(\rho_{1,52} - \rho_{0,52}) &= -0.55^* \\
(\rho_{2,5} - \rho_{0,5}) &= -0.053^* \\
(\rho_{2,52} - \rho_{0,52}) &= 0.41
\end{align*}
\]
These estimates indicate that the sorting gain for those who achieve the college level (2) with respect to the lowest level (0), \( E(U_2 - U_0 \mid S_2 = 1) \), is positive and the sorting gain for those who achieve the medium level (1) with respect to the low level, \( E(U_1 - U_0 \mid S_1 = 1) \), is negative.

Proposition 2 predicts that \( \left\{ \rho_{2,3} - \rho_{0,5} \right\} \) as empirical evidence confirms. So, empirical evidence does not reject the three assumptions under proposition 2. The estimate for \( \left\{ \rho_{2,5} - \rho_{0,5} \right\} \) is not significantly different from zero. This evidence joint with the assumption that \( \text{Cov}(U_{c1}, U_2 - U_0) < 0 \) leads us to support the assumptions that \( \text{Cov}(U_{c1}, U_2 - U_0) < 0 \) and \( \text{Cov}(U_{c2}, U_2 - U_0) \approx 0 \).

In contrast to the decreasing pattern of MTE with respect to the propensity score obtained for binary choice models (which means that those with higher probabilities of continuing studying after the level 0 have higher marginal returns), in Zamarro’s parametric model we observe increasing marginal returns with respect to both propensity scores. This result translates to higher marginal returns for those with lowest probabilities of studying further from level 0, given the chances of completing the college level, and for those with lowest probabilities of achieving the college level, given the chances of overcoming the lowest level.

The semiparametric results show departures from normality and the pattern of variability of marginal returns. For the decision of whether to continue studying after the lower level, we observe more variability in the returns of those with the lowest probabilities of deciding to do so. Conversely, when facing the decision of continuing education after a medium level, those with the highest probabilities of doing so face the highest variability in marginal returns.

5. CONCLUDING REMARKS

The debate of nature vs. nurture in the policy evaluation approach has been going on for decades, and it will take a very substantial weight of evidence to bring it anywhere near an eventual closure. The fact that for a significant portion of the duration of this debate the econometric techniques employed did not support tests of the ability-enhancing effects of education, however, has all-too-often led the debate towards a radical dichotomy (either education as a form of human capital formation or as a pure “signaling” mechanism to reveal underlying innate abilities) which does not appear to be supported by the observations. In this sense, novel statistical techniques such as those used by Carneiro, Heckman and Vytlacil (2005) or by Zamarro (2007) may actually point out the way towards a breakthrough. Although the evidence these papers provide is still far from
conclusive, our analysis highlights that, on the basis of their results, one would indeed be inclined to conclude that the value of education may to a large extent be due to its role in acting as a “catalyst” for the individual’s innate abilities i.e. in helping to bring to productive manifestation a set of latent abilities that, while innate to the individual, without exposure to education would most likely remain dormant.

Sure enough, a lot more research needs to take place before these results can be deemed “robust”. Nevertheless, at this point they already indicate that there is at least some statistical evidence that the traditional nature vs. nurture dichotomy in the policy evaluation approach may actually be more appropriately regarded as an interaction between innate and acquired abilities than as a strict opposition between the two.
REFERENCES


