College Choice:
Estimating the Determinants of Enrollment

Lidia Farré Olalla
European University Institute

November, 2005
[Job Market Paper]

Abstract

This paper estimates a model of schooling choices for a sample of high school graduates in the early 1980’s. Combining individual data from the NLSY 79 and information on labor market conditions from the CPS, I explore the decision of youths to invest in higher education in response to their economic incentives and constraints. Schooling choices are modeled using a random utility model in which decisions can be probabilistically estimated. The empirical approach employs a semiparametric double index based estimator for multiple choice models. Under the assumption that students use current wages to predict the size of their own returns to education, I find statistically, and economically, important effects from expected earnings on schooling choices. These findings suggest that changes in labor market conditions affect the composition and the rate of college enrollment.

JEL Classifications: C14, C35, I21

Keywords: Schooling Decisions, Choice Models, Semiparametric Estimation.

---

*I would like to thank Omar Licandro and Frank Vella for helpful supervision and useful discussions; I am particularly grateful to Roger Klein and Frank Vella for developing the econometric ideas, and to Lisa Lynch and Morten O. Ravn for support and guidance; I also benefited from the comments of Stephan Fahr, Liliane Karlinger, Christopher Milde, Bernat Tiana and Anzelika Zaiceva. The usual disclaimer applies. Address: European University Institute, Via della Piazzuola 43, 50133 Florence, Italy. lidia.farre@iue.it. http://www.iue.it/Personal/Researchers/farre*
1 Introduction

The US college enrollment rate of high school graduates increased slightly in the three decades from 1960 to 1990. College enrollment declined from approximately 57 percent in the late 1960's to 50 percent in the mid-1970's, stabilized over the late 1970's, and slowly rose in the 1980's and the 1990's. In the mid-1990’s however, it was only 4 percentage points above that in the 1960's, when schooling attendance increased in response to the Vietnam era draft laws. The slow growth in college attainment may have implications for the economy as a whole. The shift in the relative supply of highly educated workers, together with a steady increase in the relative demand for high skilled workers may have accelerated the rise in the college-high school wage gap and the increase in income inequality (Card and Lemieux, 2001b). Theories of growth have also identified human capital as the main engine of long term growth (Lucas, 1988). Accordingly, if the slow growth in enrollment affects the aggregate level of human capital, it might well have implications for the future rate of economic growth.

The trends in enrollment behavior over the last decades are an important focus of research for labor economists. Card and Lemieux (2001a) use aggregate time series data to document these trends, and find that the large size of the cohorts born after the 1950's partly explains the slow growth in enrollment, as students may be "crowded out" of college. However, students in the same cohort face different incentives and opportunities at the time they decide to continue with further education. Understanding the role of these factors in influencing schooling outcomes is a prerequisite for addressing enrollment growth and income distribution questions. In this paper, I use individual micro data from a sample of high school graduates in the early 1980's, to explore the schooling choices of youths in response to their economic incentives and constraints.

In the simple human capital investment model, students attend college if the present value of the expected gains from schooling are greater than or equal to the costs of college.
(Becker, 1967). A number of previous studies have examined the role of credit constraints and long term family factors in college participation decisions (see, for instance, Carneiro and Heckman 2002, Ellwood and Kane 2000). Since Roy (1951), there is also an extensive literature on the selective nature of schooling choices and the role played by "ability rents" in the expected returns to education. While these issues are also considered here, this paper focuses on an aspect of human capital accumulation that has not previously received much attention (Dellas and Koubi, 2003 being an exception). This relates to the effects of labor market conditions on the size and type of students' human capital investment upon high school graduation.

The decision to continue with further education depends upon the interaction of two factors, both of which are affected by the state of the economy: the expected returns to schooling and the opportunity cost of college. At the time students make their decisions, they use information on current wages or beliefs regarding future wages to gauge the size of their gains from schooling (Ryoo and Rosen, 2004). Accordingly, an improvement in the level of wages and job opportunities for skilled workers may encourage students to attend college. In contrast, a low level of unemployment might divert students from college as a result of the increase in employment opportunities and the amount of income forgone while going to school.

In this study, I use a sample of respondents from the National Longitudinal Survey of Youth 1979 (NLSY 79) to test the role of labor market conditions and family and individual background characteristics in enrollment decisions. The NLSY 79 is a nationally representative sample of 12,686 young men and women who were 14-22 years old when they were first surveyed in 1979. This sample is useful for the purpose of this paper because it contains detailed information on schooling paths for a sample of students who graduated from high school in the early 1980's. To estimate the sensitivity of enrollment to the expected returns to education, I use information on wages taken from the Current Population Survey (CPS) and construct lifetime earnings estimates for different educational alternatives. Although
the human capital investment model predicts that lifetime earnings are key to enrollment decisions, they have a substantial unforecastable component at the time students decide to go to college (Cuhna, Heckman and Navarro, 2005). This is due to individual heterogeneity and uncertainty about future demand and supply for labor. I experiment with different models of expectations about the returns to education that presume different sets of information on the part of the student. In each model, a set of assumptions allows me to decompose the discounted present value of lifetime earnings into two components, an expected component that students can predict and act on when they make their decisions and an unexpected component that they cannot predict.

This paper analyzes the decision to accumulate human capital upon high school graduation as previous studies have documented that this is a critical point in the student’s schooling path. Light (1995) for instance, finds that the highest school dropout rate occurs after completion of the 12th grade, and that most of the students who leave school at that time do not return afterwards. Given the substantial role of education in shaping life-cycle earnings profiles, it is important to understand the factors that influence college attainment. I estimate a multiple choice model of human capital investment in which the traditional binary university enrollment decision is extended to include the possibility of attending a community college for two years. Rouse (1994) argues that community colleges have become important providers of post-secondary education, training students who may not have attended university otherwise. In the empirical model each student upon high school graduation faces a choice among a set of three alternatives: starting in two-year college, starting in four-year college or not attending college at all.

Standard estimation techniques for multiple choice situations such as the multinomial logit and multinomial probit rely on assumptions regarding the parametric distribution of the unobserved terms in the model. If these assumptions are not satisfied, their estimates are inconsistent. Recent developments in semiparametric and nonparametric techniques allow the choice probability function for each alternative to be estimated with minimal
assumptions regarding the distribution of unobserved terms. In this paper, I extend the
semiparametric double index based estimator for binary choice models developed in Klein

The evidence here suggests important implications for enrollment behavior. First, I
find that students with a higher level of academic preparation and with more educated
parents are more likely to attend a four-year college. These factors are less important
for two-year college entrance, suggesting that community colleges offer a place in post-
secondary to students from different family backgrounds and different levels of academic
preparation. Second, I also find a statistically and economically significant effect from
expected earnings and job opportunities on schooling outcomes. More precisely, a one
standard deviation increase in the expected earnings for a particular alternative increases
the probability of choosing this alternative by approximately 7 percent (from 0.3209 to
0.3423 for four-year college, from 0.1142 to 0.1231 for two-year college, and from 0.5649
to 0.6019 for not attending college). In addition, a one standard deviation increase in the
unemployment rate increases the probability of attending a four-year college by almost 6
percent (from 0.3209 to 0.3398). These results indicate that not only family and individual
characteristics are important for enrollment, but also that labor market conditions play a
role in schooling decisions.

The paper is organized as follows: section 2 discusses basic facts about educational at-
tainment and its determinants. Section 3 presents a random utility framework for modeling
educational choices and discusses the econometric strategy that will be used. Section 4
contains a description of the data. The empirical results are outlined in section 5, followed
by some concluding remarks and possible extensions.
2 The Determinants of Schooling Attendance

Traditional human capital models posit that three main factors influence enrollment in post-secondary education: the expected returns to this education, the costs of additional education and the income of the student’s family. The relationship between family income and schooling outcomes has been widely examined in the literature. This relationship has been interpreted in two different ways. The first emphasizes that family income at the time of enrollment is critical. The second argues that the parental environment and family income available to children over their entire life cycle are more important determinants. Kane (1994) and Ellwood and Kane (2000) are examples of the literature in support of the empirical importance of short term credit constraints. They argue that the rise in college costs (tuition and fees) and the limited access to credit for low income families explain the slow enrollment response of the children of these families to the substantial increase in the college wage premium that began in the early 1980’s. Figure 1 displays the annual mean tuition costs of post-secondary education institutions and the college enrollment rate of high school graduates since the early 1970’s. This Figure reveals that college costs have been increasing since the early 1980’s. Thus, it might have represented an impediment for children of poor families in making their education choices. However, Carneiro and Heckman (2002) show that at most 8 percent of the total US population is credit constrained in a short run sense. Figure 1 also suggests that credit constraints cannot be the only factor influencing schooling outcomes as college costs continued to rise in the 1990’s while college attendance rebounded.

A number of other studies provide empirical evidence to show that long term factors associated with higher family income such as parents’ level of education and permanent income, rather than short-term credit constraints determine schooling outcomes. Cameron and Heckman (2001) nd that once family background is controlled for, minorities are more likely than whites to graduate from high school and attend college. Carneiro and
Heckman (2002) also find that conditioning on long term factors eliminates most of the effect of family income in the adolescent years on college enrollment. These authors argue that better-educated parents may not only have more resources to finance education of a higher quality but they may also be more informed about the costs and benefits associated with the various schooling options. Moreover, educated parents are better able to develop scholastic aptitude and tastes for education in their children. As a result, these children are expected to have a lower marginal cost of schooling and to go further with their education.

Although family background has an indisputable role in enrollment, the predictions of the human capital investment model indicate that other factors such as the expected returns to education and the job opportunities faced by the student upon high school graduation, influence college attendance as well. The model predicts that students choose the education path that maximizes the discounted present value of lifetime earnings, net of education costs. However, students are uncertain about their abilities; the future demand and supply for labor; their probabilities to succeed in their career aspirations; and the position that they will attain within the post-school earnings distribution. In empirical work, one possibility is to include future realized earnings in the students' information set when estimating the determinants of schooling outcomes (Willis and Rosen, 1979). Nevertheless, Cunha et al. (2005) find that, although there is an important part of ex post earnings that can be predicted by students, there is also a substantial component that students cannot predict. In particular, they estimate that if students knew their ex post earnings outcomes, around 30% would change their schooling decisions.

In this paper, I assume that students form expectations about future career prospects conditional on the wage information that they observe or on the basis of their beliefs about future wages. In this context, shifts in the demand and supply for labor that affect the rewards to education might influence enrollment. In addition, I assume that students can predict neither their position within the distribution of unobservable abilities nor their market price. This seems reasonable since students have made few career choices in their lives. Thus
they do not have the experience that allows such knowledge to be acquired. Accordingly, lifetime earnings are estimated by stratifying wages for several cross-sections in the Current Population Survey into cells on the basis of a set of observable characteristics (i.e. gender, race, age, region and urban area). The average wage for each educational alternative in each cell measures the component of earnings that students can predict and act on at the time schooling decisions are made. The difference between individual wages and the average wage in the student's cell measures the unpredictable component of earnings. This captures earnings variations due to the returns to unobservable worker attributes (e.g. ability) and to changes in transitory earnings that may result, for instance, from a weakening of labor market institutions.¹

Before turning to the empirical analysis, it is useful to consider the contribution of these two components to the changes in the distribution of wages during the last decades. In doing so, I follow many others before and decompose the total variance of wages for each year since 1979 in the between-group and the within-group variance.² I regress log weekly wages for full-time male workers on a set of education dummies, region dummies, race dummies, a dummy indicating whether the worker lives in an urban area, and his age.³ The between-group variance is the variation explained by the predicted part of this regression, while the within-group variance is the variation in the residuals. Figure 2 presents the results of this decomposition. It should be noted that, consistently with previous studies, the widening in the earnings distribution in the 1980’s and the 1990’s was accompanied by increases in wage differentials by skill group (between-group wage dispersion) and by much greater residual inequality (within-group wage dispersion).

¹These include the decline in unionization; the disappearance of government wage regulation; and more competitive markets.


³The education groups consist of: high school graduates not enrolled in college, two-year college graduates and four-year college graduates. The region dummies correspond to: Northeast, North Central, West and South. The race dummies are: black, hispanic, non-black and non-hispanic. I also include a quadratic in age.
The recent changes in the wage distribution are likely to have influenced enrollment trends. The increase in the college wage premium may have been an incentive for students to continue with their further education. However, the important rise in the dispersion of the residuals or the unpredictable component of wages represents an increase in earnings uncertainty, which may have had an adverse effect on enrollment if students were risk averse. This paper does not investigate the role of earnings uncertainty in enrollment behavior, but the changes in the wage distribution and enrollment trends over the last decades suggest that this is a promising avenue for future research.

Economic factors that vary over the business cycle such as the unemployment rate, may also affect college participation. At the theoretical level, the effect of an increase in unemployment on schooling attainment is not clear. Dellas and Koubi (2003) argue that opportunity cost considerations make schooling countercyclical, whereas the ability-to-pay has the opposite effect. If students are not credit constrained, enrollment increases during recessions as the probability of finding a job decreases and so does the opportunity cost of college. However, it is also possible that the financial situation of the family worsens during recessions if the main earner loses his or her job. Figure 3 shows the trends in college enrollment and aggregate unemployment over the last decades. The picture suggests that during the 1970's and the 1980's, periods of high unemployment were also accompanied by an increase in enrollment (e.g. in 1975 the unemployment rate was 6.9% and the enrollment rate 50.7%, and in 1983 these were 10% and 52.7%, respectively). However, the sign of this correlation is less clear in the 1990's. This investigation uses data from a sample of students who graduated from high school between 1979 and 1983. In the early 1980's the US economy was affected by an important recession, whose most dramatic consequence was the rise in unemployment. Therefore, in analyzing the educational outcomes of these students, it is important to control for the effect of variations in unemployment.

The following sections investigate in detail the determinants of schooling decisions. In doing so, I estimate a model of schooling choices in which, upon high school graduation,
students choose whether to attend a community college for two years; to attend a four-year college; or not to enroll at all in college. Two-year colleges have become an important alternative means of obtaining higher education qualification. Between 1960 and 1980, the number of public institutions offering two-year college programs rose from 332 to 869 and in the early 1980's over half of all first-time, first-year students chose to enroll in this type of college (Rouse, 1994). Community colleges are considered as institutions that provide vocational (as opposed to academic) training, mainly in job areas that are in relatively strong market demand (Betts and McFarland, 1995). While this characteristic renders two-year colleges as convenient institutions to retool and retrain workers into new occupations, it also makes workers with a two-year college degree more vulnerable to shifts in the labor demand as a result of economic changes. In addition, the low cost and the short-term investment nature of these types of institutions may divert students from four-year institutions. Separately considering two-year and four-year colleges allows the possibility to study the impact that changes in labor market conditions have on the composition of college enrollment.

3 Empirical Framework

3.1 The Decision Process

To study the determinants of enrollment decisions, I estimate the conditional probability that a student will choose a particular educational alternative from a set of alternatives. Willis and Rosen (1979) estimate a college choice model in which students are considered to be income maximizing agents that choose the alternative with the highest present value of earnings. However, the empirical evidence presented in many other studies suggests that factors such as family background and individual scholastic abilities are also relevant for making decisions regarding college enrollment. In this investigation, I allow these nonwage dimensions to enter the decision rule and analyze college choices in the framework of a utility maximization model.
In this context, a student \( i \) that chooses alternative \( l \) in the decision set \( J_i \) - no college \( (l = 1) \), two-year college \( (l = 2) \) or four-year college \( (l = 3) \) - is assumed to obtain utility:

\[
U_{li} = S_{li} + \varepsilon_{li},
\]

(1)

The utility function for each alternative contains an observed part \( (S_{li}) \) and a part that is not observed by the researcher \( (\varepsilon_{li}) \). The observed part is assumed to be characterized by a single index, \( S_{li} = X_i \beta_l \), where \( \beta_l \) is a \( q \times 1 \) vector of unknown coefficients, \( X_i \) is a \( 1 \times q \) vector of individual-specific characteristics such as socioeconomic status and measured ability, and \( i = 1,...,N \).

The student chooses the alternative that maximizes his utility. Formally, the probability of choosing each alternative is:

\[
P(d_{li} = 1) = P(\mu_{Li} > X_i \beta_{L} \mid X_i \beta_l) \text{ for all } L \text{ in } J_i \text{ such that } L \neq l,
\]

(2)

where \( \mu_{Li} = \varepsilon_{Li} \) and \( d_{li} \) is a discrete variable that takes value 1 if alternative \( l \) is chosen and 0 otherwise. The assumptions regarding the distribution of the unobserved terms in the model, the \( \varepsilon_{Li} \), lead to different functional forms for the choice probability in (2). Generally, there is little economic guidance regarding the choice of the parametric distributional assumptions. Hence it is necessary to test whether they are consistent with the data.

3.2 Econometric Strategy

Several statistical methods have been developed to estimate qualitative choice models in which the dependent variable is a discrete choice among a set of alternatives. In binary choice situations, the probit and logit models are common methods. If the set of alternatives contains more than two choices, the multinomial logit is the most widely used estimation procedure due to its computational simplicity (McFadden, 1973). However, it relies on the
assumption that the unobserved terms are independent and homoscedastic, and this is not always an accurate approximation of reality. The formal implication of this assumption, also known as the property of the Independence of Irrelevant Alternatives (IIA), is that the ratio between two probabilities is necessarily the same regardless of the other alternatives available in the decision set \( J_i \). Alternatively one can relax this property by applying the multinomial probit assuming jointly normally distributed errors (Hausman and Wise, 1978). This model allows the unobserved terms to be correlated and heteroscedastic and it estimates their corresponding covariance matrix. Keane (1992), however, shows that it can be difficult to identify the coefficients in the multinomial probit if there are no exclusion restrictions (i.e. some variables entering the utility function of one alternative that do not enter the utility function of other alternatives). The identification problem arises as movements in the model coefficients, the \( \beta_0s \), can replicate changes in the elements of the covariance matrix of the unobserved terms.

Although the multinomial logit and probit are popular in applied econometrics, the estimated coefficients are inconsistent if the assumptions regarding the distribution of the unobserved terms are not satisfied. In this paper, I propose instead to use semiparametric techniques that allow the model to be consistently estimated without specifying the distribution of the unobserved terms.

Various semiparametric estimators have been suggested for binary choice models. Klein and Spady (1993) develop a procedure to estimate binary choice single index models by maximizing a semiparametric likelihood function in which the choice probability in the objective function is replaced with the probability obtained by kernel estimators. For cases with more than two alternatives, Lee (1995) generalizes Klein and Spady's methodology and proposes a multiple index model in which the probability of choosing each alternative is an unknown function of \( m \) indices. As pointed out by Ichimura and Lee (1991), identification of the coefficients when multiple indices are involved in the estimation requires that each index contains at least a variable which is not contained in other indices. This identification
condition places limits on the type of models that can be estimated by the methodology proposed in Lee (1995). For instance, the multiple choice model in equation (2) cannot be estimated because the explanatory variables are specific to the decision maker and they do not vary across alternatives.

Alternatively Klein and Vella (2004) show that when only two indices are involved in the estimation, the choice probability functions can still be identified without imposing exclusion restrictions on the indices. In this paper, I extend the double index based estimator for binary choice models in Klein and Vella (2004) to the case in which the decision set contains three alternatives. This is possible as the multiple choice model in equation (2) has a double index structure when only three alternatives enter the decision set. To see this, write the extended form of this model:

\[
P(d_{l1} = 1 | X_i) = \text{Prob}[(X_i\beta_1 + \epsilon_{i1} > X_i\beta_2 + \epsilon_{i2}) \cap (X_i\beta_1 + \epsilon_{i1} > X_i\beta_3 + \epsilon_{i3})],
\]

\[
P(d_{l2} = 1 | X_i) = \text{Prob}[(X_i\beta_2 + \epsilon_{i2} > X_i\beta_1 + \epsilon_{i1}) \cap (X_i\beta_2 + \epsilon_{i2} > X_i\beta_3 + \epsilon_{i3})],
\]

\[
P(d_{l3} = 1 | X_i) = \text{Prob}[(X_i\beta_3 + \epsilon_{i3} > X_i\beta_2 + \epsilon_{i2}) \cap (X_i\beta_3 + \epsilon_{i3} > X_i\beta_1 + \epsilon_{i1})].
\] (3)

To identify the model one needs to normalize the utility associated to one alternative to zero; otherwise the choice of the decision maker is not affected by a multiplicative transformation of the utility. This normalization is equivalent to subtract the utility of one alternative, for instance \(U_{1i}\), from each level of utility. Then, \(P(d_{li} = 1)\) is the bivariate distribution function for the unobserved terms \(\epsilon_{2i} = \epsilon_{2i} - \epsilon_{1i}\) and \(\epsilon_{3i} = \epsilon_{3i} - \epsilon_{1i}\). Under the assumption that the unobserved term \(\epsilon_{li}\) only depends on \(X\) through the indices (index restrictions), the conditional choice probabilities in (3) can be represented as follows:

\[
P(d_{l1} = 1 | X_i) = P_1(X_i\beta_2, X_i\beta_3),
\]

\[
P(d_{l2} = 1 | X_i) = P_2(X_i\beta_2, X_i\beta_3),
\]

\[
P(d_{l3} = 1 | X_i) = P_3(X_i\beta_2, X_i\beta_3),
\] (4)

where \(\beta_2 = \beta_2 i \beta_1\) and \(\beta_3 = \beta_3 i \beta_1\).
In equation (4), the probability of choosing each alternative, \( P_l(\ldots) \), is a function of two indices. The first index, \( I_1 = X_i \beta_2^2 \), is a linear combination of the explanatory variables in the model and it captures the contribution of \( X_i \) on the level of representative utility for alternative 2 compared to its contribution on alternative 1. The second index, \( I_2 = X_i \beta_3^2 \), does the same for alternative 3 compared to alternative 1. The function \( P_l(\ldots) \) represents the way in which these two indices combine. Without additional assumptions on the distribution of the \( e_0 \)s, \( P_l(\ldots) \) is unknown. The lack of a priori information on the functional form of \( P_l(\ldots) \) prevents identification of the original model coefficients. Ichimura and Lee (1991) note that in a multiple index model an intercept cannot be identified in either index and that the slope coefficients can only be identified up to a scale factor. Accordingly, in estimation one coefficient in each index needs to be set equal to 1. In addition, to achieve identification each index must contain at least one continuous variable which is not contained in other indices (i.e. \( X_i \not\subset X_L \)). However, Klein and Vella (2004) prove that in double index model exclusion restrictions are not required to identify the choice probability functions. They show that for an appropriate transformation of the coefficients the probability of a particular event conditional on two indices equals the probability conditional on two indices in which each index leaves one of the variables out. As a result, the transformed double index model satisfies the identification condition in Ichimura and Lee (1991).\(^4\) That is:

\[
P(d_{ij} | X_i \beta_2^2, X_i \beta_3^2) = P(d_{ij} | X_i \eta_2, X_i \eta_3),
\]

(5)

It should be emphasized that the original coefficients in the model, the \( \beta_0 \)s, cannot be recovered. However, it still possible to estimate those functions of the original coefficients, the \( \eta_0 \)s, that allow the choice probability functions to be identified. This does not represent a disadvantage for the estimator as in qualitative choice models one is generally concerned with the response probabilities and the marginal effects rather than with the model coefficients.

\(^4\)For a formal proof see Klein and Vella (2004).
The choice probabilities \( P(d_{li}|X_i, \eta_2, \eta_3) \) are estimated by a nonparametric kernel regression function:

\[
\hat{P}(d_{li}|X_i, \bar{\eta}_2, \bar{\eta}_3) = \frac{1}{N} \sum_{j=1}^{N} d_{lj} K \left( \frac{x_i, \bar{\eta}_2, x_i, \bar{\eta}_3 - x_j, \bar{\eta}_2, x_j, \bar{\eta}_3}{h} \right),
\]

(6)

where \( \Pr[d_l = 1] \) is the unconditional probability of \( d_l = 1 \) (i.e. the sample proportion of individuals making the choice in question). The numerator in equation (6) is an estimator of the joint density of both indices \( (X_i, \bar{\eta}_2, X_i, \bar{\eta}_3) \) conditional on \( d_l = 1 \), and the denominator estimates the unconditional density of these two indices. \( K(.) \) is a normal kernel function of \( R^2 \) and \( h \) is the window width parameter.\(^5\)

Using the predicted probabilities, the \( \eta_i^0 \)s are estimated by choosing the \( \eta_i^0 \)s that maximize:

\[
\ln(L) = \sum_{i=1}^{N} \sum_{l=1}^{3} \tau_i \left[ d_{li} \ln(\hat{P}(d_{li}|X_i, \bar{\eta}_2, \bar{\eta}_3)) \right] \quad \text{for } l = 1, 2, 3.
\]

(7)

In estimating semiparametric models trimming is required when the density of the data is low at these observations. Then, I introduce a trimming function, \( \tau_i \), that places zero weight on observations below 5% and above 95% on the basis of the distribution of \( X \).\(^6\)

The estimator here is semiparametric in that it makes no assumptions on the form of the distribution generating the disturbances; it does, however, assume that the choice probability function depends on parametrically specified index functions and that the unobserved terms in the model satisfy \( E[e_j|X] = 0 \).

\(^5\)Klein and Vella (2004) show that using local smoothing (as opposed to using higher order kernels) as bias reduction technique improves the sample performance of the estimator. With local smoothing, the window width in the final estimation of \( \hat{P}(.) \) varies in each observation and depends on a pilot density estimator. Klei and Vella (2004) employ bivariate kernels that depend on an estimated sample covariance matrix. They orthogonalize the column vectors \( I_1 \) and \( I_2 \) and then estimate their joint density estimator as the product of two independent kernels. The same methodology is employed in this paper.

\(^6\)A set of Monte Carlo experiments have been performed in order to evaluate the finite sample performance of the semiparametric estimator and the results are available upon request.
4 The Data

This investigation combines two different data sets: information on individual schooling choices and socioeconomic characteristics are from the National Longitudinal Survey of Youth 1979 (NLSY79), and expected earnings in each schooling alternative are estimated from the Current Population Survey (CPS, 1979-2000).

4.1 The Descriptive Statistics (NLSY 79)

The NLSY 79 includes a randomly chosen sample of 6,111 young U.S. citizens and two additional samples: the first containing information on 5,296 black, hispanic, as well as non-black and non-hispanic economically disadvantaged young people, and the other containing information on 1,280 individuals serving in the military as of September 30, 1978. I restrict the analysis to individuals in the first two samples, as the individuals in the third sample are older than the average and left high school before 1979. For this reason, the information on schooling paths directly after high school graduation is not available for this group.

To capture the effect of labor market conditions on schooling decisions, one must examine the short span of time between the date that the student finishes high school, and the date when the human capital investment decision is made. The NLSY 79 only provides monthly information on schooling attendance from 1981 onwards. Since 1979, however, it contains information on the respondent’s school enrollment status as of May 1 of each survey year. This information, corresponding to the year after high school graduation, is then used to identify the human capital investment decision of each student. It should be noted that students who graduated before 1977 are excluded from the sample as their enrollment status the year after high school graduation is unavailable. After 1983 only 91 students remained at high school and I have excluded them from the sample, as this number is small compared to the number of students enrolled in high school in previous years (5,742 in 1979, 4,198 in 1980, 2,688 in 1981, 1,440 in 1982 and 368 in 1983). The investigation is then restricted to
individuals who obtained a high school diploma between 1978 and 1983. This represents approximately 43% of the total NLSY 79 sample.

This paper analyses the factors that influence human capital accumulation upon high school graduation. It has been documented that more able and motivated people progress to higher grades (Cameron and Heckman, 1998). It would then be necessary to account for the selective nature of high school graduates. In the data for this analysis 17 percent of the students left school before high school graduation. However, accounting for sample selection induced by only looking at high school graduates would greatly complicate the estimation procedure, in addition to imposing greater data demands in terms of requiring instruments. Given that I do not correct the sample selection bias in estimation, the results cannot be used to make inference for the whole population, but only for those who have a high school diploma.

The dependent variable in the model is an indicator of whether the student, after graduating from high school, decides to attend a four-year college, a two-year college or not to enroll in college at all. Some issues should be taken into account when defining the dependent variable. Firstly, a two-year college program is a stepping stone for some students into a four-year institution afterwards. Secondly, in certain states, administrative regulations require a two-year college diploma before attending a four-year college. Therefore, to avoid the risk of misclassification and to simplify the dynamics of the model, a student is only considered to be enrolled in a two-year college if he/she enrolled in a two-year college after his/her graduation from high school, and if he/she was not enrolled in a four-year college three years after. For the same reasons, a student who started in a two-year college and then continued his/her education in a four-year college, is considered to have entered a four-year college after high school graduation. Given this classification, factors that affect enroll-

---

7The dropout rate has been calculated as the percentage of students born during the same years as those who graduated from high school between 1978 and 1983, and who were not enrolled in high school and had completed less than 12th grade in 1984.
ment in two-year college may also have an effect on the probability of attending a four-year college. This should be taken into account when interpreting the empirical results.

To estimate the effects of family background on the decision to attend college, the empirical model includes the following explanatory variables: the level of education of the student's parents and the number of siblings. If a respondent does not report information on any of these variables, he/she is excluded from the sample. The level of student's observed ability or IQ is measured as a weighted sum of four tests (on reading and numeric skills) of the Armed Services Vocational Aptitude Battery (ASVAB). The tests were administered in the summer of 1980 and as the age of the students varies, it is necessary to remove the age effect from this measure.\(^8\) Approximately 1.75% of the sample provides incorrect or missing information to compute the IQ, and therefore it is not included in the analysis. Observations below the 1st and above the 99th percentile of the IQ distribution are also excluded as they are considered to be outliers. As a result of the data selection process, there are 2127 individuals who satisfy the criteria. They represent approximately 40% of the students who graduated from high school between 1979 and 1983 in the NLSY 79.

If family income is included as an explanatory variable, it affects the estimation of the coefficients for parental education and student's observed ability due to the high correlation between family income and these variables. Parental education is a good predictor of family income and several authors have considered that the level of a student's ability is strongly determined by the level of family resources (Ellwood and Kane, 2000). Therefore, I do not include family income in the model but I leave for future research testing the stability of the parameters in different parts of the family income distribution.

Table 1 describes the profiles of the individuals in the sample according to their choice after high school graduation. From them, 1195 students (56.2 percent of the sample) did not enroll in college. Another 694 (32.6 percent of the sample) chose to start in a four-year

\(^8\)This is done by regressing the IQ score on a set of age dummies and subtracting the estimated age effects from the observed IQ score.
college and 238 (11.2 percent of the sample) started in a two year college. The sum of the two enrollment rates in the sample is slightly below the college enrollment rate for high school graduates in the population in these years (approximately 50 percent).

The statistics in Table 1 indicate that females are more likely to attend college and, in particular, two-year institutions. If analyzed by race, hispanics are overrepresented in two-year institutions and blacks represent a significant proportion of those who choose not to go to college. The descriptive statistics also indicate that students who attend a four-year college come from stronger family backgrounds and have a higher level of measured ability. In addition, the family income, the parents’ education and the IQ scores of students enrolled in two-year institutions suggest that these institutions are a middle ground between not attending college at all or going to a four-year college.

4.2 The Costs and Returns to Education

Because of individual heterogeneity and uncertainty about future labor market conditions, students cannot predict an important component of their ex post earnings. In this paper, I assume that students form expectations about the gains from schooling on the basis of their observable characteristics and the level of wages in their regional economies. Accordingly, I estimate expected earnings by stratifying the data in the CPS into cells on the basis of a set of observable characteristics, and calculate the average wage for each alternative in the cell. Cells are defined for particular values including race (black, hispanic, non-black and non-hispanic), region of residence (Northeast, North Central, West and South), urban area and gender. There are 48 cells for each year in the CPS. The average cell size for high school graduates who did not attend college is approximately 1,500 observations, 250 observations for those who graduated from two-year college programs and 450 for those with a four-year college diploma. Previous studies indicate that earning profiles are strongly ordered by education levels and concave, increasing rapidly over the early career, and leveling off as
they approach 30 years of experience, with a slight decrease towards the end of the career. This evidence suggests that the most important changes in wage-experience profiles occur early in the life cycle. Therefore, I estimate expected earnings using wage information for workers who are between 25 and 40 years old.

To include expected earnings in the individual decision rule, I merge them with the individual information in the NLSY 79, using the cell's definition. Including earnings at the cell level in the educational decision process only makes sense if information at this level is relevant for students. It seems quite unlikely that educational choices are dominated by expectations relating to the cell in which one might possibly work after graduation (i.e. a student growing up in a rural area in the South of the US who anticipates earnings consequences in an urban area on the West Coast). While this cannot be ruled out, I assume that the economic environment in which the student graduates from high school is the main source of information used to calculate expected earnings.

This study tests two alternative models of expectations about the returns to education. In the first model, students have myopic expectations and use the current wage gap between college and high school graduates to gauge the size of their own returns to schooling. If this information is relevant for students’ decisions, the economic environment at the time they graduate from high school will affect enrollment. Under this form of expectations, expected earnings for each educational option are estimated as the sum of the present value of average hourly wages for different age groups (25-29, 30-34 and 35-40) in the student’s cell at the year of high school graduation.

The second model assumes perfect foresight. Under this specification, students anticipate future demand and supply conditions and their effects on future wages. For instance, an increase in the demand for college graduates will have two effects on enrollment decisions: the first one is that the present value of the expected earnings increases, and the second one is that the expected earnings at the current age increase. This study tests two alternative models of expectations about the returns to education. In the first model, students have myopic expectations and use the current wage gap between college and high school graduates to gauge the size of their own returns to schooling. If this information is relevant for students’ decisions, the economic environment at the time they graduate from high school will affect enrollment. Under this form of expectations, expected earnings for each educational option are estimated as the sum of the present value of average hourly wages for different age groups (25-29, 30-34 and 35-40) in the student’s cell at the year of high school graduation.

The second model assumes perfect foresight. Under this specification, students anticipate future demand and supply conditions and their effects on future wages. For instance, an increase in the demand for college graduates will have two effects on enrollment decisions: the first one is that the present value of the expected earnings increases, and the second one is that the expected earnings at the current age increase.

9See for instance Murphy and Welch (1992).
11A discount rate of 5% is used to generate the earning streams. Alternative choices of the discount factor does not affect the main findings about the importance of earnings on schooling choices.
sions: a direct positive effect due to higher expected wages, and an indirect negative effect as students anticipate that more college entry will depress wages in the future. Although it is likely that the net effect of the demand shock is positive, forward-looking students react slower to changes in economic prospects (Ryoo and Rosen, 2004). In this context, expected earnings are estimated as the sum of the present value of future average hourly wages for different age groups. As the students in the NLSY 79 were approximately 20 years old in 1980, wages for these students at the age of 25 are predicted using the wages for 25 years old workers in 1985 in each student’s cell at the age of 25; wages at the age of 26 are predicted using wages for workers who are 26 years old in 1986 and so on.

Table 2 displays the present value of expected earnings for each educational alternative and for each model of expectations. The mean and the standard deviation of expected earnings in both models are smaller for high school graduates not enrolled in college, as tends to be true in other samples. However, the expected earnings as a four-year college graduate for those who did not attend college, or for those who attended a two-year college were larger than the earnings for the options they chose. This suggests the existence of borrowing constraints or a high level of uncertainty regarding the returns to education (i.e. in both models, the standard deviation of expected earnings for four-year college graduates is larger than for the other two alternatives). Note also that educational earnings differentials are larger when expected earnings are estimated using wages in the late 1980’s and the 1990’s (perfect foresight) than when they are estimated using wages in the early 1980’s (myopic). This results from the expansion in earnings differentials by education groups during the 1980’s, which only stabilized in the 1990’s.

The aggregate unemployment rate during the year that students graduated from high school is also included in the empirical model to capture the opportunity cost of attending college.\footnote{Imbens and Lynch (1993) find that not only the aggregate unemployment rate but also the local rate of unemployment, affects labor market opportunities for youth. In future work, I plan to include this variable in the model as it is a more accurate measure of the individual opportunity cost of college.}
5 Empirical Results

This section discusses the estimates based on the multiple choice model of human capital investment introduced in section 3.1. In this model, the response variable is an indicator of the student’s enrollment decision and is denoted by $d_{li}$ ($l = 1$ if no college, $l = 2$ if two-year college, $l = 3$ if four-year college). The explanatory variables are a set of individual characteristics that are expected to influence human capital investment decisions, as well as the expected earnings differentials by educational alternatives and the opportunity cost of college. That is:

$$d_{li} = \begin{cases} 1 : U_{li} > U_{Ll} & \text{for all } L \text{ in } J_i \text{ such that } L \notin l, \\ 0 : \text{Otherwise} \end{cases}$$

$$U_{li} = X_i \beta_l + \varepsilon_{li} = \beta_{10} + \beta_{11} \times IQ_i + \beta_{12} \times Siblings_i + \beta_{13} \times Gender_i + \beta_{14} \times Black_i + \beta_{15} \times Hispanic_i + \beta_{16} \times Mother's \ years \ of \ education_i + \beta_{17} \times Father's \ years \ of \ education_i + \beta_{18} \times Unemployment \ rate_i + \beta_{19} \times (Wages \ (4Y) - Wages \ (HS))_i + \beta_{110} \times (Wages \ (4Y) - Wages \ (2Y))_i + \varepsilon_{li}$$

Since only the systematic component of utility ($X_i \beta_l$) and the choice of the student ($d_{li}$) are observed, decisions can only be probabilistically estimated. I employ the semiparametric double index based estimator introduced in section 3.2 to estimate the probability of a student choosing each educational option in the set of available alternatives. Table 3 reports the estimated model coefficients.\textsuperscript{13} As only relative utility matters for individual decisions, the coefficients associated with the first alternative (no college) are normalized to zero. In Table 3, the column named "first index" reports the estimated coefficients for the second alternative (two-year college) with respect to the first alternative, and the column named "second index" shows the estimates for the third alternative (four-year college) with respect to the first one. These estimates capture the contribution or the weight of

\textsuperscript{13}The programs used for estimation are adapted from those used in the simulation evidence and empirical example in Klein and Vella (2004). I acknowledge the role of Roger Klein and Frank Vella in writing these programs.
each explanatory variable to the value of each index. Given the double index structure of the model, the final effect of each variable on the choice probabilities depends upon the flexible interaction between the indices. For estimation purposes, the model needs to be transformed to an equivalent one under a nonsingular linear transformation so as to induce exclusion restrictions. Following Klein and Vella (2004), an equivalent model can be obtained by normalizing the constant term to zero and one of the coefficients in each index to one. Accordingly, I normalize the coefficient for "Siblings" in the first index and that for "IQ" in the second to one. To achieve identification, I also normalized the coefficient for "IQ" in the first index and that for "Siblings" in the second to zero. This transformation allows me to estimate those functions of the original coefficients that suffice to identify the semiparametric choice probabilities, and from these compute the marginal effects.

The model is also estimated by standard multiple choice methods such as the multinomial logit and an identity multinomial probit where the covariance matrix is restricted to the identity matrix. These estimates are likely to be inconsistent for a number of reasons, for instance if the assumptions regarding the unobserved terms in the model do not hold. I perform a number of tests to assess the validity of these assumptions. A Hausman test rejects the Independence of Irrelevant Alternatives in the multinomial logit. The results in Table 4 indicate that the estimated coefficients change systematically when one of the alternatives is excluded from the choice set, and therefore the multinomial logit is not appropriate to describe the data. To test the normality assumptions in the multinomial probit, I adapt the conditional moment test for normality in Pagan and Vella (1989) (see Appendix). This procedure relies on testing whether the marginal distribution of the unobserved term ($\varepsilon_1$) is normal. Table 5 shows the results of the test. The coefficients in this table correspond to the constant term in a linear regression of each of the moment conditions, which need to be satisfied under the hypothesis of normality, on the scores of the likelihood function for the identity multinomial probit. These results strongly reject the assumption that ($\varepsilon_2$) and ($\varepsilon_3$) have a normal distribution. Therefore, by the properties of the multivariate nor-
mal distribution, it is not possible that the unobserved terms in the model are distributed jointly normal. Thus the multinomial probit is not appropriate to describe the data, and enrollment decisions are analyzed using the estimates by the semiparametric double index estimator.

To interpret the results, I first predict a "base" probability for each individual using their original values for the explanatory variables and the estimated model coefficients displayed in Table 3. The first row in Table 6 reports the average predicted probability of each choice. Then I change the variable of interest by a specific magnitude. For continuous variables, I add one standard deviation to each individual's original value and calculate a "new" probability conditional on this value. The rest of the rows in Table 6 report the average "new" probabilities and the difference between each row and the first one is the average marginal effect. For dummy variables, I estimate the likelihood of each individual choosing an alternative with and without each of the explanatory variables and report the mean probabilities.

Let us consider first the effect of socioeconomic variables, race and measured ability. On the whole, the results in other studies are confirmed. Ceteris paribus, females are more likely than males to attend college. The estimated probability of four-year college enrollment is 0.3347 for females and 0.3067 for males. For two-year college, this probability is 0.1212 for females and 0.1070 for males. These results are consistent with the aggregate trends in enrollment by gender described in Card and Lemieux (2001a). They document that female enrollment presents an increasing trend since the early 1970's, while enrollment for males shows a discrete downward trend shift in the 1970's which is only restored in the mid-1980's.

Family background characteristics are also important determinants of individual schooling outcomes. A one standard deviation increase in the number of living siblings in the family decreases the probability of attending both types of college (from 0.1142 to 0.1020

---

14 For the sake of computation the explanatory variables in the model are standardized to have a mean of zero and a standard deviation of one.
for two-year college and from 0.3209 to 0.3104 for four-year college). This is evidence in favor of the economic view of the quality-quantity trade-off when parents decide on their number of children. There is also evidence that children from better-educated parents have a higher probability of beginning their post-secondary education in a four-year institution. The mean probability of attending a four-year college increases by 20 percent (from 0.3209 to 0.3852) when the mother’s years of education increase by one standard deviation, and by 13 percent (from 0.3209 to 0.3621) when the father’s years of education do the same. While increasing the explanatory variable by a one standard deviation is a common practice to compute the marginal effect, one must consider the size of this statistic in the raw data before interpreting the magnitude of the effect. For parental education, the descriptive statistics in Table 1 indicate that the standard deviation for years of father’s education is larger than that for mother’s education. Accordingly, the size of the marginal effects suggests that mother’s years of education have a larger positive effect on the probability of enter a four-year college. The evidence here highlights the importance of long term family background factors to college attendance decisions. Parental education is a proxy of permanent family income that captures the ability of parents to invest in their children’s education, and to produce ability and college readiness for their children from birth onwards.

Students with a higher IQ score also have a higher probability of beginning in a four-year college. This variable can be interpreted as a measure of academic skills and preparedness for college as a one standard deviation increase in the IQ score increases the probability of attending a four-year college by 0.18 percentage points (from 0.3209 to 0.4993). It should be noted that parental education and the level of measured ability are less important in determining two-year college entrance. As in Rouse (1994), this suggests that community colleges offer the possibility to participate in post-secondary education to students from different socioeconomic backgrounds and different levels of academic preparation.

According to the results in Table 6, there is a difference of 0.08 percentage points in the likelihood that blacks and non-blacks will begin in a four-year college and a difference of
0.11 that hispanics and non-hispanics will enter a two-year institution. This is consistent with the results in Cameron and Heckman (2001). They find that conditional on family background, minorities are more likely than whites to attend college. However, the large race effect here suggests that additional explanations such as the college system in the place of residence may be driving part of this result. In the US, there are important differences at the state level in terms of availability of educational choices. Cameron and Heckman (2001) find that the higher two-year entry rate of hispanics is attributable to the regional concentration of hispanics in states such as California and Texas, with extensive low-tuition community college networks.15

To estimate the effects of labor market conditions on schooling the empirical model includes as explanatory variables the aggregate unemployment rate and the expected earnings in each educational option. The results in Table 6 correspond to the human capital investment model in which expected earnings are estimated using current wages in the student’s cell at the time of high school graduation. These results indicate that variations in economic conditions affect the enrollment probabilities. In line with Della and Koubi (2003), I find that students are more likely to attend school in periods of high unemployment. More precisely, a one standard deviation increase in the unemployment rate increases the probability of beginning in a four-year college by almost 6 percent (from 0.3209 to 0.3398). This result suggests that fewer job opportunities encourage students to continue with further education, and that short term liquidity constraints are not an obstacle to pursuing education during these periods. In addition, an increase in unemployment shifts enrollment from two-year to four-year colleges. The probability of attending a two-year college decreases by 3 percent (from 0.1142 to 0.1107) when the unemployment rate increases by one standard deviation. This can be interpreted as a response to the persistence of unemployment. That is, freshmen anticipate that the unemployment rate will not decrease immediately. If they

15In future research, when I will gain access to the confidential Geocode Data, it will be necessary to control for the college composition in each state.
enroll in a two-year institution, it might well be that employment opportunities will still be scarce after graduation, and therefore they decide to stay in college for a longer period.

The empirical results indicate that students respond positively to the gains from schooling. A one standard deviation increase in the expected earnings from four-year college increases the probability of beginning in this type of institution by approximately 7 percent (from 0.3209 to 0.3423). In the same way, a one standard deviation increase in the expected earnings from some college increases the probability of attending a two-year college by 7.79 percent (from 0.1142 to 0.1231). Given that the standard deviation of those two variables is similar (see Table 2), these findings indicate that two-year students pay slightly more attention to their anticipated rewards.

Changes in the expected returns to education also affect the composition of college enrollment. An increase in earnings for four-year college graduates not only increases the likelihood of choosing this alternative but it also decreases schooling dropouts (from 0.5649 to 0.5447) and the probability of enrollment in a two-year college (from 0.1142 to 0.1130). Although the magnitude of the effect on two-year college is small, its sign indicates that students make rational decisions and choose the alternative that maximizes expected earnings. An increase in earnings for two-year college also reduces the probability of schooling dropouts (from 0.5649 to 0.5471), but it increases the probability of four-year college enrollment (from 0.3209 to 0.3298). This last result does not necessarily contradict the previous conclusion, which suggested that students choose the alternative that maximizes their earnings, and it can be explained by the criteria used to identify investment decisions. As has been discussed, students who started in a two-year college and then continued their education in a four-year institution are considered to be enrolled in a four-year college upon high school graduation. It might be that some students started in a two-year college as a response to its higher returns, but then shifted to a four-year college to continue with further education.

The evidence here also indicates that an improvement in the level of wages for low
skilled workers diverts students from college. That is, a one standard deviation increase in expected earnings for workers with only a high school diploma decreases the probability of both two-year and four-year college enrollment (from 0.1142 to 0.1070 and from 0.3209 to 0.2911 respectively). Further research should examine the effects of labor market conditions on school dropout rates. It might be that an increase in forgone income only delays the decision to attend college. According to Light (1995), this is not an issue of concern since the wage gap between young workers who postpone their schooling and those who receive it "continuously" closes over time. In contrast, if students do not return to school, favorable labor market conditions for low skilled workers at the time of high school graduation will have a negative impact on their lifetime earnings.

So far, I have assumed that students have myopic expectations and base assessments of future career prospects only on current wages. Under this form of expectations, students significantly respond to variations in the expected returns to education. However, this expectation mechanism may seem too simplistic because it assumes that students do not have knowledge about the future developments of the market structure. Therefore, I also estimate the model including as an explanatory variable the present value of expected earnings estimated using wages some years after college graduation (see section 4.2). Table 7 shows the estimates for this alternative model of enrollment decisions. The socioeconomic and demographic estimates are similar to the results in Table 3, however the expected returns to education become statistically insignificant.

A possible explanation for this last result may be related to the substantial changes in the dynamics of the wage process during this period: increases in earnings gaps by education groups and a much greater increase in the dispersion of the idiosyncratic component of earnings. Then, realized future wages in the late 1980's and in the 1990's are weak proxies for students' expectations at the time they decide to go to college in the early 1980's (i.e. large expectational errors explain this result). More work must be done to identify the elements that belong to the information set of the students. However, the results in this
paper indicate that labor market conditions, which affect wages and job opportunities at the time students decide to go to college, are relevant for their investment decisions.

The evidence here also shed light on the potential explanations for the slow growth in US college enrollment rates in the 1980's and the 1990's. There is evidence that students respond to the expected gains from schooling, and that the earnings differentials by education groups increased during the 1980's (see Section 2). Then, one would have expected a higher college enrollment rate. However, the variance of the residuals or the unpredictable component of earnings increased steadily over the period. This implies an increase in the difficulty of predicting the returns to education, that might have discouraged students from investing in additional human capital. A key issue for future research is to quantify the effects of the increase in earnings uncertainty on enrollment decisions, and the welfare implications of the recent changes in the wage structure.

6 Conclusions

This paper examines the economic incentives and constraints that influence investment in higher education. I estimate a model of schooling decisions for a sample of students from the NLSY 79. In the empirical model, each student upon high school graduation faces a choice among a set of three alternatives: starting in two-year college; starting in four-year college; or not attending college at all. A number of tests reject the assumptions in standard multiple choice models (i.e. multinomial logit and probit) regarding the parametric distribution of the unobserved terms. Hence, I employ a semiparametric double index estimator, which allows a flexible functional form of the choice probability functions.

The empirical results are interesting on several fronts. I find that parental education and the level of student's observed ability strongly influence enrollment in four-year college, while they are less important for two-year college entry. This suggests that two-year colleges provide equality of opportunities. I also find that labor market conditions at the time
students make their decisions affect schooling outcomes. That is, higher wages for skilled workers encourage students to continue with further education; but also favorable labor market conditions for low skilled workers divert students from college or, at least, postpone their decision to accumulate human capital. Finally, there is evidence that during periods of high unemployment students are more likely to stay in college. These results indicate that education policies aimed at influencing enrollment trends should consider the interactions between economic conditions and enrollment behavior.

Further research should examine the contribution of the changes in the wage distribution to the trends in US college enrollment. In particular, it might be that the substantial increase in the dispersion of the idiosyncratic component of wages has discouraged students from pursuing higher education, explaining the slow growth in college enrollment over the last decades.
References


Figure 1: Tuition and Fees in Institutions of Higher Education and College Enrollment Rate

Source: U.S. Department of Education (October 2001)
Figure 2: Decomposition (Within and Between) of Std. Dev. of log wages

Source: CPS (1979-2000)
Figure 3: Unemployment rate and College Enrollment Rate

Table 1
Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>No College</th>
<th>Two-year College</th>
<th>Four-year College</th>
<th>All the Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>54.14</td>
<td>47.90</td>
<td>49.71</td>
<td>52.00</td>
</tr>
<tr>
<td>Black</td>
<td>25.69</td>
<td>25.21</td>
<td>25.50</td>
<td>25.58</td>
</tr>
<tr>
<td>Hispanic</td>
<td>11.63</td>
<td>22.27</td>
<td>7.78</td>
<td>11.57</td>
</tr>
<tr>
<td>IQ</td>
<td>57.40</td>
<td>64.28</td>
<td>75.41</td>
<td>64.05 (18.77)**</td>
</tr>
<tr>
<td>Family income*</td>
<td>15.72</td>
<td>16.99</td>
<td>21.13</td>
<td>17.62 (14.20)</td>
</tr>
<tr>
<td>Father’s education</td>
<td>10.41</td>
<td>11.40</td>
<td>13.03</td>
<td>11.38 (3.74)</td>
</tr>
<tr>
<td>Mother’s education</td>
<td>10.64</td>
<td>11.21</td>
<td>12.58</td>
<td>11.33 (2.93)</td>
</tr>
<tr>
<td>Siblings</td>
<td>3.80</td>
<td>3.10</td>
<td>2.90</td>
<td>3.43 (2.40)</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>7.25</td>
<td>7.23</td>
<td>7.31</td>
<td>7.27 (1.34)</td>
</tr>
<tr>
<td>No. Observations</td>
<td>1195</td>
<td>238</td>
<td>694</td>
<td>2127</td>
</tr>
</tbody>
</table>

*Family income is in thousands of dollars deflected by the CPI (base 1980)
**Standard Deviation in parenthesis
Table 2
Present Value of Expected Earnings

<table>
<thead>
<tr>
<th>Chosen Alternative</th>
<th>No College</th>
<th>Two-year College</th>
<th>Four-year College</th>
</tr>
</thead>
<tbody>
<tr>
<td>No College</td>
<td>18.54 (3.93)*</td>
<td>20.64 (4.28)</td>
<td>23.67 (4.71)</td>
</tr>
<tr>
<td>Two-year College</td>
<td>18.88 (4.15)</td>
<td>20.98 (4.12)</td>
<td>23.96 (4.61)</td>
</tr>
<tr>
<td>Four-year College</td>
<td>18.90 (4.12)</td>
<td>21.08 (4.16)</td>
<td>24.18 (4.63)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chosen Alternative</th>
<th>No College</th>
<th>Two-year College</th>
<th>Four-year College</th>
</tr>
</thead>
<tbody>
<tr>
<td>No College</td>
<td>24.92 (4.84)</td>
<td>28.75 (5.26)</td>
<td>38.95 (6.73)</td>
</tr>
<tr>
<td>Two-year College</td>
<td>25.61 (4.67)</td>
<td>29.64 (5.04)</td>
<td>39.55 (6.23)</td>
</tr>
<tr>
<td>Four-year College</td>
<td>25.62 (5.10)</td>
<td>29.55 (5.55)</td>
<td>40.13 (7.18)</td>
</tr>
</tbody>
</table>

*Standard deviation in parenthesis
Table 3
Estimates of Enrollment Decisions

<table>
<thead>
<tr>
<th></th>
<th>First Index</th>
<th>Second Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siblings</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>IQ</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Gender</td>
<td>0.5390 (3.4266)</td>
<td>0.0591 (1.3282)</td>
</tr>
<tr>
<td>Black</td>
<td>0.8601 (5.1619)</td>
<td>0.4897 (9.2482)</td>
</tr>
<tr>
<td>Hispanics</td>
<td>1.4549 (4.1664)</td>
<td>0.0799 (0.7169)</td>
</tr>
<tr>
<td>Father’s education</td>
<td>0.1336 (0.5941)</td>
<td>0.2345 (3.5925)</td>
</tr>
<tr>
<td>Mother’s education</td>
<td>0.1957 (0.5689)</td>
<td>0.3781 (4.9736)</td>
</tr>
<tr>
<td>W(4-year)-W(No college)</td>
<td>0.4547 (1.7267)</td>
<td>0.1548 (1.9984)</td>
</tr>
<tr>
<td>W(4-year)-W(2-Year)</td>
<td>0.6232 (2.2045)</td>
<td>0.0211 (0.2536)</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>0.3647 (1.8181)</td>
<td>0.1294 (2.6836)</td>
</tr>
<tr>
<td>In(likelihood function)</td>
<td>1.60733</td>
<td></td>
</tr>
</tbody>
</table>

$t$-statistics in parenthesis
Table 4

<table>
<thead>
<tr>
<th>Excluded Category</th>
<th>Base Category</th>
<th>4-year College</th>
<th>2-year College</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-year College</td>
<td>(3.78)(^*)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No College</td>
<td>(79.75)</td>
<td>(0.53)</td>
<td></td>
</tr>
</tbody>
</table>

\(^*\)Chi-squared statistic

Table 5

<table>
<thead>
<tr>
<th>( \varepsilon_{ij} )</th>
<th>( \alpha_1 )</th>
<th>( t)-statistic</th>
<th>( \alpha_2 )</th>
<th>( t)-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \varepsilon_{i2} )</td>
<td>0.0098</td>
<td>(4.4247)</td>
<td>0.0364</td>
<td>(4.0581)</td>
</tr>
<tr>
<td>( \varepsilon_{i3} )</td>
<td>0.0239</td>
<td>(2.1682)</td>
<td>0.0835</td>
<td>(1.2033)</td>
</tr>
</tbody>
</table>
Table 6
Changes in the Probability of Attending Two- or Four-Year College

<table>
<thead>
<tr>
<th></th>
<th>Four-Year</th>
<th>Two-Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated Mean</td>
<td>0.3209</td>
<td>0.1142</td>
</tr>
<tr>
<td>Male</td>
<td>0.3067</td>
<td>0.1070</td>
</tr>
<tr>
<td>Female</td>
<td>0.3347</td>
<td>0.1212</td>
</tr>
<tr>
<td>Black</td>
<td>0.4112</td>
<td>0.1205</td>
</tr>
<tr>
<td>Non-black</td>
<td>0.3273</td>
<td>0.1107</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.3469</td>
<td>0.1379</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>0.3226</td>
<td>0.1081</td>
</tr>
<tr>
<td>Mother’s years educ.</td>
<td>0.3852</td>
<td>0.1137</td>
</tr>
<tr>
<td>Father’s years educ.</td>
<td>0.3621</td>
<td>0.1171</td>
</tr>
<tr>
<td>Siblings</td>
<td>0.3104</td>
<td>0.1020</td>
</tr>
<tr>
<td>IQ</td>
<td>0.4993</td>
<td>0.1129</td>
</tr>
<tr>
<td>ln(W(4-year))</td>
<td>0.3423</td>
<td>0.1130</td>
</tr>
<tr>
<td>ln(W(2-year))</td>
<td>0.3298</td>
<td>0.1231</td>
</tr>
<tr>
<td>ln(W(high school grad.))</td>
<td>0.2911</td>
<td>0.1070</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>0.3398</td>
<td>0.1107</td>
</tr>
</tbody>
</table>
Table 7
Estimates of Enrollment Decisions
(Perfect Foresight)

<table>
<thead>
<tr>
<th></th>
<th>First Index</th>
<th>Second Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siblings</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>IQ</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Gender</td>
<td>0.3506 (1.4576)</td>
<td>0.0565 (1.1929)</td>
</tr>
<tr>
<td>Black</td>
<td>1.2168 (3.0313)</td>
<td>0.5108 (6.0837)</td>
</tr>
<tr>
<td>Hispanics</td>
<td>2.0584 (3.7825)</td>
<td>0.2217 (1.3979)</td>
</tr>
<tr>
<td>Father’s education</td>
<td>0.5361 (1.1983)</td>
<td>0.2859 (3.1048)</td>
</tr>
<tr>
<td>Mother’s education</td>
<td>0.3860 (1.0533)</td>
<td>0.4286 (4.4681)</td>
</tr>
<tr>
<td>W(4-year)-W(No college)</td>
<td>1.0287 (1.2444)</td>
<td>0.0551 (0.2732)</td>
</tr>
<tr>
<td>W(4-year)-W(2-Year)</td>
<td>0.3387 (0.4531)</td>
<td>0.0685 (0.3370)</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>0.4642 (1.9051)</td>
<td>0.1405 (2.6795)</td>
</tr>
<tr>
<td>ln(likelihood function)</td>
<td>1.64798</td>
<td></td>
</tr>
</tbody>
</table>

*t* statistics in parenthesis
Normality Test: Multiple Choice Models

This section discusses how the conditional moment procedure in Pagan and Vella (1989) can be extended to test assumptions on the distribution of the unobserved terms in a multiple probit model. Let us .rst write the choice probability for each alternative as a function of the data, the parameters of the model and a set of parameters γ:

\[ P_1 = \Phi(U_{12}, U_{13}; r_1), \]
\[ P_2 = \Phi(\alpha U_{12}, U_{23}; r_2), \]
\[ P_3 = \Phi(\alpha U_{13}, \alpha U_{23}; r_3), \]

where \( U_{12} = X_i(\beta_2/\sigma_2), \)
\( U_{13} = X_i(\beta_3/\sigma_3), \)
\( U_{23} = (X_i(\beta_2/\sigma_2 + \beta_3/\sigma_3)^2 + \gamma_1(X_i(\beta_2/\sigma_2 + \beta_3/\sigma_3)^2 + \gamma_2(X_i(\beta_2/\sigma_2 + \beta_3/\sigma_3)^2 + X_i(\beta_3/\sigma_3 + \beta_3/\sigma_3)^2)), \)
\( \Phi() \) is the bivariate normal distribution for \((\varepsilon_{i2}, \varepsilon_{i3})\) and \( r_1 \) the correlation coefficient.

The sample moment conditions in the normality test for \((\varepsilon_{i2})\) are:

\[ A_1 = N^{i,1} \sum_{i=1}^{N} X_i^\gamma \mu \left[ \frac{d_{ji}}{P_j(X_i, \beta, \gamma_1 = 0, \gamma_2 = 0)} \right] \left[ \frac{\partial P_j(X_i, \beta, \gamma_1 = 0, \gamma_2 = 0)}{\partial \gamma_1} \right] = 0, \quad (A.1) \]

and

\[ A_2 = N^{i,1} \sum_{i=1}^{N} X_i^\gamma \mu \left[ \frac{d_{ji}}{P_j(X_i, \beta, \gamma_1 = 0, \gamma_2 = 0)} \right] \left[ \frac{\partial P_j(X_i, \beta, \gamma_1 = 0, \gamma_2 = 0)}{\partial \gamma_2} \right] = 0. \quad (A.2) \]

Note that \( A_1 \) and \( A_2 \) are exactly the scores of the log likelihood function with respect to \( \gamma_1 \) and \( \gamma_2 \), calculated at \( \gamma_1 = \gamma_2 = 0 \). Note also that for the multinomial probit, the derivatives of the \( P_j \) with respect to the \( \gamma \) parameters have a simple form comprising standardized normal densities and univariate distributions. That is:

\[ \frac{\partial P_j(X_i, \beta, \gamma_1 = 0, \gamma_2 = 0)}{\partial \gamma_1} = \phi(U_{12}) \Phi(\alpha U_{13}, \alpha U_{23}; / q \frac{1}{r_1^2}) (X_i)^2/2. \]

Similar moment conditions can be derived for \((\varepsilon_{i3})\). Due to the required normalizations, the properties of the marginal distribution for the unobserved term in the .rst alternative cannot be tested.

The test for normality is a simple \( t \) test of whether the constant term is zero in a linear regression of each of the moment conditions that need to be satis.ed if the model is correctly speci.ced on that constant and the scores of the likelihood function for the multinomial probit under the assumption of normality, which implies \( \gamma_1 = \gamma_2 = 0 \).