Demand for Post-compulsory Education: The Choice Between Academic and Vocational Tracks*

Cristina Lopez-Mayan†

CEMFI and UIMP

January 31, 2010

Abstract

I develop a discrete choice structural model of individual schooling decisions following compulsory education, to analyse the effect of expected life-time earnings. The model is estimated using Spanish data on educational histories. I study the two paths available to a Spanish young person considering post-compulsory education: academic high school and vocational high school. In my model the options open to an individual following completion of one of these two high school tracks are linked to their expected life-cycle earnings. I use the estimated model to assess how school attendance and dropout rates respond to changes in expected earnings. I find that the largest impact in decreasing the dropout rate in post-compulsory high school is associated with raising the annual wage corresponding to a vocational diploma. I also find that the Spanish wage structure in 2006 discourages post-compulsory education attendance. It also affects the qualification of high school graduates by reducing the share of workers with a vocational qualification.

Keywords: High school education, vocational education, expected life-cycle earnings, dropout rate, discrete choice structural model, NPL algorithm.

JEL codes: I21, I28, J24.

---

*I would like to thank Manuel Arellano, Samuel Bentolila, Pedro Mira and Victor Aguirregabiria for very helpful comments. Special thanks are also due to Dante Amengual, Laura Crespo, Laura Hospido and Michael Manove. I acknowledge financial support from Fundacion Ramon Areces. For the most recent version of the paper, please see: http://www.cemfi.es/∼clopezmayan/

†Contact: crislopezmaya@gmail.com

1 Introduction

This paper studies the effect of expected life-cycle earnings on schooling choices of Spanish youth following compulsory education. To accomplish this, I formulate and estimate a sequential decision model that accomodates the choices available to an individual after completing compulsory schooling. Each year, the individual chooses the option that maximizes her expected life-time utility. I use a rich microdataset on schooling histories elaborated by the Spanish Statistics Institute in 2005. This dataset has not been used for this purpose in previous research.

In the Spanish system, following compulsory schooling, the individual chooses between entering the labor market and attending high school education. In the latter, she chooses between two possible paths: vocational high school and academic high school. Individuals attend education to accumulate completed grades towards the diploma. Graduation is a probabilistic outcome because, when the individual decides what option to choose, she faces uncertainty about the probability of completing the grade. This probability depends on previous performance and on observed individual characteristics.

Once the individual completes academic or vocational high school, she has available a new set of choices that include the option of entering the labor market and the option of continuing in further education. In the latter case, the alternatives differ depending on the diploma obtained. I model the first decision following high school education because it is important in order to understand the preceding choices. For example, most individuals choose to attend academic high school because this diploma is required to enter university. In my model, the options available to an individual following completion of one of the two high school tracks are also linked to their expected life-cycle earnings.

The contributions of this paper to the literature on human capital investment decisions are twofold. This is the first time that a structural model on schooling decisions reflecting the different tracks (academic and vocational) available in post-compulsory education is estimated. In addition, as suggested by human capital theory, the model establishes the link between those decisions and the expected life-time earnings corresponding to each of the options available following completion of these tracks. This model is of interest because it allows to quantify whether changes in the expected life-time earnings generate changes in the track choice and/or in the dropout rate in (post-compulsory) high school education. This paper is the first study that provides empirical evidence on these aspects.

From a methodological perspective, I estimate the model using Aguirregabiria and Mira (2002)'s nested pseudo likelihood algorithm (NPL), which allows the researcher to estimate discrete choice structural models without having to compute repeated solutions of the dynamic problem. I assume that individuals finish compulsory schooling with differences in their preferences and ability to progress in each track that influence subsequent choices. I follow a finite-mixture (nonparametric) approach to model this permanent unobserved heterogeneity. This assumption implies that the support points of the finite-mixture distribution and the probability of each point are parameters to be estimated. This is the first time that the NPL algorithm is used to estimate a finite horizon model with permanent unobserved heterogeneity where those mass points and probabilities are estimated without imposing a distributional assumption[1].

The estimated model is used to assess how individuals respond to a change in expected life-

[1]Aguirregabiria and Mira (2007) implement the NPL in an infinite horizon model where they assume that the only variable that presents permanent unobserved heterogeneity follows a discretized version of a Gaussian variate with zero mean and constant variance. The support points are rescaled by the standard deviation and the distribution is fully characterized by only estimating a single parameter.

cycle earnings. I make two types of exercises. The first one analyses how a separate change in the life-time earnings corresponding to university, vocational college, academic high school, vocational high school and compulsory schooling, respectively, affects schooling outcomes. The main conclusion from these exercises is that the largest impact on the completion rate in upper secondary education is found after an increase in the earnings received by workers with a vocational qualification. Increasing the annual wage of vocational college by ten percent every year reduces the dropout rate by three percentage points. The decrease is uniquely driven by the lower dropout rate in the technical track. This higher completion rate is accompanied by an increase of fifteen percentage points in the proportion of individuals who decide to attend the vocational track instead of the academic one. Moreover, these exercises also show that a ten percent increase in the life-cycle wages corresponding to compulsory schooling discourages participation in post-compulsory education, reducing the attendance rate by six percentage points.

In the second type of exercises, I use the Spanish Wage Structure Survey of 2006 to show how individuals would behave if they faced that wage structure in the moment they make their schooling decisions. Except for university, the premium of each schooling level with respect to compulsory education decreased between 2002 and 2006. The model predicts that this wage structure will discourage the participation in post-compulsory education, reducing the proportion of individuals who attend the technical track. Additionally, it will increase by almost eight percentage points the dropout rate in this track. As a consequence, the proportion of individuals who have high school as their maximum education level will decrease. The evidence from the Spanish Labor Force Survey confirms this prediction. Among high school graduates, the share with a vocational diploma will be lower (from 19% in the baseline to 11%). So, there will be a change in the qualification of the high school graduates towards a reduction in the share of workers with a technical qualification.

The model of human capital investment assumes that the expected monetary reward drives the decision to invest in education. One issue that arises when testing the human capital model is that earnings are only observed after the schooling investment has been completed. When individuals make their education decisions, they face uncertainty about the wage they can earn from each choice. To deal with this issue, the literature uses two different approaches. The first one assumes that individuals have rational expectations. This assumption implies that youths know their potential earnings profile and use it to predict their expected earnings. One of the most representative study using this approach is Willis and Rosen (1979). The second approach, suggested by Manski (1993), considers that young people form their expectations about their potential earnings observing the incomes realized by members of the preceding generations.

In this paper, I adopt the last approach and the reason is twofold. First, I consider that young people do not form their expectations as the rational approach assumes. Instead, I consider they observe the wages of workers similar to them to predict the expected earnings associated to each education level. Second, the dataset is short to observe individuals completing further education following high school education and working in the labor market. As a consequence, I assume that youths observe the earnings of other individuals who are working in the moment they make their schooling decisions and use them, conditioned on some observed individual characteristics, to infer the life-cycle earnings of each schooling choice. The life-cycle profiles are constructed using earnings from the Spanish Wage Structure Survey 2002. If the individual decides to attend further education following high school completion, the corresponding life-time earnings are weighted by the probability of graduation.

There are other papers that use the structural framework to analyse schooling decisions in post-compulsory education but they only consider the choice between attending or dropping out of high school, without distinguishing the kind of track. For example, Eckstein and Wolpin (1999) construct a model of high school progression where individuals can work and attend school at the same time to analyse the effect of working on high school performance. Keane and Wolpin (1997) formulate a model of occupation and schooling decisions in which individuals decide among attending high school, staying at home and three occupation options.

Other papers, such as those by Arcidiacono (2004, 2005), analyse the decision of attending college. However, unlike the model I present, they use a sample of high school graduates and do not consider the link between the choice of attending college and the schooling decisions in the high school level. Keane and Wolpin (1997) considers in some way this link because, in their model, the individual can choose to attend education until she accumulates sixteen years of schooling and becomes a college graduate. However, they do not consider that there is uncertainty in the schooling progression and that the individual accumulates a complete grade just by attending school that year. Regarding this, I include uncertainty in the schooling progression following the approach of Eckstein and Wolpin (1999). In contrast, I model the first decision following high school completion while these authors assume that the value of the high school diploma embeds all the decisions about attending college.

There are other papers that analyse the decision between attending academic or vocational high school but using a reduced-form approach. However, they do not analyse the effect of the expected life-time earnings on those decisions. For example, Bradley and Lenton (2007) analyses the effect of individual's prior attainment, family background characteristics and state of the labor market on the decisions of attending low and high levels of academic and vocational high school and on the decision of entering the labor market. They find that individuals with the best prior attainment enter high academic courses and are less likely to drop out while the least qualified enroll in low level courses but are more likely to drop out. Using German data, Dustmann (2004) analyses the influence of parental background on the track that the children choose when they finish primary school. In the German system the individual can choose among three paths: two are the basis for apprenticeship training and the third is the basis for attending university. He finds that parental background is strongly related to the track chosen by the child and, given the rigidity of the German education system to move among tracks, this translates in substantial earnings differentials later in life.

The rest of the paper is organized as follows. The next section explains the Spanish education system. Section 3 presents the dataset and some descriptive statistics. Section 4 provides evidence on data correlations from reduced-form estimates. Section 5 presents the theoretical model. Section 6 explains how the expected life-time earnings and the current wages included in the model are computed. Section 7 explains the solution and estimation method. Section 8 shows the parameter estimates and the model fit, and Section 9 presents the exercises used to quantify the impact of expected earnings on schooling choices. Finally, Section 10 concludes.

2 The Spanish education system

In the last decades, there has been a debate in Spain about the need to reform the education system. This debate is motivated by the following facts:

- The Programme for the International Student Assessment (PISA) has shown that the performance of fifteen-year-old Spanish students is below the average of the OECD countries[2].

[2] In the PISA study (OECD, 2003), the scores of Spain in maths, reading, science and problem solving were 485, 481, 487 and 482 respectively, while the averages of the OECD countries were 500, 494, 500 and 500 respectively.

- There is a high proportion of individuals who drop out from both the compulsory and the upper secondary levels. As shown in Figure 1, these high dropout rates result in a high proportion of Spanish youths with compulsory schooling as their maximum educational level in comparison to other European countries. Only Malta and Portugal have higher percentages than Spain.

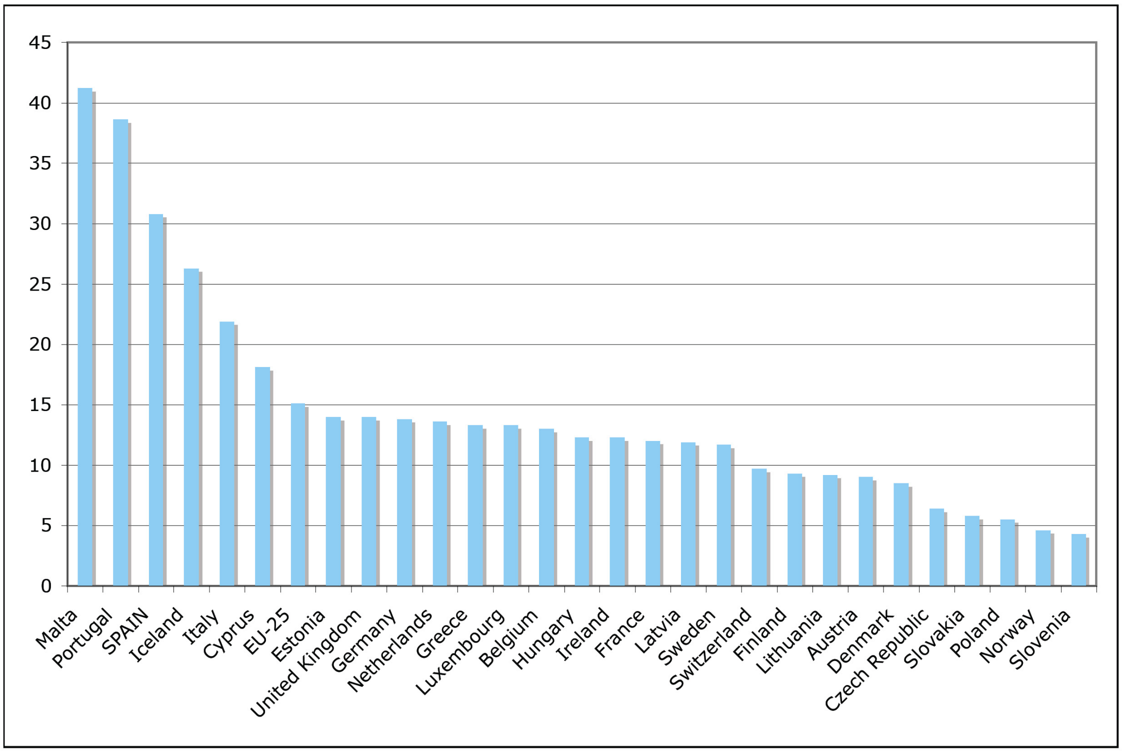

Figure 1: Percentage of people aged 18-24 years old with at most compulsory schooling and who have declared they are not receiving any education or training (2005). Source: Eurostat.

Of the 30.8% Spanish youths with at most compulsory schooling in 2005, 19% have only primary education (MEC, 2006). Thus, this group contains individuals who dropped out of compulsory education. The rest (81%) have compulsory schooling. This second category includes individuals who entered the labor market after finishing compulsory schooling and individuals who started in upper secondary education but dropped out.

Education in Spain is run by regional governments. However, the central government establishes some general aspects that have to be applied in all the territory of Spain: compulsory schooling age, schooling levels, basic content of the curriculum, number of pupils per class, etc.

Both compulsory and (post-compulsory) high school education are public and free although the individual can choose to attend education in private and semi-private schools (*colegios concertados*). Private schools are more expensive than semi-private ones because the latter receive government subsidies.

The education law relevant for the schooling decisions analyzed in this paper was the law passed in 1990[3] and that regulated the Spanish education system until 2006. One of the most

[3] *Ley Orgánica 1/1990 de Ordenación General del Sistema Educativo* (LOGSE).

important changes introduced by this law was the increase of the minimum compulsory schooling age from fourteen to sixteen years old.

Figure 2 represents the schooling levels. Compulsory schooling attendance covers ten years,

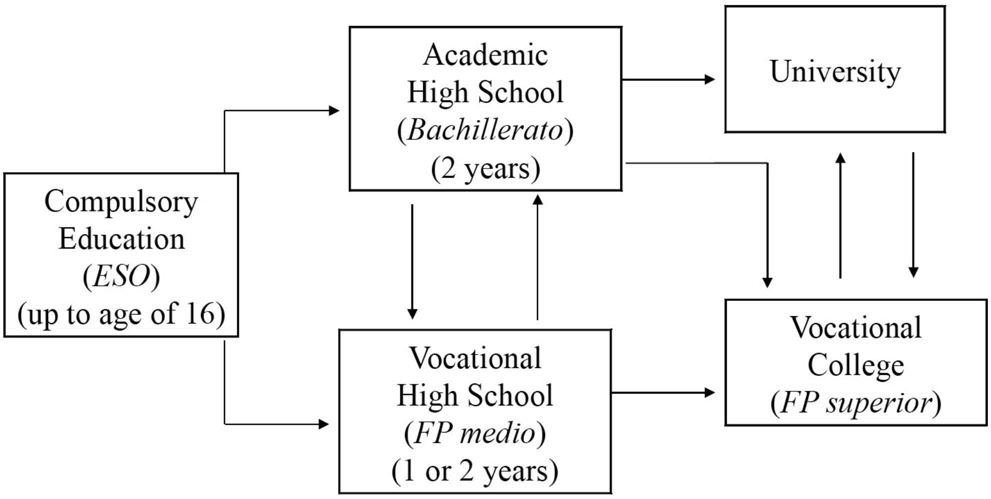

Figure 2: Schooling levels

up to the age of sixteen. If the individual passes all grades, she gets the compulsory diploma. The law establishes that the individual can fail up to two grades in this stage. In case she fails more than twice, she has to leave school and pass an official test provided by the administration to obtain the compulsory education certificate.

The next schooling level is upper secondary or high school education. It is non-compulsory and the only requirement to start in this level is to have the compulsory diploma. In this stage, the individual can choose between attending academic high school and vocational high school[4].

Academic high school has two grades, from sixteen to eighteen years old. The law establishes a maximum of four years for graduation. That is, it allows the individual to fail up to twice. If she fails more times, she has to leave the school and pass a test provided by the administration in order to get the academic high school diploma.

The education received by the student in the academic high school is more specific than in the previous level and it is predominantly oriented to the access to the tertiary level (mainly university attendance).

Since the first grade, the individual has to choose one program among four different possibilities: social sciences and humanities, arts, health and natural sciences, and technology. In addition to the program-specific subjects, all the students have several common subjects.

Access to the second grade requires that the number of failed subjects is equal to or less than two. Otherwise, the individual has to repeat the complete grade. Graduation from academic high school requires to have passed all the subjects both from the first and the second grades.

Vocational education in Spain is a schooling-based training with apprenticeship in firms up to twenty-five percent of the total time of the program. Depending on the program, the vocational track can have one or two grades. The individual can choose among a wide range of technical

[4]Individuals without the compulsory diploma and, at least with 18 years old, can attend vocational high school if they previously pass a test provided by the regional administration.

programs.

Training is oriented to give an individual a specific qualification for the access to the labor market. The kind of occupations that individuals have access to with a technical qualification are mainly blue-collar occupations (such as machinery mechanics and repairers; electronics mechanics and servicers; agricultural, forestry and fishery labourers,...) but also some white-collar jobs (such as clerical workers or sales workers).

The law allows to switch between academic and vocational high school. Depending on the origin and destination tracks, it is possible to waive some subjects.

After graduation in academic or vocational high school, the individual can attend tertiary education. In Spain, there are two possibilities at this level: university and vocational college. Access to university requires the academic high school diploma but this is only a sufficient condition. It is also necessary to pass a general (not university-specific) test that determines, jointly with the average grade in the academic track, the final grade to apply for admission in a university field.

The academic high school diploma also allows access to vocational college and, unlike for university, the individual does not have to take any admission test (although it could be required that the academic program is related to the field chosen in vocational college).

In order to enter vocational college with a technical high school diploma, the individual has two options: to validate the technical program with an academic program and, so, to get the academic high school diploma, or if she is eighteen or older, that validation process can be substituted by an admission test.

Following completion of the academic track, an individual can start in the vocational track and viceversa (an individual can attend academic high school following completion of a technical program).

3 The data

The data used in this paper come from a survey produced by the Spanish Statistics Institute in 2005, *Encuesta de Transición Educativo-Formativa e Inserción Laboral* (ETEFIL). The objective of this survey is to know the education and labor decisions of individuals who graduate from any non-university education level in the school year 2000/2001. Specifically, the sampled groups are individuals who complete in 2000/2001:

- Compulsory education.
- Academic high school.
- Vocational high school.
- Vocational college.

Additionally to these four groups, individuals who dropped out of compulsory education in 2000/2001 are also sampled.

For each of these groups, samples are obtained using a stratified two-step method: first step units are schools and second step units are pupils. Information is collected through a retrospective interview: individuals are asked about their education decisions and labor activities since 2000/2001 until the moment of the interview (May-July 2005). So, the dataset contains four observations for each individual (one by year).

In this paper, I restrict my attention to the sample of individuals who obtained the compulsory diploma in 2000/2001 (8098 individuals). It covers the period corresponding to the upper secondary education and it also covers one or two years following high school completion. For example, individuals who obtained an academic high school diploma in two years (therefore, without failing any grade) are observed two years in the option they chose following graduation. Only the individuals who graduated in the last year are right-censored in their next decision but they represent a small percentage of the total sample as I will show below.

The data contain three kinds of information:

- **Personal characteristics.** Individuals are asked about their date of birth, gender, nationality, mother and father's education and region of residence.

- **Education.** With respect to the compulsory level, individuals report the age at which they obtained the compulsory diploma and whether they attended a private, semi-private or public school. With respect to the post-compulsory levels, year by year the survey asks whether the individual decided to leave the school or to attend education. In the first case, the individual reports the leaving-school dates. In the second case, she is asked about the kind of track she chose, the grade and the program she attended and whether or not she graduated that year. The individual also indicates whether she attended a private or public school and whether she has changed to a new school in the current year. Individuals also report their decision following completion of high school education.

- **Work.** On a monthly basis, all individuals are asked about their employment or unemployment status. If they work, they report whether the job is part-time or full-time. A detailed questionnaire on the job/unemployment characteristics is asked to those individuals who are in some of the following situations:

  1. They work in a full-time job or they are unemployed in the moment of the interview.
  2. They worked in a full-time job in the same firm or they were unemployed for at least six consecutive months in the past.

  Individuals have to fill in as many questionnaires as times they are in any of the previous situations.

  The questions about the job refer to the activity of the firm, occupation, net monthly wage in an interval basis, type of contract, hours worked, necessary qualification in the job, starting and finishing dates, the means that the individual used to find the job and the reason why the individual left the job (only for jobs in the past).

  The questionnaire for the unemployment situation asks about the means that the individual uses to search for a job, whether she receives unemployment benefits, the starting and finishing dates of the unemployment spell and whether she received any job offer.

One piece of information not included in the database is individual subject grades both for the compulsory and post-compulsory levels. With respect to the lack of an average grade for compulsory schooling, the database contains the age at which the individual finished the compulsory schooling and this variable can be used as a proxy for her performance because, generally, repeating will be the main reason to finish with more than sixteen years old[5].

[5]Other explanations such as moving to a new city or being sick for a long time could also be possible but infrequent.

The absence of grades from academic and vocational high school, in principle, could be a limitation to construct the individual grade level progression. However this is not the case because from the data it is possible to infer whether an individual passed a grade or, on the contrary, she failed and had to repeat in the next year. Since attainment of a high school diploma only requires accumulation of grades (it is not necessary to achieve any average grade requisite), I consider that the information in the data captures the most essential features of the grade level progression. In appendix A, I explain how I construct the individual grade level progression and Table A1 shows the percentage of individuals who pass a grade in each school-year and in each track.

3.1 Descriptive statistics

From the initial sample, I decide to drop some individuals (see appendix A) remaining a sample with 7945 individuals. As I mentioned previously I observe the decisions following completion of upper secondary education for almost all individuals. There are only 195 individuals (2.45% of the selected sample) who are right-censored because they graduate at the end of the academic year 2004/2005. Due to the censoring, these individuals are excluded from the sample used to estimate the model. This finally contains 7750 individuals.

Following compulsory education, the individual can decide to attend the academic track or the vocational track. Regarding this, obtaining a technical diploma can require to pass one or two grades depending on the kind of program attended. Those individuals who not attend school can stay at home or enter the labor market. Each year, I decide to group the decisions following compulsory schooling in four categories: academic high school, one-grade vocational high school, two-grades vocational high school and labor market. The first three correspond to the decision of attending the academic track or the vocational track of one or two grades respectively. The last category includes individuals who not attend school that year. In order to simplify the analysis, I do not consider to stay at home as a separate category. Instead, I assume that all individuals who not attend school participate in the labor market. This is not a restrictive assumption given that most individuals who not attend school are observed working at least one month during the school months (see column fifth in Table 8).

Now I comment some descriptive aspects of the sample. All the percentages are calculated using the population weights provided by the database.

Table 1 shows several data distributions conditioning on different characteristics. In the first row we can see that more than half of the individuals who graduated from compulsory education are females. The next rows show the distribution of individuals by the age at which they obtained the compulsory diploma. Around 19% and 8% of the individuals finished with one and two years of delay, respectively. Table 2 below shows that among males the percentage that graduated with delay is higher than among females.

The next rows of table 1 show that most individuals attended compulsory education in a public school and only 5% went to a private school. With respect to parents' education, more than 40% of the parents have compulsory education, around 8% of them have some kind of vocational diploma and between 15-18% have a university degree. Fathers always present higher education than mothers. The last rows of the table show the distribution of the programs chosen in academic and vocational high school, respectively.

Table 3 contains the choice distribution by school-year with respect to the total number of individuals who start that year without having completed upper secondary education previously. The first important fact is that just after finishing compulsory education 80% of the individuals choose academic high school while only 15% choose the vocational track. And only 5% of them leave education altogether.

In the second year following compulsory schooling, there is a small increase in the percentage who chooses labor market reflecting the fact that some individuals decided to drop out after one year in upper secondary education. Between the second and the third year, however, there is an important decrease in the percentage choosing the academic track. Graduation from academic high school takes two years and by the begining of the school-year 2003/04 many individuals have graduated. Most of the non-graduated individuals will have to repeat a complete grade in that year. On the other hand, some individuals in the vocational track also graduated at the end of the second year. The fact that the percentage of the labor market option increases and that the percentage choosing vocational high school is almost constant reflects that many individuals dropped out of or transferred from the academic to the vocational track in the year three. In the last year, the situation is very similar to the previous one but with a higher percentage in labor market, reflecting again the dropping out of both types of tracks.

The decisions present differences by gender. The proportion of males who choose vocational and academic high school is 56.57% and 43.61%, respectively. These differences by gender are clearly reflected in Table 4. Around 83% of females chose the academic track just after getting the compulsory diploma and only 11% of them chose vocational high school. However, among males, 18% decided to enroll in vocational education.

Table 4 also reflects that the decisions following compulsory education are very different by the age at which the individual graduated: 91% of individuals who finished with sixteen years old entered academic high school while this percentage is lower among individuals who finished with delay, (specially for people with two years of delay).

Table 5 shows the transition matrices. They provide evidence of persistence in decisions. Each matrix indicates the one-period transition rates, that is, the percentage of transitions from origin (decision in year $t-1$) to destination (decision in year t). The transitions in the first three matrices are calculated over the total number of individuals who start the school-year t without having completed high school education (7627, 3045 and 1814 individuals respectively). The last matrix is calculated over the total number of observations in each transition across years (12486 observations).

The general conclusion from Table 5 is that there is a strong state dependence in data, especially in academic high school and labor market options. This means that most individuals who choose an option in year $t-1$ stay in the same option in year t. This persistence decreases across time for academic and vocational high school and remains more constant for the labor market option (except in the third transition in which the persistence increases). The strong state dependence in the labor market option implies that reenrollment after a period working is a rare event.

The transitions between the academic and the vocational tracks are not very frequent. Around 4% of those individuals who attend the academic track one year decide to change to the technical track in the next year. The transition from vocational to academic high school is even less frequent except between 2003/04 and 2004/05. The transitions from schooling options to labor market increase across years and they are higher in the case of vocational high school. These transitions reflect the dropout decisions.

By the moment of the interview (four years following completion of compulsory education) 24% of the individuals do not have a high school diploma. This rate varies by the delay in finishing compulsory schooling. It is 15% among individuals who finished compulsory schooling with sixteen years old and 43% and 55% among individuals who finished with one and two years of delay, respectively. That 24% includes both individuals who never participate in upper secondary education and individuals who participate but drop out.

Table 6 contains the graduation rates. They are computed as the rate of graduated over enrolled individuals (so, the counterpart are the dropout rates). For the total sample, the graduation rate is 80.07%. So, almost 20% dropped out of upper secondary education. By tracks, the completion rate in academic high school is higher than in the vocational track. And, again, there are differences by gender: females have a higher rate than males in both tracks (the biggest difference is 10.54 percentage points in vocational high school).

Among all the individuals who complete a high school track, 17.31% obtained a technical diploma. Table 7 presents the distribution of the decisions following completion of each track. According to the Spanish system, with an academic diploma it is possible to choose among university, vocational college, vocational high school or entering the labor market. As we can see in Table 7 the most chosen option is university (73%) followed by vocational college (19%). With a technical diploma, the available options are vocational college, academic high school, vocational high school (enrollment in a different program) or entering the labor market. Table 7 shows that almost 78% choose to enter the labor market. Thus, from this evidence I can conclude that individuals attend the academic track as an intermediate step to continue with tertiary education while vocational high school is seen as a way of acquiring a specific qualification before entering the labor market.

With respect to data on employment, 57.47% of the individuals work at least one month during the sample period. Table 8 gives an idea of the percentage of people working and attending school at the same time. For each year, it shows the percentage of individuals who report to work in any month between October and May, both included. These months correspond to the school year in Spain[6]. There is a clear jump in the proportion of people with a job and attending academic high school between the second and the third year following compulsory schooling. At the end of year two, most individuals graduate and those who failed or who enrolled later in the academic track work more. The proportion of people working and attending the vocational track is more constant and higher than in academic high school. Given that the technical path is more oriented to the labor market, this fact can be evidence that it is easier to make compatible education and work in vocational education than in academic high school. Finally, more than 90% of individuals who choose not to attend education report to work at least one school month.

4 Evidence from reduced-form estimates

Before explaining the model I show reduced-form estimates to give evidence of the correlations in data. The tables with the results are shown in appendix B.

In Tables B1 and B2, I show the estimates, for every period, of logit models for the probability of choosing academic high school and vocational high school while the individual has not completed upper secondary education. The number of accumulated passed grades (n) has a clear positive impact all the periods whereas the variable $d_{t-1} = LM$ has a negative effect. The positive sign of the first variable indicates that the number of accumulated passed grades increases the utility attached to that option because it reduces the effort towards graduation. The positive sign can also indicate that learning generates utility to the individual. The negative sign of the second variable indicates that attending school after a period working implies an effort cost for

[6] More exactly, the school year goes from middle September to middle June. But as I do not know the starting day of the job, I prefer do not include those months as school ones.

the individual and this cost is higher if she decides to attend academic high school.

In Table B1 we can see that being a female affects positively to the utility of attending the academic track whereas having finished compulsory schooling with delay has a negative and persistent effect all the periods although the main impact is in year one. The sign of these variables reverses in Table B2. The same happens with the dummies for parents' education: the higher the schooling level, the more positive impact on the probability of choosing academic high school. The greatest effect of parents' education is in the two first periods.

With respect to the type of school, having attended compulsory schooling in a public school has a negative impact on the probability of choosing academic high school while it has a positive effect on the probability of choosing the vocational track. However, when the individual attends school in the upper secondary level, the sign of the type of school changes. Attending a private school increases the utility of the vocational high school option but it reduces the utility of choosing the academic track.

Tables B3 and B4 show the estimations of multinomial logit models for the decisions following an academic or a vocational high school diploma, respectively. To obtain each table I make a pool of all the periods. In Table B3 the reference category is attending university while in Table B4, the reference is entering the labor market.

In Table B3 we can see that the expected present discounted value of the life-time earnings has a negative impact on decisions relative to the value corresponding to the university option[7]. On the other hand, if the individual finished compulsory schooling with more than sixteen years old or if she repeated some grade in academic high school is more likely that she chooses to attend vocational education or to enter the labor market relative to attend university. That is, a bad performance in compulsory schooling or in academic high school has a negative effect on the probability of choosing university. Having attended a private school in the academic track has no significant effect with the exception of a negative impact on the probability of attending vocational high school relative to university. The type of school attended in compulsory education in general has no significant effects. The exceptions are the negative impact of the semi-private school on the probability of vocational college and the positive effect of the private school on the probability of entering labor market relative to choose university. Although this last effect is only significant at the ten percent level, it is interesting the positive sign because, usually, parents consider that a private school provides a more qualified education than public or semi-private schools in order to continue with further schooling after graduation. With respect to parents education, the only significant and negative effects appears when the parents have the university level. Having a mother with a university degree has more impact on the probability of attending university than having a father with university education.

In Table B4, in general, the effects are less significant than in Table B3. The expected present discounted value has no effect relative to decide to enter the labor market with the exception of a negative impact on the probability of attending vocational high school again. A bad performance in compulsory schooling or in the technical track increases the probability of entering the labor market after graduation but the effects are not so clear as in Table B3. We have seen in Table 7 that labor market is the option chosen by most of the vocational high school graduates. The fact that Table B4 does not show important effects of the present value or of the previous performace confirms that people attends the technical track to acquire a suitable qualification before entering the labor market and not to attend further education although this

[7]Life-time earnings are computed using the annual wages of workers who have the corresponding option as her maximum education level. Wages are obtained from the Spanish Wage Structure Survey of 2002. See Section 6 for a more detailed explanation of the way I compute these values.

implies to receive higher earnings. Finally, with respect to the effect of parents' education it is interesting to note that if the mother has a vocational high school diploma or a university degree this has a positive impact on the probability of choosing vocational college with respect to labor market.

5 The model

The model is a sequential model of schooling decisions following completion of compulsory education. In this section, I present the structure and the specific parameterizations of the dynamic programming model and in the next sections, I explain the solution and estimation method.

5.1 The choice set

Individuals are followed from the year they finish compulsory schooling (the same for everybody) until their year of the upper secondary graduation or until the last observed year if they do not graduate.

At the begining of each year, the individual can choose between entering the labor market or continuining education. In this last case, she decides whether to attend an academic or a vocational track. In the latter, there are programs of one or two grades. Given this different requirements for graduation, I decide to consider vocational high school of one or two grades as two different options in the choice set.

According to the dataset, each individual is observed four years following completion of compulsory education. In consequence, in the model, I consider that the individual can get a high school diploma within that period. That is, if someone does not get a diploma after four years, I assume she never graduates and her maximum educational level will be compulsory schooling forever. This is not a restrictive assumption for two reasons. First, I am allowing to fail up to three times in one-grade vocational high school and up to twice in two-grade vocational high school or in academic high school (the 1990 law only allows to fail twice in the academic track). Second, the percentage of individuals who continue in upper secondary education four years following compulsory education is very small. According to the Spanish Labor Force Survey, the percentage of individuals aged 20-24 years old in 2005 with compulsory education as their maximum schooling level and who declares attending high school is only around 2%.

I also model the first decision following completion of each one of the two tracks. Although the focus of the paper is in upper secondary education, decisions in this level only can be understood by considering the choices following graduation. The data shows that the academic track is viewed as an intermediate step to continue with tertiary education while vocational high school is a previous step before entering the labor market. Given the length of the database I only can model that first decision.

The set of decisions following completion of a high school track depends on the kind of diploma obtained.

To summarize, the choice set is described as follows:

- While the individual has not graduated from high school education, she can choose among four mutually alternatives: $d_t \in \{AHS, VHS1, VHS2, LM\}$, that is, attend academic high school (AHS), vocational high school of one or two grades ($VHS1$ and $VHS2$, respectively) or not attend school and entering the labor market (LM).

- Once she completes high school, the possibilities depend on the type of diploma obtained:

– With an academic high school diploma, the individual chooses among four alternatives. Three of them refer to continuing with education (vocational high school, vocational college or university) and the fourth option is leaving education and entering the labor market.

– With a vocational high school diploma, the alternatives are some different. Again, the individual can choose to go to the labor market or to continue studying. However, the educational options are different. She can enroll in vocational college, in academic high school and also in vocational high school if she decides to attend a different program (in this case, I do not distinguish between programs of one or two grades).

I assume that the problem ends once the individual makes a decision following high school education and I assign a terminal value to that decision. Those individuals who do not graduate are assigned also a terminal value corresponding to the expected value of the compulsory diploma. These terminal values will be explained in detail later.

In the database, some individuals work and attend school at the same time. However, with the purpose of constructing a manageable version of the model, I abstract from this possibility.

5.2 The maximization problem

The individual chooses a sequence of discrete choices $\{d_t\}_{t=1}^{T_i}$ in order to maximize the expected present discounted value of utility over a finite horizon.

The maximization problem is:

$$V_t(S_t) = \max_{\{d_t\}} E\left[\sum_{\tau=t}^{T_i} \beta^{\tau-t} U_t(d_t) \mid S_t\right]$$

where $V_t(S_t)$ is the maximal expected present value at t (value function), given the state space S_t at t and given the discount factor β. U_t is the instantaneous utility function and d_t is the decision in period t.

The contemporaneous utility for each alternative is a linear function on its components:

$$U_t = \begin{cases} \overline{educ}_t^{AHS}(X_t) + \epsilon_t^{AHS} \quad if \quad d_t = AHS \\ \overline{educ}_t^{VHS1}(X_t) + \epsilon_t^{VHS1} \quad if \quad d_t = VHS1 \\ \overline{educ}_t^{VHS2}(X_t) + \epsilon_t^{VHS2} \quad if \quad d_t = VHS2 \\ wage_t(X_t) + \epsilon_t^{LM} \quad if \quad d_t = LM \end{cases}$$

The utility of not attending school is given by the wage the individual can earn working in the labor market and by a shock realized at the begining of the period t. The individual knows the wage that she can earn each year[8] but she does not know future shocks.

The current-period reward for the education options is the consumption value of school attendance. The main reward of each schooling option is the life-time earnings associated to it. However, the individuals also attach a value to the current-period utility of attending school because they can value learning per se. This value will depend on the individual effort which has both a systematic ($\overline{educ}$) and a stochastic component (ϵ) that fluctuates randomly in each period. Given the additive nature of the rewards in the labor market option, the current utility of school attendance is denominated in wage units (euros).

[8]Note that in this specification I am abstracting from the posibility of being unemployed. This is something to explore in future refinements of the model.

The stochastic components of the utility are observed by the individual at time t (not before) but not by the econometrician. They generate variation in individual decisions conditional on the variables observed by the econometrician. I assume that shocks are jointly distributed as a generalized extreme value distribution.

The state space includes all the information known to the individual at time t and affecting her decisions. S_t has two kinds of components: X_t and the vector of random shocks $\epsilon_t = \{\epsilon_t^{AHS}, \epsilon_t^{VHS1}, \epsilon_t^{VHS2}, \epsilon_t^{LM}\}$, $S_t = (X_t, \epsilon_t)$. Vector X_t includes two kinds of variables. The first are variables observed both by the individual and the econometrician at time t (such as performance) and that evolve according to previous decisions. And the second one contains variables only known by the individual (such as ability).

The value function can be written as the maximum over alternative-specific value functions. Given the choice set explained above, I define three different value functions depending on whether the individual completed high school education and the type of diploma she obtained:

1. While the individual has not completed high school, the value function is the maximum over the following alternative-specific value functions:

$$V_t^1(S_t) = \max\{V_t^{AHS}(S_t), V_t^{VHS1}(S_t), V_t^{VHS2}(S_t), V_t^{LM}(S_t)\}$$

2. Following completion of the academic track, the choice set changes and the value function is the maximum over the values attached to university, vocational college, vocational high school and labor market:

$$V^2(X_{T_i}) = \max\{Y^{UN}(X_{T_i}), Y^{VC}(X_{T_i}), Y^{VHS}(X_{T_i}), Y^{LM}(X_{T_i})\}$$

3. Following completion of the vocational track, again, the individual faces a new choice set and the value function is the maximum over the values associated to vocational college, vocational high school, academic high school and labor market[9]:

$$V^3(X_{T_i}) = \max\{Z^{VC}(X_{T_i}), Z^{VHS}(X_{T_i}), Z^{AHS}(X_{T_i}), Z^{LM}(X_{T_i})\}$$

The functions $Y(X_{T_i})$ and $Z(X_{T_i})$ implicitly embed all the decisions in the following years in further education and in the rest of the life time and, thus, the individual problem ends with these functions. These terminal values are equal to the present discounted value of life-time earnings corresponding to each terminal option. They vary by gender and region of residence (variables in X_{T_i}). In subsection 5.6, I explain this in more detail.

$V_t^k(S_t)$, $k = \{AHS, VHS1, VHS2, LM\}$, are the alternative-specific value functions corresponding to periods in which the individual has not obtained a diploma yet. They satisfy the Bellman equation and depend on t because of the finite horizon over which a diploma can be obtained.

For the alternatives corresponding to attend education, the alternative-specific value function is as follows:

$$V_t^k(S_t) = U_t^k + \beta \Big[p(c_t^k = 1 \mid X_t, d_t = k) E\Big(V_{t+1}(S_{t+1}) \mid S_t, d_t = k, c_t^k = 1\Big)$$

$$+ p(c_t^k = 0 \mid X_t, d_t = k) E\Big(V_{t+1}(S_{t+1}) \mid S_t, d_t = k, c_t^k = 0\Big)\Big]$$

[9]In the Spanish system, the options following completion of the vocational track are the same independently of whether the technical diploma is from a program of one or two grades. Thus, I do not distinguish between completing a vocational program of one or two grades when I model the terminal decision. The only important fact is that the individual has a vocational diploma.

where c_t^k is a dummy variable equal to one if the individual passes the grade in option k at the end of year t and $p(c_t^k = 1 \mid X_t, d_t = k)$ is the probability of this event. The utility of choosing option k is equal to the instantaneous utility of that option plus the future discounted expected value. This value is given by the state reached at $t+1$ which depends on the realization of c_t^k.

The maximum expected value if the individual passes the grade is given by:

$$E\left(V_{t+1}(S_{t+1}) \mid S_t, d_t = k, c_t^k = 1\right) = AHSDip_{t+1}V^2(X_{t+1}) + VHSDip_{t+1}V^3(X_{t+1})$$

$$+NoDip_{t+1}\left[I\left(t+1<4\right)E\left(V_{t+1}^1(S_{t+1}) \mid S_t, d_t = k, c_t^k = 1\right) + I\left(t+1=4\right)V^0(X_{t+1})\right]$$

$AHSDip_t$ and $VHSDip_t$ are dummy variables equal to one if the individual starts year t with an academic or a vocational diploma, respectively. In this case, in $t+1$, she obtains $V^2(X_{t+1})$ or $V^3(X_{t+1})$ depending on the type of diploma. $NoDip_{t+1}$ takes the value one if the individual does not have a high school diploma in year $t+1$. This is the case if the individual passes in year t but one grade is not enough to graduate[10]. In case she still can graduate in $t+1$, her maximum expected value is $E\left(V_{t+1}^1(S_{t+1}) \mid S_t, d_t = k, c_t^k = 1\right)$ (Emax function), where the expectation is taken over the distribution of shocks. If the next year is the last, the maximum expected value will be $V^0(X_{t+1})$ which is the present discounted value of the life-time earnings corresponding to a compulsory schooling diploma.

The maximum expected value if the individual does not pass the grade is given by:

$$E\left(V_{t+1}(S_{t+1}) \mid S_t, d_t = k, c_t^k = 0\right) = I\left(t+1<4\right)E\left(V_{t+1}^1(S_{t+1}) \mid S_t, d_t = k, c_t^k = 0\right)$$

$$+I\left(t+1=4\right)V^0(X_{t+1})$$

Therefore, when the individual decides, she faces two kinds of uncertainty. The first one comes from the shocks to the utility function because the individual only observes them at the begining of period t but not before. The second source of uncertainty refers to the fact that passing a grade is a probabilistic outcome at the begining of period t. As a consequence, the individual faces uncertainty about the relevant value function for $t+1$. This uncertainty is represented by $p(c_t^k = 1 \mid X_t, d_t = k)$. I model this probability using a logit specification and its parameterization is explained in detail in subsection 5.3.

Finally, for the alternative of not attending school and entering the labor market, the value function is:

$$V_t^{LM}(S_t) = U_t^{LM} + \beta\left[I(t+1<4)E\left(V_{t+1}^1(S_{t+1}) \mid S_t, d_t = LM\right) + I(t+1=4)V^0(X_{t+1})\right]$$

If the individual decides to enter the labor market at time t, she gets the instantaneous utility attached to this option plus the future discounted expected value. This value is given by $E\left(V_{t+1}^1(S_{t+1}) \mid S_t, d_t = LM\right)$ or $V^0(X_{t+1})$ depending on whether the individual still can complete upper secondary education. As above, the expectation is taken over the distribution of the utility shocks.

5.3 The schooling progression

Individuals make educational choices within given institutional settings. The structural approach requires the specification of the educational environment in terms of the grade level progression and the graduation requirements.

[10]Note that in VHS1, if the individual passes a grade she graduates and, therefore, this second component of the expected value dissapears.

In the model, graduation is a probabilistic outcome because the individual faces uncertainty about her final performance at the end of year t.

As I explained previously, the dataset does not have information on grades, so it is not possible to model the grade distribution in so detail as Eckstein and Wolpin (1999) does. However, this is not an important limitation because I infer from the data if an individual passes a grade. Since attainment of a high school education diploma only requires accumulation of completed grades, I consider that the information present in the data captures the most essential features of the grade level progression and can be used to construct the schooling progression of each individual[11].

Completion of the academic track requires passing two grades while completing the vocational track requires passing one or two grades depending on the program type. I define n_t^k, $k = \{AHS, VHS1, VHS2\}$, as the number of accumulated passed grades at the begining of year t in option k. $AHSDip_{t+1}$ and $VHSDip_{t+1}$ are dummy variables equal to one if the individual starts period $t+1$ with an academic or vocational high school diploma respectively. These variables can be expressed in terms of n_t^k:

$$AHSDip_t = \begin{cases} 0 & if \quad n_t^{AHS} < 2 \\ 1 & if \quad n_t^{AHS} = 2 \end{cases}$$

$$VHSjDip_t = \begin{cases} 0 & if \quad n_t^{VHSj} < A_j \\ 1 & if \quad n_t^{VHSj} = A_j \end{cases}$$

where $A_j = j$ with $j = \{1, 2\}$.

The dummy variable $VHSDip_t$ is equal to one if either $VHS1Dip_t$ or $VHS2Dip_t$ is equal to one.

An individual starts period t with an academic high school diploma if she has passed two grades by the end of period $t-1$. Otherwise, she begins year t without the diploma. The same for the vocational diploma.

The law of motion of n_t^k is as following:

$$n_t^k = n_{t-1}^k + c_{t-1}^k, \quad k = \{AHS, VHS1, VHS2\}$$

where c_{t-1}^k is a dummy variable equal to one if the individual passed one grade at the end of school year $t-1$ in choice k. Note that if the individual failed, $c_{t-1}^k = 0$ and $n_t^k = n_{t-1}^k$.

5.4 Individual heterogeneity

Given my data, potentially, I might consider to include individual heterogeneity in the current-period utilities, in the progression probabilities and in the terminal values. The heterogeneity in progression rates will be identified through the information on individual schooling progression. However, it would not be possible to identify simultaneously individual heterogeneity both in current-period utilities and in terminal values. The first one is identified with the panel dimension in observed decisions but to identify the second one I would need to observe a larger panel with individual decisions following high school completion. As I do not observe that, I only consider individual heterogeneity in utilities and in progression rates. Thus, I assume that individuals

[11] See Joensen (2009) and Beffy et al. (2009) for a similar environment to model attainment of university degrees.

finish compulsory schooling with different preferences for the academic and vocational tracks and with different ability to progress in each track.

This type of individual heterogeneity is a source of persistence in individual decisions and a source of self-selection. It is unobservable for the econometrician. The presence of this unobserved variables in the alternative-specific utilities implies that, although utility shocks were uncorrelated, the sum of the permanent heterogeneity and the shocks is a serially correlated state variable.

In this kind of dynamic programming models it is standard to account for permanent heterogeneity considering that there are M different types in the population. The idea behind this approach is that there are groups of identical or very similar individuals, so with M number of types (M smaller than the number of individuals) it is possible to measure all the heterogeneity present in the population.

In a first approach, I only consider individual heterogeneity in preferences for the vocational track. Thus, an individual of type m is described by the parameter μ_m^{VHS} which captures her preference for vocational high school. In a next step, I am interested in including also heterogeneity in the schooling progression to take into account unobserved differences in individual ability.

The distribution of μ_m^{VHS} has discrete support (M points) with π_m ($m = 1...M$) being the probability of each point. π_m is a parameter to be estimated and represents the proportion of type m individuals in the population ($\sum_{m=1}^{M} \pi_m = 1$).

5.5 Parameterizations

Here, I explain the parameterization of the consumption values of school attendance and of the probability of completing a grade.

<u>*Current value of school attendance*</u>:

The systematic component of the current utility of attending school is parameterized as follows:

$$\overline{educ}_t^{AHS}(X_t) = \mu^{AHS} + \beta_1^{AHS} n_t^{AHS} + \beta_2^{AHS} I(d_{t-1} = LM)$$

$$\overline{educ}_t^{j}(X_t) = \mu_m^{VHS} + \beta_1^{VHS} n_t^{j} + \beta_2^{VHS} I(d_{t-1} = LM)$$

with $j = \{VHS1, VHS2\}$. Given that VHS1 and VHS2 are both technical options that only differ in the number of grades towards graduation, I impose that the coefficients of n_t^j and $I(d_{t-1} = LM)$ are the same in $\overline{educ}_t^{VHS1}(.)$ and $\overline{educ}_t^{VHS2}(.)$.

Attending school rewards individuals with the present discounted value of the corresponding life-time earnings. However, schooling also can have a positive or a negative utility in the current period because attending school entails an effort. On the other hand, the individual can value learning per se (understanding a mathematical problem may generate positive utility to an individual).

The variables that I include in the value attached to attend school account for those facts. They give the net consumption value associated to each option. The number of accumulated passed grades increases the current utility because, on the one hand, it measures the learning value and, on the other hand, it reduces the effort needed for graduation in that period.

$I(d_{t-1} = LM)$ is an indicator function equal to one if the individual worked in the previous period. The coefficient of this variable measures the cost of attending education after a period working. This cost includes both the opportunity cost associated with the forgone wage and also the effort cost that supposes to come back to school after a period in the labor market. These costs can differ between academic and vocational high school.

The effort that supposes attending each track also depends on individual preferences. In the current parameterization, I include heterogeneity in preferences for the vocational track. Given that in the data, most of the individuals choose the academic track following completion of compulsory schooling, with this parameterization, I am considering that individuals who choose to attend the vocational track are those with strong preferences for the technical path.

With this parameterization for the current value of school attendance I control for different sources of persistence in the schooling decisions[12].

Probability of passing:

At the begining of each year the individual faces uncertainty about the probability of passing. The accumulation of passed grades can be interpreted as if there is a latent variable $\overline{c}_t^k$, that reflects the increase in human capital during year t. And if this increase reaches certain treshold, then the individual passes:

$$c_t^k = \begin{cases} 0 & if \quad \overline{c}_t^k < threshold \\ 1 & if \quad \overline{c}_t^k > threshold \end{cases}$$

The function $p(c_t^k = 1 \mid X_t, d_t = k)$ reflects the probability of passing a grade by the end of the academic year t conditional on the state space and on the schooling decision at time t. This function is different for academic and vocational high school. I use a logit specification to model this probability:

$$p(c_t^k = 1 \mid X_t, d_t = k) = \frac{e^{x'_{k,t}\gamma_t^k}}{1 + e^{x'_{k,t}\gamma_t^k}}$$

where

$$x'_{AHS,t}\gamma_t^{AHS} = \gamma_{0,t}^{AHS} + \gamma_{1,t}^{AHS} n_t^{AHS} + z'_{AHS}\overline{\gamma}_t^{AHS}$$

$$x'_{j,t}\gamma_t^j = \gamma_{0,t}^j + \gamma_{1,t}^j n_t^j + z'_j\overline{\gamma}_t^{VHS}$$

with $j = \{VHS1, VHS2\}$.

This probability depends on the number of passed grades at the begining of the year t (previous performance) and on observed individual characteristics included in the vector z (gender and a dummy variable equal to one if the individual finished compulsory schooling with more than sixteen years old)[13].

5.6 Terminal values

Once the individual completes upper secondary education, the choice set changes. Each option of the new choice set has attached a terminal value which embeds life-cycle decisions. Thus, the individual, given her state space at the moment of the graduation will choose the option with the higher terminal value.

These terminal values are given by the present discounted value of the life-time earnings associated to each option. In this specification I am abstracting from the unemployment probability

[12] I am interested in continuing exploring the parameterization of these values to allow for individual heterogeneity also in the academic track or to include some observed individual characteristics, such as gender or whether the individual completed compulsory schooling with delay.

[13] I am also interested in exploring other specifications for this probability and to include the type of school, the type of program attended and unobserved ability.

during the life-cycle. However, it is important to take into account that probability because there are differences in the unemployment rate by age, gender and schooling level. Thus, in the next future, I will refine the computation of the expected life-time earnings by including unemployment probabilities.

In the current specification of the terminal values, the only source of variation are the variables included in X_T which are dummies for female and region. I am also interested in considering idiosyncratic shocks in these values. Nevertheless, the current specification generates an estimated model that reproduces the main facts observed in data.

The terminal values for the options corresponding to an academic or a vocational diploma are, respectively:

$$Y^k(X_T) = \begin{cases} EW^k_{AHS} & for \quad k = \{UN, VC, VHS\} \\ W^{LM}_{AHS} & for \quad k = LM \end{cases}$$

$$Z^k(X_T) = \begin{cases} EW^k_{VHS} & for \quad k = \{VC, VHS, AHS\} \\ W^{LM}_{VHS} & for \quad k = LM \end{cases}$$

If the individual does not complete high school education within four years following completion of compulsory schooling, I assume she never graduates and her terminal value is the value of the compulsory diploma:

$$V^0(X_T) = W^{CompDip}$$

I assume that individuals observe earnings of workers with similar observable characteristics to infer their potential earnings in each schooling path.

W^{LM}_{AHS} and W^{LM}_{VHS} are the present discounted value of the life-time earnings corresponding to the option of entering the labor market just following completion of academic or vocational high school, respectively. These values are calculated using the earnings of individuals currently working who have those diplomas as their maximum schooling level.

If following high school completion, the individual decides to attend further education, the present discounted value of the corresponding life-time earnings is weighted by the probability of graduation.

For the options following graduation in academic high school, the expected present discounted values are:

$$EW^k_{AHS} = p^k_{AHS} W^k + (1 - p^k_{AHS}) W^{LM}_{AHS} \quad for \quad k = \{UN, VC, VHS\}$$

where p^k_{AHS} is the probability of graduating from option k following completion of the academic track. W^k is the present discounted value of the life-time earnings of individuals who are currently working and who have option k as their maximum schooling levels.

For the options following graduation in vocational high school, the expected present discounted values are:

$$EW^k_{VHS} = p^k_{VHS} W^k + (1 - p^k_{VHS}) W^{LM}_{VHS} \quad for \quad k = \{VC, VHS, AHS\}$$

where p^k_{VHS} is the probability of graduating from option k following completion of the vocational track. Again, W^k is the present discounted value of life-time earnings of individuals who are currently working and who have option k as their maximum schooling levels. The only difference between W^{LM}_{AHS} and W^{AHS} and between W^{LM}_{VHS} and W^{VHS} is that W^{AHS} and W^{VHS} imply that the life-cycle is shorter than in W_{AHS}^{LM} and W_{VHS}^{LM}, respectively, because the first values include the years spent in further education.

In the same way, $W^{CompDip}$ is the present discounted value of life-time earnings of workers who have compulsory education as their maximum educational level.

6 Computation of the wages included in the model

Wages in the labor market utility. To estimate the model, I need individual wages for the labor market option. As already mentioned, the dataset does not contain information on wages for all the individuals who are observed working. This information is only available for around 52% of those individuals.

However, even if I had that information, the estimation requires computing the utility of the labor market option for every individual. Thus, I need also wages even for those not observed working in the dataset.

To impute wages for the labor market option, I use the subsample of individuals who choose not to attend school and who have data on wages. The survey asks about the net monthly wage using a closed interval question. Thus, monthly wages have been calculated at the midpoint of the corresponding interval. I assume that an individual works in a year if she reports working at least one month within that year. With respect to this, to make the correspondence with the model, the year corresponds to the school year and I assume that a person works in a year if she has a job at any month from July of year t to August of year $t+1$ (both included). For those individuals who report two different jobs within a school year, I use the wage paid in their first job.

Monthly wages are multiplied by twelve to get annual equivalent earnings and they are deflated with the regional CPI in base 2001[14].

For each year, I make an OLS regression of the logarithm of the individual real wage on a dummy for female and on a set of dummies for the region of residence[15]. Then, I use the estimated coefficients to impute a real annual wage for every individual in each year.

Expected life-cycle earnings in the terminal values. As explained above, the present discounted value attached to each terminal option is calculated using the earnings of individuals currently working with the corresponding level of education as their maximum schooling attainment. For the options that imply continue in further education, that value is weighted by the probability of graduation.

The data on earnings used to compute those present values come from the Wage Structure Survey carried out by the Spanish Statistics Institute every four years. For this paper I use the survey of 2002 which it is the one that better corresponds to the sample period. This survey is very appropiate because the information on the worker's educational level is very detailed.

I use the subsample of Spanish workers aged 25-50 years old and I deflate their gross annual wages with the CPI of 2002 by region in base 2001. Then, I regress the logarithm of the real annual

[14]The annual CPI of year t by region is calculated as an average of the monthly CPI by region corresponding to the months between July of year t and August of year $t+1$. I use the monthly CPI series by region in base 2001 for the months from July 2001 to July 2005 obtained from the website of the Spanish Statistics Institute. The CPIs from July 2001 to December 2001 are expressed in base 2001 using the link coefficients provided in this website.

[15]I also estimated wage regressions controlling for more individual characteristics such as parents' education and the age when the individual finished compulsory schooling. They were not significant and, to avoid that the state space was too big, I decide not to include them in the final regression to impute the wages.

wage on age, age squared and on dummies for female and region. I run these OLS regressions for the workers who have university, vocational college, academic high school, vocational high school and compulsory schooling as their maximum education level.

Then I use the OLS coefficients to calculate the life-cycle earnings for each possible terminal option to every individual in the ETEFIL sample. To construct the life-cycle income, I assume that the age of retirement is 65 years old for all individuals. I also assume that, depending on the option the individual chooses after graduation, the length of her life cycle is: 40 years for university; 42 for vocational college, academic high school and vocational high school (if these last two options are chosen following high school completion); 45 years for academic and vocational high school if following high school graduation the individual chooses to enter the labor market. Finally, I consider 45 years to calculate the present value attached to the compulsory diploma. In this way, the present discounted value of each terminal option takes into account the opportunity cost of continuing with further schooling following upper secondary education. I compute these present discounted values assuming a discount factor of 0.97.

For the terminal options that imply attending further education, their present values are weighted by the probability of graduation. Those probabilities are calculated using the ETEFIL samples of academic and vocational high school graduates in 2000/2001. I select the two samples using the same rules explained in Appendix A for the sample of compulsory schooling graduates.

I compare the distributions of the personal characteristics and of the decisions following high school graduation in these two samples with the corresponding distributions in the sample of compulsory schooling graduates (comparison not shown in the paper). The subsamples are very similar, so I can assume that the probabilities of graduation in the terminal options for the sample I use are the same as those calculated from the samples of academic and vocational high school graduates.

Thus, I use the subsample of vocational high school graduates to calculate the probabilities of graduating from vocational college, vocational high school and academic high school following completion of the vocational high school track. Similarly, the subsample of academic high school graduates is used to compute the probabilities of graduation from university, vocational college and vocational high school following completion of the academic track. Those probabilities are calculated as the rate of graduated over enrolled individuals. The length of the dataset implies that few individuals are observed finishing university because these studies can have three, four or five grades. In consequence, to calculate the probability of graduating from university I proceed as follows. I use the group of individuals who enrolled in university in 2001/2002 (91% of the total individuals enrolled in university over the four years). Year by year, I compute the ratio between the number of individuals who passed the *correct* grade[16] to the number of individuals enrolled in that grade. Finally, I average these ratios and use this mean as the probability of graduating from university[17].

All the probabilities are calculated for males and females separately and they are shown in Table 9.

Table 10 shows the premium of the life-cycle earnings corresponding to each of the possible final schooling levels relative to the life-cycle earnings attached to compulsory education. As we can see, having more than compulsory schooling always provides a positive premium. For the alternatives following an academic diploma, the relative premium attached to have a university degree dominates the relative premium corresponding to vocational college, vocational high

[16]For *correct* grade I mean grade 1 in year 2001/2002, grade 2 in 2002/2003, grade 3 in 2003/2004.

[17]The graduation rate obtained with this method is comparable to the graduation rate calculated with data from the Ministry of Education.

school and academic high school. With respect to the options available following a vocational diploma, the relative premium corresponding to vocational college presents the highest value in the last row (total). However, if we split the premium by regions, academic high school in the Northwest and in the South has the highest relative premium. By gender, the table shows that the relative premium for females dominates the relative premium for males.

7 Solution and estimation of the model

The solution of the dynamic discrete choice model implies finding the sequence of optimal decisions that maximize the expected present discounted utility. The model is solved by backwards recursion starting from the terminal period. One of the main computational burdens of this kind of models is that it is necessary to solve the model for all the possible points of the state space at each period.

Rust (1987) established some simplifying assumptions to reduce this curse of dimensionality. He assumed i.i.d. preference shocks with a type I extreme value distribution, additive separability in the utility between the observed and unobserved components and conditional independence of the future state variables on current shocks. These assumptions provide an analytical expression for both the Emax function and the conditional choice probabilities (CCPs). This generates important computational gains not only in solving the model but also in the estimation because it avoids the simulation of multi-level integrals.

I also assume the additive separability and the conditional independence but, with respect to the preference shocks, I assume they follow a generalized extreme value (GEV) distribution that yields the nested logit model and which includes the type I extreme value distribution as a particular case. With the first two assumptions, the utility of alternative k can be decomposed as the sum of the deterministic and the stochastic components, $U_t^k = u(k, X_t) + \epsilon_t^k$. And the alternative-specific value function can be expressed as $V_t^k(S_t) = v(k, X_t) + \epsilon_t^k$, where $v(k, X_t) = u(k, X_t) + \beta \sum_{c=0}^{1} p(c_t^k = c \mid X_t)\overline{V}(X_{t+1})$.

$\overline{V}(X_{t+1})$ is the expectation of the value function over the distribution of the unobservable shocks, conditional on the observable state variables:

$$\overline{V}(X_t) = \int \max_k \left[u(k, X_t) + \epsilon_t^k + \beta \sum_{c=0}^{1} p(c_t^k = c \mid X_t)\overline{V}(X_{t+1}) \right] dG_\epsilon(\epsilon_t)$$

The conditional choice probability that individual i chooses option k conditional on the observable state variables and on the vector of parameters of the model (θ) is:

$$P(d_{it} = k \mid X_{it}, \theta) = \int I \left\{ v(k, X_{it}) + \epsilon_{it}^k > v(k', X_{it}) + \epsilon_{it}^{k'} \forall k' \right\} dG_\epsilon(\epsilon_{it})$$

The assumption that shocks follow a GEV distribution that yields a nested logit model has the advantage that both the Emax and the CCPs have closed form expressions and it is not necessary to simulate multidimensional integrals, reducing substantially the computational burden of the model. In the nested logit model the set of alternatives is partitioned into subsets (nests) that group together choices having some observable characteristics in common. The individual selects with a certain probability one of the nests and, then, she chooses among the options included in that nest. This model allows correlation across the shocks of alternatives belonging to the same nest. Furthermore, the IIA property holds within a nest but not when alternatives belong to different nests. When all correlations are zero, the generalized extreme value distribution becomes the product of independent extreme value distributions and the nested logit model becomes a standard logit.

In what follows, I show the expressions for the Emax and for the CCPs corresponding to a nested logit model. Note that just by setting all $\sigma_s = 1\ \forall s$ and S equal one, we can obtain the expressions for the standard multinomial logit, and, so, these expressions are valid both for estimating a nested or a multinomial logit discrete choice model[18]. The Emax closed form is:

$$E\left[V^1_{t+1}(S_{t+1}) \mid S_t, d_t\right] = \gamma + \log\left[\sum_{s=1}^{S}\left[\sum_{k=1}^{K_s}\exp\left(\frac{v(k,X_t)}{\sigma_s}\right)\right]^{\sigma_s}\right]$$

where γ is the Euler's constant (0.577215665), S is the number of nests, K_s is the number of options in the nest s and the parameter σ_s is a measure of the degree of correlation among the shocks of alternatives belonging to nest s. A higher value means greater independence and less correlation.

The closed form expression for the conditional choice probability is:

$$P(d_{it} = k \mid X_{it}, \theta) = \frac{\exp\left(\frac{v(d_{it},X_{it})}{\sigma_s}\right)}{\sum_{k=1}^{K_s}\exp\left(\frac{v(k,X_{it})}{\sigma_s}\right)} \times \frac{\left[\sum_{k=1}^{K_s}\exp\left(\frac{v(k,X_{it})}{\sigma_s}\right)\right]^{\sigma_s}}{\sum_{s=1}^{S}\left[\sum_{k=1}^{K_s}\exp\left(\frac{v(k,X_{it})}{\sigma_s}\right)\right]^{\sigma_s}} \qquad (1)$$

The probability of choosing alternative k in nest s is equal to the product of two terms. The first one is the probability of choosing k given that the individual is in nest s ($p_{k|s}$) and the second term is the probability of selecting nest s (p_s).

The model is estimated using the procedure proposed in Aguirregabiria and Mira (2002). They develop a new algorithm called the nested pseudo likelihood algorithm to estimate discrete decision models in which the additive and conditional independence assumptions hold. This algorithm is based on the reformulation of the value functions in terms of the CCPs. In this way, the problem of solving the model by searching for the fixed point in the value function space is reformulated as a problem of searching for the fixed point in the probability space.

The NPL procedure overcomes the main limitations of the nested fixed point algorithm (NFXP) proposed by Rust (1987) and of the CCP estimator proposed by Hotz and Miller (1993). The NFXP is a very computation demanding method because it requires solving the model for each trial value of the structural parameters. However, its main advantage is that it gives the partial maximum likelihood estimator (MLE). The CCP estimator is characterized by its computational simplicity because it only requires solving the model once. However, this estimator is less efficient than the MLE both in finite samples and asymptotically.

The NPL algorithm also avoids having to compute repeated solutions of the dynamic problem but, unlike the Hotz and Miller's estimator, the NPL gives the partial MLE, and so, there are no efficiency losses.

The NPL algorithm works on the Hotz and Miller's representation of the alternative-specific value functions in terms of choice probabilities. With a liner-in-parameters utility[19], the Hotz

[18] I develop these more general expressions because I am interested in estimating both multinomial and nested logit specifications.

[19] With a linear-in-parameters utility, it is possible to express the deterministic component of the utility as $u(k, X_t) = z(X_t, k)'\theta_u$, where $z(X_t, k)$ is the vector of state variables for option k.

and Miller's representation of $v(k, X_t)$ is

$$v(k, X_t, \theta) = \tilde{z}^P(X_t, k)'\theta_u + \tilde{e}^P(X_t, k)$$

where θ_u is the vector of parameters in the utility functions.

That is, the alternative-specific value function can be decomposed in two components. The first one, $\tilde{z}^P(X_t, k)'\theta_u$, is the expected and discounted sum of the current and future utilities which can be originated from state X_t if the individual chooses option k in period t and behaves optimally in the future. The second one, $\tilde{e}^P(X_t, k)$, is the expected and discounted sum of the future shocks which may occur starting from state X_t and option k if the individual behaves optimally in the future. More formally,

$$\tilde{z}^P(X_t, k) = z(X_t, k) + \sum_{j=1}^{T_i-t} \beta^j E_{X_{t+j}|X_t, d_t=k} \left[\sum_{k'} P(d_{t+j} = k' \mid X_{t+j}, \theta) z(X_{t+j}, k') \right]$$

$$\tilde{e}^P(X_t, k) = \sum_{j=1}^{T_i-t} \beta^j E_{X_{t+j}|X_t, d_t=k} \left[\sum_{k'} P(d_{t+j} = k' \mid X_{t+j}, \theta) e(X_{t+j}, k') \right]$$

where $E_{X_{t+j}|X_t, d_t=k}$ is the expectation over the distribution of X_{t+j} conditional to X_t and $d_t = k$ and it is calculated assuming that the individual always chooses the optimal option in the future. $P(d_{t+j} = k' \mid X_{t+j}, \theta)$ is the conditional probability of choosing option k' in the future.

$e(X_t, k)$ is the expectation of the shock ϵ_t^k conditional on X_t and on $d_t^* = k$ being the optimal decision: $E(\epsilon_t^k \mid X_t, d_t^* = k)$. Hotz and Miller showed that this expectation only depends on k, on the choice probabilities and on the distribution function of the shocks. Similarly to the Emax function, when shocks are i.i.d. type I extreme value distributed (logit model) or when they follow GEV distribution that yields a nested logit model, $e(X_t, k)$ has a closed form. Specifically, the functional form in the nested logit model is as follows:

$$e(X_t, k) = \gamma - \sigma_s \log p_{k|s} - \log p_s$$

where $p_{k|s}$ is the probability of option k conditional to be in nest s and p_s is the probability of choosing nest s (see equation (1) above).

To estimate the model is necessary to compute $\tilde{z}^P(X_t, k)$ and $\tilde{e}^P(X_t, k)$ at each period and at each state X_t. As Aguirregabiria and Mira (2008) shows, in models with finite horizon the sequence of those values is calculated by backwards induction using the following recursive expressions:

$$\tilde{z}_t^P(X_t, k) = z(X_t, k) + \beta \sum_{c=0}^{1} p(c_t^k = c \mid X_t) \left[\sum_{k'} P_{t+1}(d_{t+1} = k' \mid X_{t+1}) \tilde{z}_{t+1}^P(X_{t+1}, k') \right]$$

$$\tilde{e}_t^P(X_t, k) = \beta \sum_{c=0}^{1} p(c_t^k = c \mid X_t) \left[\sum_{k'} P_{t+1}(d_{t+1} = k' \mid X_{t+1})(e_{t+1}(X_{t+1}, k') + \tilde{e}_{t+1}^P(X_{t+1}, k')) \right]$$

with $k = \{AHS, VHS1, VHS2\}$. For the LM option, the expressions are equal to these but without the transition probability.

If in $t + 1$ the state did not evolve to a state with an upper secondary diploma, the set of available options is again $k' = \{AHS, VHS1, VHS2, LM\}$.

If the state did evolve to another one with an upper secondary diploma, the available options in $t+1$ are $k' = \{UN, VC, VHS, LM\}$ or $k' = \{AHS, VC, VHS, LM\}$ depending on whether the diploma is academic or vocational. In these cases, the expressions for $\tilde{z}^P_{t+1}(.)$ are given by

$$\tilde{z}^P_T(X_T, k) = Y^k(X_T), \quad \text{with} \quad k = \{UN, VC, VHS, LM\}$$

$$\tilde{z}^P_T(X_T, k) = Z^k(X_T), \quad \text{with} \quad k = \{AHS, VC, VHS, LM\}$$

Finally, if t is the last period, those states that does not evolve to a one with an upper secondary diploma obtain in $t+1$ the value attached to the compulsory diploma. $\tilde{z}^P_{t+1}(.)$ is simply equal to:

$$\tilde{z}^P_T(X_T) = V^0(X_T)$$

In all the terminal states both $e_{t+1}(.)$ and $\tilde{e}^P_{t+1}(.)$ are equal to zero because the terminal values still do not have idiosyncratic shocks.

The NPL algorithm is based on a nested procedure: the inner algorithm searches for the vector of parameters that maximizes the pseudo-likelihood function and the outer algorithm is a fixed point algorithm that updates the estimates of the CCPs using the parameter estimations obtained in the inner algorithm.

When the model presents permanent unobserved heterogeneity, the log-likelihood function is a finite mixture with the probabilities of being of type m as the weights:

$$L(\theta, \omega_m) = \sum_{i=1}^{N} \log \sum_{m=1}^{M} L_i(\theta_{u,m}, \theta_c)\pi_m$$

where $L_i(\theta_{u,m}, \theta_c) = \prod_{t=1}^{T_i} P(d_{it} \mid X_{it}, \theta_{u,m}, \theta_c) \prod_{t=1}^{T_i-1} p(c_{it} = 1 \mid X_{it}, \theta_c, d_{it})$ and $\theta_{u,m}$ is the vector of parameters in the utility function by type. Given that the transition probabilities do not depend on the permanent unobserved heterogeneity, the log-likelihood can be decomposed in two terms:

$$L(\theta_{u,m}, \theta_c) = \sum_{i=1}^{N} \sum_{t=1}^{T_i-1} p(c_{it} = 1 \mid X_{it}, \theta_c, d_{it}) + \sum_{i=1}^{N} \log \sum_{m=1}^{M} \pi_m \prod_{t=1}^{T_i} P(d_{it} \mid X_{it}, \theta_{u,m}, \theta_c)$$

The mixture only affects to the second term. In consequence, it is possible to obtain consistent estimates of θ_c, the vector of parameters in the transition functions, without having to solve the dynamic problem. Then, the NPL algorithm is applied to the finite-mixture likelihood to obtain estimations of $\theta_{u,m}$ and π_m given $\hat{\theta}_c$[20].

The algorithm is initialized with a consistent estimation of θ_c, $\hat{\theta}_c$, and with an initial guess for the CCPs, $\hat{P}^0$, for $\theta_{u,m}$ and for π_m, $\forall\ m$. Note that $\hat{P}^0$ contains the conditional choice probability for each state point in each year and by type.

All those elements are used to construct $\tilde{z}^{\hat{P}^0}(X_t, k)$ and $\tilde{e}^{\hat{P}^0}(X_t, k)$ for each type. Then, we compute $v^{\hat{P}^0}(k, X_t, \theta, m) = \tilde{z}^{\hat{P}_0}(X_t, k)'\hat{\theta}^0_{u,m} + \tilde{e}^{\hat{P}0}(X_t, k)$ for all the state points and types.

[20] Arcidiacono and Miller (2008) develops a NPL class estimator that allows to deal with the presence of permanent unobserved heterogeneity in the transition probabilities by combining the EM algorithm with the iteration procedure in the space of conditional choice probabilities. In the current specification of the model, there is no permanent unobserved heterogeneity in the transitions and I estimate the finite-mixture likelihood by just applying the NPL algorithm.

The next step is to construct the pseudo finite-mixture likelihood function:

$$Q(\theta_{u,m}, \pi_m, \hat{P}^0, \hat{\theta}_c) = \sum_{i=1}^{N} \log \sum_{m=1}^{M} \pi_m \prod_{t=1}^{T_i} P(d_{it} \mid X_{it}, \theta_{u,m}, \hat{\theta}_c, \hat{P}^0)$$

where $P(d_{it} \mid X_{it}, \theta_{u,m}, \hat{\theta}_c, \hat{P}^0)$ is given by expression (1):

$$P(d_{it} \mid X_{it}, \theta_{u,m}, \hat{\theta}_c, \hat{P}^0) = \frac{\exp\left(\frac{v^{\hat{P}^0}(d_{it}, X_t, \theta, m)}{\sigma_s}\right)}{\sum_{k=1}^{K_s} \exp\left(\frac{v^{\hat{P}^0}(k, X_t, \theta, m)}{\sigma_s}\right)} \times \frac{\left[\sum_{k=1}^{K_s} \exp\left(\frac{v^{\hat{P}^0}(k, X_t, \theta, m)}{\sigma_s}\right)\right]^{\sigma_s}}{\sum_{s=1}^{S}\left[\sum_{k=1}^{K_s} \exp\left(\frac{v^{\hat{P}^0}(k, X_t, \theta, m)}{\sigma_s}\right)\right]^{\sigma_s}}$$

Then, the inner algorithm gives

$$\{\hat{\pi}, \hat{\theta}_u^1\} = \arg\max_{\theta_u \in \Theta} Q(\theta_{u,m}, \pi_m, \hat{P}^0, \hat{\theta}_c)$$

$\hat{\theta}_u^1$ is used to update the CCPs,

$$\hat{P}^1 \equiv \hat{P}(d_t = k \mid X_t, \hat{\theta}_{u,m}^1, \hat{\theta}_c, \hat{P}^0) =$$

$$\frac{\exp\left(\frac{\tilde{z}^{\hat{P}^0}(X_t,k)'\hat{\theta}_{u,m}^1 + \tilde{e}^{\hat{P}^0}(X_t,k)}{\sigma_s}\right)}{\sum_{k=1}^{K_s} \exp\left(\frac{\tilde{z}^{\hat{P}^0}(X_t,k)'\hat{\theta}_{u,m}^1 + \tilde{e}^{\hat{P}^0}(X_t,k)}{\sigma_s}\right)} \times \frac{\left[\sum_{k=1}^{K_s} \exp\left(\frac{\tilde{z}^{\hat{P}^0}(X_t,k)'\hat{\theta}_{u,m}^1 + \tilde{e}^{\hat{P}^0}(X_t,k)}{\sigma_s}\right)\right]^{\sigma_s}}{\sum_{s=1}^{S}\left[\sum_{k=1}^{K_s} \exp\left(\frac{\tilde{z}^{\hat{P}^0}(X_t,k)'\hat{\theta}_{u,m}^1 + \tilde{e}^{\hat{P}^0}(X_t,k)}{\sigma_s}\right)\right]^{\sigma_s}}$$

This iterative procedure is repeated until convergence in the probability space (outer algorithm). After convergence, we obtain the estimators $\hat{\theta}_{u,m}$ and $\hat{\pi}_m \ \forall \ m$.

8 Estimation results

The results presented here are obtained with the discount factor (β) set to 0.97, with the number of types (M) equal to two and with $\sigma_s = 1 \ \forall \ s$ (that is, assuming that shocks follow a type I extreme value distribution[21]).

8.1 Parameter values

Table 11 presents the estimates and standard errors of the utility parameters. All these estimates are expressed in monetary equivalent wage units (euros of 2001). The last row contains the estimation and standard error of the proportion of type one individuals. There is 61% type one individuals in the sample.

The contemporaneous utility of attending high school is negative for both tracks although it is more negative in vocational high school. In this case, the value depends on the individual's type. A person of type one attaches a lower consumption value to attend the vocational track (-83,202.01) than a person of type two (-51,318.17). The contemporaneous utility of each schooling alternative augments in 56,986.68 when the individual completes a grade in academic high school

[21] I am also interested in estimating the model using the nested logit specification in order to compare the results with those obtained in the current (multinomial logit) approach.

and in 82,825.56 if she completes a grade in vocational high school. These positive effects are consistent with the idea that the more human capital an individual accumulates in a path, the lower the effort of attending that path and, so, the higher the utility of choosing it again in the next period. This is confirmed with the positive effect that the number of accumulated passed grades has in the probability of passing the current grade in both tracks (see Table 12). The effect is bigger in the technical track and this can explain why the individual attaches more value to a passed grade in academic high school than in vocational high school. The higher value of a passed grade in the technical track can be due to the fact that with only one grade the individual graduates if she attended a one-grade program.

The consumption utility that the individual attaches to attend academic or vocational high school is reduced by 14,790.45 and 22,650.61, respectively, if she was working in the labor market in the previous period. Those quantities measure the cost of attending school after a period working and, so, they measure both the opportunity cost associated with the forgone wage and the effort cost associated to come back to school after a year without studying. This cost is higher if the individual decides to attend the vocational track. It is reasonable to assume that the effort cost associated to enter academic high school from the labor market should be higher than the cost of entering vocational high school because, in the latter, education is more oriented to the labor market. Thus, the fact that the estimated cost is lower for the academic track is interpreted as that the effect of the foregone wage dominates the effect of the effort cost. That is, the opportunity cost of attending vocational high school is higher than the opportunity cost of attending academic high school after a period working. This can be due to the fact that the increase in the future wages associated to a vocational high school diploma is lower than the increase associated to an academic high school diploma (as we can see comparing last and fourth columns in Table 10).

Finally, Table 12 presents the estimates of the parameters included in the transition probabilites. The first part of the table shows the estimates corresponding to the probability of passing a grade in the academic track each school-year while the second part shows the estimates corresponding to passing a grade in vocational high school. I have imposed the same coefficients for the probability of passing a grade in any of the two types of technical programs that the individual can attend. I have commented above the positive effect of the number of accumulated passed grades on those probabilities. The other variables that I include are a dummy for female and a dummy equal to one if the individual finished compulsory schooling with more than sixteen years old. Being a female has a positive effect on the probability of passing a grade, with the exception of the third year in academic high school and the first year in vocational high school in which it has no significant effect. The probability of passing an academic grade is negatively affected if the individual completed compulsory schooling with more than sixteen years old. The effect on the probability of passing a vocational grade is less negative, maybe, because most individuals who finished compulsory schooling with delay prefer to choose that option.

8.2 Model fit

Tables 13-17 show the fit of the model. They are based on a simulation of schooling histories for 7750 individuals using the parameter estimates.

Table 13 compares the actual and predicted values of several statistics. The model predicts that the percentage of high school graduates is a bit higher than in data. The proportion of vocational graduates and the percentages of females among high school and academic graduates are well fitted. The model underpredicts in six percentage points the share of females among vocational gradutes. The last rows of Table 13 show that the model captures the general behaviour of the individuals in the sample because it presents a good fit of the percentage of individuals who attend academic or vocational high school, who never attends upper secondary education and who switch between the two tracks.

Table 14 shows the actual and predicted choice distributions of those individuals who start each period without having completed upper secondary education. The model captures the patterns observed in data, including the academic high school graduation effect in the third year. The only significant deviations refer to the labor market option which is underpredicted and overpredicted in the third and last year respectively.

Table 15 presents the predicted transition matrix calculated using the total number of observations in each transition. We can see that the model captures the high persistence in the academic and labor market options observed in data. With respect to vocational high school, the model predicts more persistence than what it is observed in data. The transitions between the academic and vocational tracks are well captured as well as the transitions from labor market to high school.

Table 16 compares the actual graduation rates with those predicted by the model. The total completion rate in high school education is some overpredicted while the graduation rate in academic high school is well fitted both for the full sample and by gender. The fit of the completion rate in vocational high school is not so good and this explains why the total completion rate is also overpredicted. In the current specification I do not include unobserved ability in the progression rate and this probably is the reason to not adjust well the completion rate in vocational high school. This shows evidence that in future specifications of the model will be important to consider that kind of heterogeneity.

Finally, Table 17 shows the distribution of the choices predicted by the model following completion of each one of the two high school tracks. According to the premium attached to each terminal schooling level (see Table 10), all the individuals who complete the academic track decide to attend university whereas the individuals who complete vocational high school attend vocational college or academic high school depending on the region they live. In data we observe that most individuals with an academic diploma choose university while with a vocational diploma the most chosen option is labor market. The main conclusion from this evidence is that the present discounted values of life-time earnings in the way they are computed allow to adjust the observed patterns in upper secondary education but they can not capture well individual decisions following high school completion. To fit these decisions I will have to refine the computation of life-cycle wages by including unemployment considerations and idiosyncratic shocks.

8.3 Unobserved heterogeneity

Tables 18 and 19 present several schooling outcomes on the basis of a simulation of schooling histories for 7750 individuals using the parameter estimates. There is a clear distinction in the tracks chosen by the two population types. Most individuals of type I (92%) attend the academic track whereas the proportion of type II who attends the vocational track is 40%. It is interesting to note that, in addition to the higher proportion attending vocational high school, type II's are those with the highest attendance rate in post-compulsory education because there is only a 1% who never attends high school.

With respect to the success, type I's present a lower completion rate in high school than type II's. By tracks, we can see that the dropout rate in academic high school is the same for both types. However, type I's present a higher dropout rate in the vocational track than type II's. As a consequence of the highest completion rate, the share of high school graduates among type II's is 88% while this share is 75% among type I's. The differences in the track attendance by types generate differences in the composition of the high school graduates. As we can see in Table 19, 39% of type II's who complete high school present a technical qualification while among type I's this percentage is only 4%.

9 Quantifying the impact of expected earnings on schooling choices

In this section, I present two different exercises to show the effects of changes in life-cycle earnings on schooling choices and dropout rates. In the first type of exercises, I use the previous estimates to analyse, separately, the effects of a ten percent increase in the annual wage corresponding to university, vocational college, vocational high school, academic high school and compulsory education. I denote by $w_t^{UN}, w_t^{VC}, w_t^{VHS}, w_t^{AHS}, w_t^{CompDip}$ the annual wages in year t for each of these schooling levels. After increasing by ten percent the annual wages corresponding to each education level, I compute the new expected present discounted values for each of the five scenarios. Then, I use these values and the estimates to simulate schooling histories. Table 20 shows the effects of these changes on several schooling outcomes and Table 21 contains the average increase of the expected present discounted values associated with each scenario.

The biggest effects on reducing the high school dropout rate appear after an increase in the annual wage corresponding to vocational education (columns two and three of Table 20). The high school completion rate augments around two or three percentage points with respect to the baseline scenario after an increase in the earnings attached to vocational high school and vocational college respectively. This change is due to the higher graduation rate in the technical track since the academic completion rate remains constant. The vocational track is more attractive than in the baseline scenario because the expected life-cycle earnings are higher and this translates into a higher proportion of individuals who attend this track (30% and 33% for the increase in the vocational high school and vocational college wages respectively). As a consequence, the share of vocational graduates among high school graduates rises from 19% to 31-34%. Finally, there is also a small decrease in the share of individuals who never attend post-compulsory education.

After an increase in the wage corresponding to an academic diploma, the schooling outcomes are very similar to those in the baseline model. The most important effect is the decrease by almost two percentage points of the share of individuals who never attend high school education. Increasing the life-cycle earnings attached to a university degree rises by more than two percentage points the proportion of individuals who decide to attend upper secondary education and rises also the proportion who chooses the academic track (from 83% to 91%). This last effect is not surprising given that university is only an option available following an academic diploma. There is also a decrease in the completion rate corresponding to vocational high school but, since the academic completion rate remains constant and the proportion attending the academic track is higher, there is almost no effect on the high school completion rate.

I also rise the annual wage corresponding to the compulsory diploma. In this case, the main effect is that the percentage of high school graduates over the total number of individuals reduces from 80%, in the baseline scenario, to 74%. This effect is due to the increase by almost six percentage points of the proportion of individuals who never attend post-compulsory education. So, as expected, the main effect of rising the wage that the individual can earn with a compulsory diploma is to discourage the participation in post-compulsory education.

In the second exercise, I use the wages from the Spanish Wage Structure Survey of 2006 to compute life-cycle profiles and see how individuals would behave if they faced that wage structure in the moment they make their schooling decisions[22]. Table 22 shows the premium attached to each terminal option following high school graduation relative to have compulsory schooling as the maximum education level. If we compare those numbers with the corresponding in Table 10, we can see that, except for university, the premium of each schooling level with respect to compulsory education decreased between 2002 and 2006. For instance, the premium of having a vocational college degree following an academic diploma decreased by eight percentage points. And the premium of remaining with the academic or vocational diploma following high school completion decreased by twelve and ten percentage points respectively.

I use these life-cycle profiles and the parameter estimates to simulate the individual behaviour under the wage structure corresponding to 2006. Table 23 shows several schooling outcomes predicted by the model in this scenario. The wage structure of 2006 discourages the participation in post-compulsory education: the share of individuals who decide to not participate is around 7% while in the baseline it is around 4%. The lower attendance rate is uniquely driven by the decrease in the proportion of individuals who attend the vocational track (from 18% to 11%). In contrast, the proportion who attends academic high school increases by around three percentage points.

Thus, given the premium associated with the wage structure of 2006, more individuals decide to attend the academic track in order to attend university afterwards because its life-time earnings are the only one that do not get worse with respect to the earnings of compulsory education. The premium of all the options available following vocational high school gets worse and, in consequence, the number of individuals who decide to not attend the technical track is lower.

With respect to the dropout rate, Table 23 shows that the high school completion rate decreases by two percentage points. This effect is only driven by the decrease in the completion rate in the vocational track since the completion rate in academic high school remains constant.

As a consequence of the lower attendance and completion rates in post-compulsory education, the proportion of high school graduates among the individuals who finished compulsory schooling decreases. This means that the proportion of Spanish youth who have compulsory education as their maximum schooling level will increase. Figure 3 confirms this prediction. As we can see, between 2000 and 2006, the proportion of individuals aged 20-24 years old with compulsory education as their maximum schooling level increased. Individuals in the ETEFIL sample belong to this age interval because they are 21-23 years old in 2006.

10 Concluding remarks

I develop a discrete choice structural model of individual schooling decisions following compulsory education, to analyse the effect of expected life-time earnings. This model considers the two schooling paths (academic and vocational) that the individual can choose in post-compulsory education. It also links all options available to a young people following completion of one of these two high school tracks with the corresponding expected life-cycle earnings.

The model is estimated using a Spanish microdataset on schooling histories. The parameter estimates are used to assess how individuals respond to a change in expected life-cycle earnings. I make two types of exercises. In the first one, I analyse how schooling decisions are affected

[22] I assume that the probabilities of graduation from further education are the same as in 2002.

by a change in the annual wage corresponding to university, vocational college, vocational high school, academic high school and compulsory schooling, respectively. In the second exercise, I use the Spanish Wage Structure Survey of 2006 to show how the individuals would behave if they faced these wages in the moment they make their schooling decisions. The results can be summarized as follows:

1. In the first exercise I find that the largest impact in the school attendance and completion rates in high school education is found after an increase in the earnings received by workers with a vocational qualification. Rising the annual wage of vocational college by ten percent every year reduces the dropout rate by three percentage points. The decrease is uniquely driven by the lower dropout rate in the technical track, since the percentage of individuals who drop out of academic high school remains constant. There is also an increase of fifteen percentage points in the proportion of individuals who decide to attend the vocational track instead of the academic one.

   I also find that a ten percent increase in the life-cycle wages corresponding to compulsory schooling discourages participation in post-compulsory education. The attendance rate is reduced by six percentage points.

2. The Spanish Wage Structure Survey of 2006 shows that, except for university, the premium of each schooling level with respect to compulsory education decreased between 2002 and 2006. The model predicts that this wage structure will discourage the participation in post-compulsory education, reducing the proportion of individuals who attend the technical track. Additionally, it will increase the dropout rate in this track. As a consequence, the proportion of individuals who have high school as their maximum education level will decrease. The evidence from the Spanish Labor Force Survey confirms this prediction. Among high school graduates, the share with a vocational diploma will be lower. So, there will be a change in the qualification of the high school graduates towards a reduction in the share of workers with a technical qualification.

The model does a good job in fitting observed schooling choices and provides empirical evidence of how individuals respond in their track and dropout decisions to life-cycle earnings. However, in the current specification of the model I do not consider two important aspects that I would like to include in future specifications. The first one refers to the presence of unobserved ability in the progression rates. As I comment, the absence of this type of heterogeneity can explain why the completion rates in vocational high school are overpredicted. Secondly, in the computation of both life-cycle earnings and current wages, I am abstracting from unemployment considerations. In data we observe that most individuals enter the labor market following a vocational diploma but the current specification of life-cycle wages can not reproduce this behaviour. To include the unemployment possibility could be important to improve the fit of these terminal decisions.

References

Aguirregabiria, V. and P. Mira (2002): "Swapping the Nested Fixed Point Algorithm: a Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 70, 1519–1543.

——— (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1–53.

——— (2008): "Dynamic Discrete Choice Structural Models: A Survey," *Journal of Econometrics*, forthcoming.

Arcidiacono, P. (2004): "Ability Sorting and the Returns to College Major," *Journal of Econometrics*, 121, 343–375.

——— (2005): "Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?" *Econometrica*, 73, 1477–1524.

Arcidiacono, P. and R. A. Miller (2008): "CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," Manuscript. Duke University.

Beffy, M., D. Fougère, and A. Maurel (2009): "Choosing the Field of Study in Post-Secondary Education: Do Expected Earnings Matter?" IZA Discussion Paper No. 4127.

Bradley, S. and P. Lenton (2007): "Dropping out of post-compulsory education in the UK: an analysis of determinants and outcomes," *Journal of Population Economics*, 20, 299–328.

Dustmann, C. (2004): "Parental background, secondary school track choice, and wages," *Oxford Economic Papers*, 56, 209–230.

Eckstein, Z. and K. I. Wolpin (1999): "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica*, 67, 1295–1339.

Hotz, J. and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60, 497–529.

Joensen, J. S. (2009): "Academic and Labor Market Success: The Impact of Student Employment, Abilities, and Preferences," Working Paper.

Keane, M. P. and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.

Manski, C. (1993): *Adolescent Econometricians: How Do Youths Infer the Returns to Schooling?*, in Studies of Supply and Demand in Higher Education, edited by Charles T. Clotfelter and Michael Rothschild. Chicago: University of Chicago Press, 43-60.

MEC (2006): "Sistema estatal de indicadores de la educacion. Prioritarios 2006," Instituto de Evaluacion, Ministerio de Educacion y Ciencia.

OECD (2003): "Learning for Tomorrow's World. First Results from PISA 2003," OECD Publications.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.

Willis, R. J. and S. Rosen (1979): "Education and Self-Selection," *Journal of Political Economy*, 87, S7–S36.

TABLES:

Table 1: Sample distributions (%)

Female	53.86
Age at the end of compulsory education:	
16 years old (no delay)	72.34
17 years old (1 year delay)	19.35
18 years old (2 year delay)	8.30
Compulsory education school type:	
Public	61.98
Private	4.98
Semi-private	33.04
Mother's education:	
Without education	10.67
Compulsory	45.74
Academic high school	12.16
Vocational high school	5.52
Vocational college	2.31
University	15.20
Don't know	8.40
Father's education:	
Without education	9.25
Compulsory	41.19
Academic high school	11.92
Vocational high school	4.81
Vocational college	3.65
University	18.30
Don't know	10.89
Private high school	26.33
Types of programs in academic high school:	
Arts	2.81
Health and Natural Sciences	34.25
Technology	18.59
Social Sciences	44.35
Types of programs in vocational high school:	
Program 1 (Skilled agricultural, forestry and fishery workers)	1.91
Program 2 (Electrical and electronic trades workers)	16.73
Program 3 (Clerical support workers)	29.15
Program 4 (Machinery workers, machine operators, building finishers)	20.24
Program 5 (Garment, wood working and handicraft workers)	2.86
Program 6 (Service and sales workers)	29.10

End of compulsory education is in 2001. 7750 individuals (population weights)
The last three distributions are calculated over the individuals who attend high school education, academic high school and vocational high school any time (7386, 6198 and 1541 individuals, respectively)

Table 2: Age at the end of compulsory education

	Male	Female
16 years old (no delay)	68.74	75.42
17 years old (1 year delay)	21.60	17.43
18 years old (2 year delay)	9.66	7.14

The end of compulsory education is in 2001
7750 individuals (population weights)

Table 3: Choice distribution of individuals who start in each period without having completed high school education

School-year	AHS	VHS1	VHS2	LM	Individuals
2001/02	79.57	3.66	11.11	5.66	7750
2002/03	74.72	2.58	11.41	11.29	7627
2003/04	51.26	2.56	11.49	35.69	3045
2004/05	27.63	1.06	6.48	64.83	1814

AHS: Academic high school, VHS1: 1-grade vocational high school,
VHS2: 2-grade vocational high school, LM: labor market
Population weights

Table 4: First decision following compulsory education

	AHS	VHS	LM	Total
By gender:				
Females	83.30	10.74	4.96	100
Males	75.22	18.32	6.47	100
By age at end of compulsory education:				
16 years old (no delay)	91.20	6.63	2.17	100
17 years old (1 year delay)	54.87	33.16	11.98	100
18 years old (2 year delay)	35.81	42.93	21.26	100

AHS: Academic high school. VHS: Vocational high school. LM: labor market.
7750 individuals (population weights)

Table 5: Transition matrices

	Choice (2002/03)		
Choice (2001/02)	Academic high school	Vocational high school	Labor market
Academic high school	92.25	3.68	4.07
Vocational high school	0.46	79.51	20.03
Labor market	1.94	5.65	92.41
	Choice (2003/04)		
Choice (2002/03)	Academic high school	Vocational high school	Labor market
Academic high school	85.45	3.46	11.09
Vocational high school	1.49	73.10	25.41
Labor market	3.61	3.80	92.59
	Choice (2004/05)		
Choice (2003/04)	Academic high school	Vocational high school	Labor market
Academic high school	79.83	3.57	16.59
Vocational high school	4.15	59.60	36.25
Labor market	2.26	2.71	95.03
	Choice (t)		
Choice (t-1)	Academic high school	Vocational high school	Labor market
Academic high school	89.94	3.63	6.43
Vocational high school	1.08	75.96	22.97
Labor market	2.69	3.64	93.68

Percentage of transitions from origin (choice in $t-1$) to destination (choice in t). The first three transition matrices are calculated using the total of individuals who start the school-year t without having completed high school education. The number of individuals is 7627, 3045 and 1814 respectively. The percentages in the last matrix are calculated over total observations in each transition across years (12486 observations). Population weights

Table 6: Completion rates

	Total	AHS	VHS
Full sample	80.07	78.62	66.85
Males	76.27	75.16	62.27
Females	83.27	81.30	72.81

Graduated over enrolled individuals
7750 individuals (population weights)
AHS: Academic high school
VHS: Vocational high school

Table 7: Choice distribution following high school completion

	UNIV	VC	VHS	AHS	LM	Total
Academic high school diploma	72.64	18.97	1.52	-	6.87	100
Vocational high school diploma	-	7.50	7.72	7.04	77.74	100

7750 individuals (population weights). UNIV: University, VC: Vocational college, AHS: Academic high school, VHS: Vocational high school, LM: Labor market

Table 8: People with a job during the school months (%)

School-year	AHS	VHS1	VHS2	LM
2001/02	8.13	27.58	18.78	89.78
2002/03	9.18	35.27	25.60	91.77
2003/04	18.95	33.53	38.59	96.62
2004/05	43.02	23.06	41.87	96.77

School months: October to May (both included)

7750 individuals (population weights)

AHS: Academic high school, VHS1: 1-grade vocational high school, VHS2: 2-grade vocational high school

LM: labor market

Table 9: Completion rates following a high school diploma

	Academic high school		Vocational high school	
	Males	Females	Males	Females
University	59.37	67.33	-	-
VC	77.67	84.81	67.82	68.53
VHS	67.34	51.56	75.90	79.89
AHS	-	-	53.57	58.31

Source: ETEFIL samples of academic and vocational high school graduates in 2000/2001 (population weights)

AHS: Academic high school, VHS: Vocational high school

VC: Vocational college

Table 10: Life-cycle premium

	Academic diploma				Vocational diploma			
	UNIV	VC	VHS	LM	VC	AHS	VHS	LM
Region:								
Northwest	1.62	1.39	1.42	1.42	1.39	1.43	1.42	1.40
Northeast	1.56	1.42	1.34	1.27	1.43	1.33	1.37	1.36
Centre	1.81	1.54	1.42	1.52	1.48	1.46	1.36	1.34
East	1.61	1.44	1.35	1.39	1.41	1.38	1.33	1.31
South	1.75	1.50	1.44	1.50	1.46	1.47	1.39	1.38
Gender:								
Male	1.54	1.37	1.33	1.34	1.36	1.34	1.32	1.31
Female	1.88	1.60	1.50	1.54	1.56	1.52	1.45	1.44
Total	1.67	1.46	1.39	1.41	1.43	1.41	1.37	1.35

Present discounted value of life-cycle earnings corresponding to compulsory schooling equal to one. Source: Spanish Wage Structure Survey 2002
UNIV: university, VC: Vocational college, VHS: Vocational high school
AHS: Academic high school, LM: Labor market
Northwest: Galicia, Asturias, Cantabria. Northeast: Pais Vasco, La Rioja, Navarra, Aragon. Centre: Castilla-Leon, Castilla-La Mancha, Extremadura, Madrid. East: Cataluña, Valencia, Islas Baleares. South: Andalucia, Murcia, Islas Canarias, Ceuta y Melilla

Table 11: Estimates of the utility function parameters

Number of accumulated passed grades in AHS	56,986.68	(1,498.07)
Number of accumulated passed grades in VHS	82,825.56	(1,211.50)
Reenrollment cost in AHS after a period working	-14,790.45	(1,054.20)
Reenrollment cost in VHS after a period working	-22,650.61	(979.67)
Constants:		
AHS (μ^{AHS})	-55,140.28	(783.54)
VHS type 1 (μ_1^{VHS})	-83,202.01	(1,326.87)
VHS type 2 (μ_2^{VHS})	-51,318.17	(805.38)
Type 1 probability (π_1)	0.61	(0.02)

$M = 2$ and $\beta = 0.97$. Estimates expressed in monetary equivalent units (euros of 2001)
Standard errors are in parentheses. They are calculated using numerical second derivatives
Likelihood = -15,678.39. AHS: Academic high school, VHS: Vocational high school

Table 12: Estimates of the transition probability parameters

1. Probability of passing a complete grade in academic high school ($c_t^{AHS} = 1$)

	School-year					
	2001/02		2002/03		2003/04	
Female	0.26	(0.07)	0.28	(0.07)	0.12	(0.11)
Finished compulsory schooling with delay	-1.51	(0.07)	-1.24	(0.08)	-0.39	(0.12)
n^{AHS}	-	-	0.10	(0.09)	0.48	(0.23)
Intercept	1.65	(0.05)	1.25	(0.10)	0.19	(0.24)
Individuals enrolled in the academic track	6171		5727		1577	

2. Probability of passing a complete grade in vocational high school ($c_t^{VHS} = 1$)

	School-year					
	2001/02		2002/03		2003/04	
Female	-0.20	(0.12)	0.34	(0.16)	0.64	(0.25)
Finished compulsory schooling with delay	-0.41	(0.14)	-0.65	(0.16)	-0.28	(0.24)
n^{VHS}	-	-	0.90	(0.15)	0.96	(0.24)
Intercept	0.97	(0.13)	1.07	(0.17)	0.37	(0.23)
Individuals enrolled in the vocational track	1171		1079		402	

Estimates obtained using a logit specification for both probabilities. The dependent variable takes value one in the school-year t if the individual passed the grade. Standard errors are in parentheses
n^{AHS} and n^{VHS} are the number of accumulated passed grades in academic and vocational high school, respectively. *Finished compulsory schooling with delay* is a dummy variable equal to one if the individual completed compulsory schooling with more than sixteen years old

Table 13: Predicted statistics (%)

	Actual	Predicted
High school graduates	76.00	79.66
Vocational graduates among high school graduates	17.31	19.03
Females among high school graduates	55.21	55.12
Females among academic high school graduates	57.31	58.83
Females among vocational high school graduates	45.20	39.32
Individuals who attend academic high school	79.93	82.72
Individuals who attend vocational high school	19.68	18.22
Individuals who never attend post-compulsory education	5.08	4.45
Individuals who switch between academic and vocational HS	4.69	5.39

Predicted statistics obtained by simulating schooling histories for 7750 individuals using the parameter estimates from Tables 11 and 12

Table 14: Choice distribution of individuals who start in each period without having completed high school education

	AHS		VHS		LM	
School-year	Actual	Predicted	Actual	Predicted	Actual	Predicted
2001/02	79.57	82.00	14.77	12.45	5.66	5.55
2002/03	74.72	75.27	13.99	15.28	11.29	9.46
2003/04	51.26	54.49	13.05	17.79	35.69	27.72
2004/05	27.63	16.31	7.55	4.06	64.83	79.63

AHS: Academic high school, VHS: Vocational high school, LM: labor market
Predicted statistics obtained by simulating schooling histories for 7750 individuals using the parameter estimates from Tables 11 and 12

Table 15: Predicted transition matrix

	Choice (t)					
Choice (t-1)	Acad. high school		Voc. high school		Labor market	
Acad. high school	86.46	(89.94)	4.55	(3.63)	8.98	(6.43)
Voc. high school	1.07	(1.08)	87.33	(75.96)	11.60	(22.97)
Labor market	3.42	(2.69)	3.27	(3.64)	93.31	(93.68)

Percentage of transitions from origin (choice in $t-1$) to destination (choice in t).
The percentages are calculated over total observations in each transition across years.
They are based on a simulation of schooling histories for 7750 individuals using the parameter estimates from Tables 11 and 12. The actual values are shown in parentheses

Table 16: Predicted completion rates (%)

	Total		AHS		VHS	
	Actual	Predicted	Actual	Predicted	Actual	Predicted
Full sample	80.07	83.38	78.62	77.98	66.85	83.22
Males	76.27	80.39	75.16	74.22	62.27	81.95
Females	83.27	85.98	81.30	80.84	72.81	85.24

Graduated over enrolled individuals. AHS: Academic high school, VHS: Vocational high school. Percentages obtained by simulating schooling histories for 7750 individuals using the parameter estimates from Tables 11 and 12

Table 17: Predicted choice distribution following high school completion

	UNIV	VC	VHS	AHS	LM	Total
Academic high school diploma	100	0	0	-	0	100
Vocational high school diploma	-	79.49	0	20.51	0	100

Percentages obtained by simulating schooling histories for 7750 individuals using the parameter estimates from Tables 11 and 12. UNIV: University, VC: Vocational college, AHS: Academic high school, VHS: Vocational high school, LM: Labor market

Table 18: Completion rates by type (%)

	Total	AHS	VHS	Percent of Sample
Type I	80.02	77.86	72.56	61.00
Type II	88.41	78.22	85.13	39.00

Graduated over enrolled individuals.
AHS: Academic high school, VHS: Vocational high school.
Simulation of schooling histories for 7750 individuals using parameter estimates from Tables 11 and 12.

Table 19: Statistics by type (%)

	Type I	Type II
High school graduates	74.74	87.50
Vocational graduates among high school graduates	4.39	38.94
Individuals who attend academic high school	91.78	68.30
Individuals who attend vocational high school	4.52	40.02
Individuals who never attend post-compulsory education	6.60	1.04
Individuals who switch between academic and vocational HS	2.90	9.36
Percent of Sample	61.00	39.00

Predicted statistics obtained by simulating schooling histories for 7750 individuals using the parameter estimates from Tables 11 and 12

Table 20: Schooling outcomes after a 10% increase in each annual imputed wage (%)

	Baseline	$10\%\Delta w_t^{VHS}$	$10\%\Delta w_t^{VC}$	$10\%\Delta w_t^{CompDip}$	$10\%\Delta w_t^{UN}$	$10\%\Delta w_t^{AHS}$
High school graduates	79.66	83.10	83.94	73.50	80.75	81.95
Vocational graduates among high school graduates	19.03	30.71	34.10	15.84	11.57	19.63
Individuals who attend academic high school	82.72	73.90	71.19	78.80	90.56	84.44
Individuals who attend vocational high school	18.22	29.66	32.93	14.85	10.88	19.36
Individuals who switch between academic and vocational tracks	5.39	6.60	7.23	3.79	4.48	6.44
Individuals who never attend post-compulsory education	4.45	3.05	3.11	10.14	2.04	2.65
High school completion rate	83.38	85.71	86.63	81.79	82.43	84.17
Academic high school completion rate	77.98	77.91	77.71	78.50	77.99	78.00
Vocational high school completion rate	83.22	86.07	86.91	78.37	78.61	83.13

For each scenario of life-cycle earnings, statistics are computed by simulating schooling histories for 7750 individuals using the parameter estimates from Tables 11 and 12

Table 21: Average increase on the expected present discounted values (%)

	Academic diploma				Vocational diploma				
	UNIV	VC	VHS	LM	VC	AHS	VHS	LM	Comp. Diploma
$10\%\Delta w_t^{VHS}$	0	0	5.87	0	3.00	4.23	10.00	10.00	0
$10\%\Delta w_t^{VC}$	0	8.18	0	0	7.00	0	0	0	0
$10\%\Delta w_t^{CompDip}$	0	0	0	0	0	0	0	0	10.00
$10\%\Delta w_t^{UN}$	6.90	0	0	0	0	0	0	0	0
$10\%\Delta w_t^{AHS}$	3.10	1.82	4.13	10.00	0	5.77	0	0	0

Average increase in the present discounted value of life-cycle earnings of each terminal option after a 10% increase in the annual wages corresponding to vocational high school, vocational college, compulsory schooling, universtiy and academic high school. w_t is the annual imputed wage

UNIV: university, VC: Vocational college, VHS: Vocational high school, AHS: Academic high school, LM: Labor market

Table 22: Life-cycle premium (Spanish Wage Structure Survey 2006)

	Academic diploma				Vocational diploma			
	UNIV	VC	VHS	LM	VC	AHS	VHS	LM
Region:								
Northwest	1.63	1.35	1.26	1.24	1.34	1.26	1.26	1.26
Northeast	1.54	1.32	1.20	1.18	1.31	1.20	1.20	1.20
Centre	1.80	1.46	1.33	1.38	1.42	1.36	1.30	1.29
East	1.65	1.39	1.29	1.31	1.37	1.31	1.27	1.27
South	1.70	1.40	1.29	1.35	1.36	1.32	1.25	1.25
Gender:								
Male	1.54	1.34	1.24	1.24	1.32	1.25	1.23	1.23
Female	1.85	1.46	1.33	1.38	1.42	1.35	1.29	1.29
Total	1.66	1.38	1.27	1.29	1.36	1.29	1.25	1.25

Present discounted value of life-cycle earnings corresponding to compulsory schooling equal to one.

UNIV: university, VC: Vocational college, VHS: Vocational high school

AHS: Academic high school, LM: Labor market

Northwest: Galicia, Asturias, Cantabria. Northeast: Pais Vasco, La Rioja, Navarra, Aragon. Centre: Castilla-Leon, Castilla-La Mancha, Extremadura, Madrid. East: Cataluña, Valencia, Islas Baleares. South: Andalucia, Murcia, Islas Canarias, Ceuta y Melilla

Table 23: Schooling outcomes using the Spanish Wage Structure Survey 2006 (%)

	Baseline	Predicted
High school graduates	79.66	75.21
Vocational graduates among high school graduates	19.03	10.86
Individuals who attend academic high school	82.72	85.82
Individuals who attend vocational high school	18.22	10.85
Individuals who switch between academic and vocational tracks	5.39	3.96
Individuals who never attend post-compulsory education	4.45	7.29
High school completion rate	83.38	81.13
Academic high school completion rate	77.98	78.12
Vocational high school completion rate	83.22	75.27

Statistics computed by simulating schooling histories for 7750 individuals using the parameter estimates from Tables 11 and 12. Life-time earnings are computed using the wages from the Spanish Wage Structure Survey of 2006

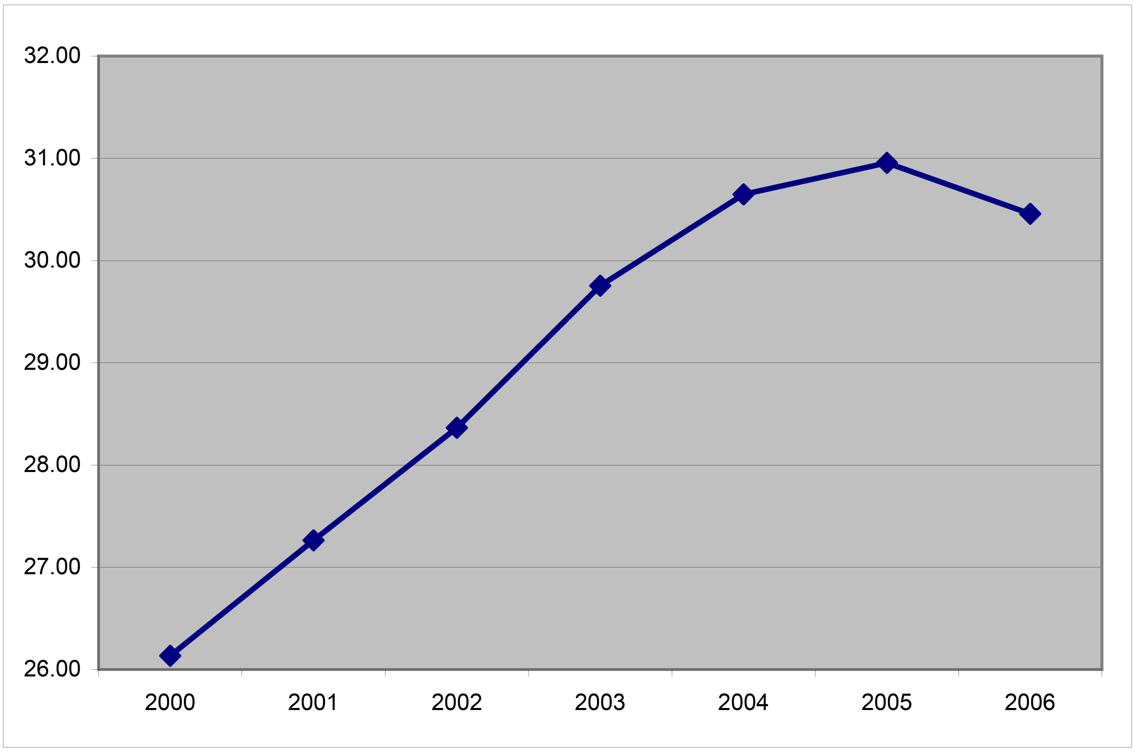

Figure 3: Percentage of people aged 20-24 years old with compulsory education as their maximum schooling level. Source: Own elaboration based on the Spanish Labor Force Survey.

Appendices

A Grade progression and sample selection

To construct the grade level progression I use the information on the grades that the individual attended for two consecutive years. If the grade in the second year is higher than the one in the first year and she is attending the same track in both years, then the individual passed. The same conclusion is drawn if the individual graduated at the end of one year. Otherwise, she failed in the first year and was repeating during the second year.

For those individuals who dropped out in the second year following compulsory schooling, there are two possible situations. The first one is that they reenroll in the track that they were attending in the first year. In this case, I use the grade in which they reenroll to infer if they had passed the grade at the end of the first year. The second possibility is that they never reenroll in the previous track. In this case, I assume they had failed. This is not a restrictive assumption because the bad performance is usually the main reason to drop out.

Table A1 shows the percentage of individuals who pass a grade in each school-year and track:

Table A1: Individuals who pass a grade (%)

School-year	AHS	VHS1	VHS2
2001/02	79.95	41.95	71.57
2002/03	78.20	67.85	79.66
2003/04	64.74	76.93	72.00

Individuals who pass over enrolled individuals
7750 individuals (population weights)
AHS: Academic high school, VHS1: 1-grade vocational high school, VHS2: 2-grade vocational high school

The initial sample contains 8098 individuals. I eliminate those individuals:

- who finished compulsory schooling being older than 18 years old (26 individuals). The 1990 law allows students to fail up to two complete grades, so the maximum age of finishing compulsory education is 18 years old. A recorded age above eighteen years old is either measurement error or it indicates that there are still some students who started education under the previous education law.

- who report a non-Spanish nationality (44 individuals). It would be interesting to take into account nationality in the model but given that there are so few non-Spanish people it would not be possible and, thus, I prefer to drop them.

- who are observed choosing artistic education (dance or music) or a program of vocational high school corresponding to the previous education law (43 individuals). I decide to drop the individuals in the first case because the artistic track is very different from the general educational path in terms of admission rules and motivation to choose it. And in the second case, I drop them because those vocational programs disappear with the 1990 law.

- who have inconsistent responses over the survey (40 individuals). For example, they are observed attending university but previously they do not report an academic high school diploma.

In total I eliminate 153 individuals and the final sample size is 7945. The eliminated individuals only represent 1.89% of the initial sample and dropping them does not generate any difference between the initial and the final samples.

B Reduced form estimations

Table B1: Probability of choosing academic high school each year

	$t=1$		$t=2$		$t=3$		$t=4$	
Variable	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
n^{AHS}	-	-	5.84**	(0.25)	4.72**	(0.18)	3.41**	(0.23)
$d_{t-1} = LM$	-	-	-2.92**	(0.35)	-1.43**	(0.24)	-2.65**	(0.24)
Private school in AHS	0.02	(0.12)	-0.39**	(0.14)	0.40*	(0.19)	0.65*	(0.28)
Female	0.59**	(0.07)	0.52**	(0.10)	0.41*	(0.16)	0.21	(0.18)
+ 16 years old	-2.40**	(0.07)	-1.52**	(0.10)	-0.77**	(0.16)	-0.47*	(0.19)
School type in compulsory education (ref: public school):								
Private	1.18**	(0.28)	0.44	(0.31)	-0.39	(0.66)	0.15	(0.49)
Semi-private	0.36**	(0.10)	0.28*	(0.11)	0.13	(0.18)	-0.24	(0.21)
Mother education (ref: compulsory diploma):								
Without education	-0.44**	(0.13)	-0.33†	(0.17)	-0.49†	(0.30)	-0.01	(0.39)
AHS	0.63**	(0.13)	0.63**	(0.18)	0.22	(0.30)	0.50	(0.30)
VHS	0.42*	(0.17)	0.21	(0.21)	-0.14	(0.37)	0.01	(0.48)
VC	0.34	(0.28)	0.61	(0.39)	0.70	(0.62)	0.10	(0.53)
UNIV	1.04**	(0.19)	0.73**	(0.21)	0.41	(0.39)	0.48	(0.39)
Don't know	-0.05	(0.17)	-0.31	(0.24)	-0.38	(0.46)	-0.18	(0.45)
Father education (ref: compulsory diploma):								
Without education	0.10	(0.14)	0.37*	(0.18)	-0.29	(0.32)	0.21	(0.40)
AHS	0.55**	(0.12)	0.37*	(0.17)	0.10	(0.27)	0.16	(0.32)
VHS	0.02	(0.16)	0.17	(0.24)	0.57	(0.36)	1.18†	(0.72)
VC	0.64**	(0.21)	0.50†	(0.27)	0.59	(0.49)	-0.34	(0.42)
UNIV	1.19**	(0.16)	0.89**	(0.20)	1.02**	(0.32)	0.14	(0.35)
Don't know	-0.32*	(0.16)	0.15	(0.21)	0.11	(0.44)	0.24	(0.41)
Regions (ref: south):								
Northwest	-0.02	(0.11)	-0.10	(0.17)	-0.14	(0.26)	-0.24	(0.28)
Northeast	-0.23*	(0.11)	-0.78**	(0.17)	-0.09	(0.27)	-0.36	(0.33)
Center	0.22*	(0.10)	0.17	(0.14)	0.50*	(0.24)	0.28	(0.24)
East	-0.76**	(0.10)	-0.60**	(0.13)	-0.16	(0.22)	-0.01	(0.25)
Intercept	1.83**	(0.10)	-0.29*	(0.13)	-2.25**	(0.26)	-2.10**	(0.31)

Significance levels: † : 10% ∗ : 5% ∗∗ : 1%. Robust standard errors. 7750 individuals

Logit specification

Table B2: Probability of choosing vocational high school each year

	$t=1$		$t=2$		$t=3$		$t=4$	
Variable	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
n^{VHS}	-	-	8.87**	(1.01)	4.36**	(0.20)	2.99**	(0.24)
$d_{t-1} = LM$	-	-	-1.21**	(0.24)	-0.61*	(0.24)	-1.47**	(0.29)
Private school in VHS	0.60**	(0.12)	0.96**	(0.21)	1.21**	(0.24)	1.79**	(0.38)
Female	-0.57**	(0.07)	-0.29**	(0.10)	-0.24	(0.15)	-0.24	(0.24)
+ 16 years old	2.10**	(0.08)	1.74**	(0.11)	0.33*	(0.17)	0.34	(0.25)
School type in compulsory education (ref: public school):								
Private	-1.59**	(0.33)	-1.15**	(0.40)	-0.42	(0.48)	-1.59	(1.00)
Semi-private	-0.54**	(0.11)	-0.64**	(0.19)	-0.05	(0.20)	-0.33	(0.27)
Mother education (ref: compulsory diploma):								
Without education	0.08	(0.14)	0.18	(0.19)	0.46	(0.30)	0.15	(0.36)
AHS	-0.60**	(0.14)	-0.45*	(0.20)	-0.00	(0.29)	-0.06	(0.44)
VHS	-0.28†	(0.17)	0.06	(0.23)	0.46	(0.33)	1.15**	(0.43)
VC	-0.48†	(0.28)	-0.40	(0.43)	-0.29	(0.60)	0.10	(0.88)
UNIV	-1.01**	(0.20)	-0.60*	(0.25)	0.07	(0.32)	0.60	(0.43)
Don't know	0.08	(0.18)	0.56*	(0.27)	0.39	(0.40)	0.77	(0.56)
Father education (ref: compulsory diploma):								
Without education	0.05	(0.15)	0.02	(0.20)	-0.21	(0.32)	-0.80*	(0.39)
AHS	-0.32*	(0.13)	-0.41*	(0.19)	-0.70*	(0.32)	-0.79†	(0.44)
VHS	0.06	(0.17)	-0.27	(0.26)	-0.23	(0.34)	-0.01	(0.62)
VC	-0.40†	(0.21)	-0.40	(0.32)	-0.07	(0.35)	-0.56	(0.72)
UNIV	-1.06**	(0.17)	-1.12**	(0.24)	-0.64†	(0.33)	-0.20	(0.43)
Don't know	0.05	(0.17)	-0.19	(0.25)	-0.57	(0.40)	-0.46	(0.53)
Regions (ref: south):								
Northwest	0.03	(0.12)	0.05	(0.18)	0.27	(0.27)	0.51	(0.34)
Northeast	0.42**	(0.12)	0.21	(0.19)	0.33	(0.25)	0.04	(0.37)
Center	-0.20†	(0.11)	0.13	(0.15)	0.32	(0.22)	0.44	(0.29)
East	0.65**	(0.11)	0.80**	(0.15)	0.55*	(0.23)	0.20	(0.32)
Intercept	-2.21**	(0.11)	-3.16**	(0.15)	-3.05**	(0.22)	-2.83**	(0.34)

Significance levels: † : 10% ∗ : 5% ∗∗ : 1%. Robust standard errors. 7750 individuals

Logit specification

Table B3: Multinomial logit for the decision following an academic high school diploma (reference outcome: UNIV)

	VC		VHS		LM	
Variable	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Life-time earnings (own option)	-0.08**	(0.01)	-0.10**	(0.02)	-0.03**	(0.01)
Finished comp. schooling with delay	1.18**	(0.13)	1.90**	(0.29)	1.03**	(0.17)
Repeated some grade in AHS	1.60**	(0.09)	1.19**	(0.27)	1.46**	(0.13)
Private school in AHS	0.02	(0.14)	-1.76**	(0.50)	-0.25	(0.20)
Type of AHS program (ref: social sciences):						
Arts	0.63*	(0.26)	0.1260	(0.67)	0.26	(0.36)
Health and Nature Sciences	-0.17†	(0.10)	0.01	(0.29)	-0.40*	(0.20)
Technology	0.78**	(0.12)	0.85*	(0.35)	-0.18	(0.20)
School type in compulsory education (ref: public school):						
Private	-0.07	(0.26)	1.10	(0.73)	0.54†	(0.31)
Semi-private	-0.24*	(0.12)	0.12	(0.31)	-0.08	(0.17)
Mother education (ref: compulsory diploma):						
Without education	0.24	(0.19)	0.66	(0.42)	0.28	(0.25)
AHS	-0.25†	(0.14)	-0.49	(0.44)	0.25	(0.19)
VHS	-0.24	(0.18)	-0.81	(0.64)	-0.36	(0.29)
VC	-0.41	(0.30)	-0.20	(0.78)	-0.20	(0.43)
UNIV	-0.54**	(0.15)	-1.18†	(0.63)	-0.55*	(0.23)
Don't know	0.10	(0.23)	-0.04	(0.61)	0.26	(0.31)
Father education (ref: compulsory diploma):						
Without education	0.14	(0.19)	0.37	(0.47)	0.32	(0.26)
AHS	-0.34*	(0.14)	0.61†	(0.37)	-0.15	(0.20)
VHS	-0.31	(0.20)	0.63	(0.53)	-0.45	(0.34)
VC	0.03	(0.20)	-0.10	(0.79)	-0.36	(0.35)
UNIV	-0.67**	(0.15)	-0.02	(0.52)	-0.07	(0.20)
Don't know	0.31	(0.20)	1.03†	(0.55)	0.49†	(0.28)
Intercept	1.99**	(0.24)	-0.31	(0.73)	-1.11**	(0.36)

Significance levels: † : 10% ∗ : 5% ∗∗ : 1%. Robust standard errors. 7750 individuals.

UNIV: University, VC: Vocational college, VHS: Vocational high school, LM: labor market.

The expected present discounted values are rescaled by 10000.

Table B4: Multinomial logit for the decision following a vocational high school diploma (reference outcome: LM)

	AHS		VC		VHS	
Variable	Coefficient	Std. Error	Coefficient	Std.Error	Coefficient	Std. Error
Life-time earnings (own option)	-0.03	(0.02)	0.03	(0.02)	-0.07*	(0.03)
Finished comp. schooling with delay	-0.44	(0.27)	-0.70**	(0.24)	-0.33	(0.26)
Repeated some grade in VHS	-0.90*	(0.41)	0.06	(0.28)	-1.43**	(0.47)
Private school in VHS	-0.48	(0.40)	-0.02	(0.31)	0.34	(0.29)
Type of VHS program (ref: program 3):						
Program 1	0.81	(0.73)	-0.26	(0.83)	1.75*	(0.74)
Program 2	-0.14	(0.53)	-0.25	(0.39)	1.22*	(0.58)
Program 4	-0.60	(0.55)	-0.63†	(0.37)	0.93†	(0.56)
Program 5	-	-	0.51	(0.59)	1.30†	(0.72)
Program 6	-0.15	(0.32)	-0.56†	(0.32)	0.51	(0.34)
School type in compulsory education (ref: public school):						
Private	-	-	0.79	(0.74)	-0.00	(1.16)
Semi-private	0.77**	(0.29)	0.17	(0.27)	0.02	(0.30)
Mother education (ref: compulsory diploma):						
Without education	0.08	(0.41)	0.11	(0.42)	0.59	(0.42)
AHS	0.61	(0.54)	-0.26	(0.57)	0.34	(0.50)
VHS	0.56	(0.58)	1.06*	(0.43)	0.61	(0.58)
VC	0.90	(0.94)	-0.51	(0.97)	0.00	(1.06)
UNIV	0.97	(0.59)	1.08*	(0.54)	0.73	(0.66)
Don't know	-0.01	(0.66)	-0.04	(0.65)	-0.58	(0.64)
Father education (ref: compulsory diploma):						
Without education	0.21	(0.43)	-0.42	(0.43)	-0.66	(0.51)
AHS	-0.28	(0.53)	-0.57	(0.50)	-0.16	(0.46)
VHS	-0.71	(0.73)	-0.24	(0.53)	-0.07	(0.60)
VC	-0.01	(0.80)	-0.45	(0.71)	-0.07	(0.86)
UNIV	0.57	(0.59)	-0.40	(0.58)	-0.08	(0.70)
Don't know	0.42	(0.60)	-0.34	(0.59)	0.82	(0.54)
Intercept	-1.09	(0.75)	-2.68**	(0.74)	-0.10	(0.97)

Significance levels: † : 10% ∗ : 5% ∗∗ : 1%. Robust standard errors. 7750 individuals.

AHS: Academic high school, VC: Vocational college, VHS: Vocational high school, LM: labor market.

The expected present discounted values are rescaled by 10000.