The Field and the Length of Studies in the French Post-Secondary Education: The Effect of Expected Earnings

Magali Beffy* Denis Fougère† Arnaud Maurel‡

July 2, 2007
Preliminary version

Abstract

This paper examines the determinants of the choice of the field of study and of the education level achieved in the chosen track, by using a framework in which students entering higher education are assumed to anticipate their future market earnings. For that purpose, we use French data coming from the *Generation* surveys collected by the CEREQ (*Centre d'Etudes et de Recherches sur les Qualifications,* France). Our econometric approach is based on a structural three-equations model, which is identified thanks to some exclusion restrictions. Once graduated from high school, individuals are supposed to sequentially choose their post secondary major field, then their education level (i.e. the length of their studies in the chosen field), and finally they enter the labor market. We account for the potential correlation between the unobserved individual-specific preferences that affect the values of each post secondary field of study, as well as between the unobserved individual-specific factors that affect the choice equation of the educational level and the labor market earnings equation. Simulating for each given field of study a 10 percent increase of the expected earnings results in a significant impact on the allocation between fields. We also get significant effects on the allocation between education levels when simulating a 10 percent increase of the expected post-college earnings, while the same increase at the graduate level has a negligible effect.

JEL Classification: J24, C35

*CREST-INSEE, Paris, and IZA, Bonn. E-mail: magali.beffy@insee.fr
†CNRS, CREST-INSEE, Paris, CEPR, London, and IZA, Bonn. E-mail: fougere@ensae.fr
‡CREST-INSEE, Paris. E-mail: arnaud.maurel@ensae.fr

1 Introduction

Over the last years, the French higher education system has been the object of much discussion and sharp criticism. Aghion and Cohen, in public for the French *Conseil d'Analyse Economique* (Aghion and Cohen, 2004), emphasize the main difficulties the French higher education system, and especially the French university, has to cope with. Pointing out, among others things, the high dropout rate in French universities, they argue that the French higher education system needs urgently to be reformed. In this context, it seems crucial to understand students' educational choices in terms both of the major and of the level of their post-secondary studies.

In this paper, we focus on the effect of expected labor market income on individual post-secondary schooling choices. In particular, we assess the sensitivity of students' schooling choices to expected earnings by estimating a structural model of post-secondary educational choices. In the existing applied literature, several papers explicitly consider the impact of expected labor market earnings on schooling and career choices. A first set of papers study these issues by using a rational expectations framework. For instance, Willis and Rosen (1979) allow the demand for college education to depend on expected future earnings. Assuming that students form rational (i.e. unbiased) expectations, these authors show that the expected flow of post-education earnings are strong determinants of college attendance. Berger (1988) also focuses on the impact of expected earnings on the individual demand for post-secondary education: his results show that, when choosing college majors, students are more influenced by the (rationally) expected flow of future earnings than by their expected initial earnings.[1] Then, following Keane and Wolpin (1997), several econometricians have estimated structural dynamic models of schooling decisions (Cameron and Heckman, 1998, 2001; Eckstein and Wolpin 1999; Belzil and Hansen, 2002; Keane and Wolpin, 2001). Their papers assume that students form rational earnings expectations conditional on schooling decisions, and that the expecetd earnings affect in turn schooling choices. Papers by Arcidiacono (2004, 2005) estimate sequential models of college attendance, in which the value of each alternative also depends on the corresponding expected flow of earnings.

A second set of papers examines the validity of the rational expectations assumption in the context of educational choices. More precisely, they consider the specification and the estimation of schooling decision models in which the rational expectations assumption is relaxed. For instance, Buchinsky and Leslie (2000) use a dynamic schooling decision framework in which they compare the predictions of

[1] See also Altonji (1993), who estimates a sequential model in which schooling decisions depend on expected returns to education.

models assuming different forecasting behaviors (*myopic*, rational or adaptative); their results show that assuming adaptative (i.e. Bayesian) earnings expectations leads to more realistic predictions in terms of the impact on educational attainment of the changes in the wage structure observed from 1980 to 1994 in the U.S. Previously, Freeman (1971, 1975) and Manski (1993) have proposed models assuming that individuals have *myopic* expectations relatively to their potential labor market earnings. Within such a framework, students are assumed to form their wage expectations by observing the earnings of comparable individuals who are currently working. According to Manski's terminology, such expectations are computed "in the manner of practicing econometricians". More recently, Boudarbat and Montmarquette (2007) examine the effect of expected earnings on the choice of fields of study in Canada; for that purpose, they estimate a mixed multinomial logit model applied to the choice of major, using a sample of Canadian university graduates. These authors also relax the assumption of rational expectations; as in Manski's paper (1993), the predicted earnings are computed from the wages of young individuals who have the same education level and who are currently working.

Our model relies also on a *myopic* expectations assumption. Unlike the previous papers, it concentrates on the effects of expected earnings on the individual choices of the field and the length of post-secondary studies, which have never been considered simultaneously. It is also the first microeconometric study devoted to these issues in France.

The remainder of the paper is organized as follows. Section 2 describes the theoretical model. The econometric counterpart of this model and the likelihood function are discussed in Section 3. Section 4 describes the data and presents some preliminary statistics, while Section 5 contains the estimation and simulation results.

2 The model

In this section, we present a theoretical model of post-secondary education. After graduating from high-school, individuals are assumed to choose simultaneously their field of study and their education level (i.e. the length of their studies) in the chosen field. Note that we restrict our analysis to individuals who attend university. [2] Once they leave the post-secondary education system, they are supposed to enter the labor market. Thus we consider a sequence of three individual decisions:

- Stage 1: When entering college, each student chooses his/her post-secondary

[2] The argument justifying our choice to focus on individuals attending university is detailed in the section devoted to the data.

major.

- Stage 2: He/she keeps on studying in the field chosen in stage 1, until he/she reaches an endogenously determined level of education.

- Stage 3: He/she leaves the post-secondary education system and participates in the labor market.

2.1 Stage 1: Choice of the major

After graduating from high-school (and getting the final high-school diploma, called “Baccalauréat” in France), the individual who decides to continue studying, must choose the college major, hereafter indexed by j^*.[3] . We assume that this choice is made among a set of M college majors. Furthermore, we assume that the chosen field (j^*) depends on the individual's expectations concerning both the education level he/she will achieve within this major (see stage 2) and the expected flow of his/her labor market earnings, which is assumed to depend on his/her educational level (see stage 3). An important underlying assumption is that the flow of future earnings and the highest level of education reached in field j^* are partly uncertain.[4]

Let us denote by V_j the individual utility associated with the choice of field j ($j = 1, \ldots, M$). This utility is assumed to be composed of three additive elements, respectively denoted by v_{0j} and v_{1j}. The first term v_{0j} represents the *intrinsic* value (i.e. the consumption value) of the major, while v_{1j} may be considered as the *investment value* of a post-secondary education in field j. It is a function of the sum of the expected flows of future labor market earnings which are associated with the $L + 1$ educational levels that can be reached within field j, each of these expected values being weighted by the probability $\Pr(K = k \mid J = j)$ to reach the k-th educational level $(k = 0, ..., L)$ within field j $(j = 1, ..., M)$. Here $k = L$ denotes the highest educational level that can be reached within major j, and $k = 0$ corresponds to the case where the student drops out from the major before terminating the first year of college. Then the utility V_j of major j can be written as :

$$V_j = v_{0j} + v_{1j}, \text{ for } j = 1, ...M$$

[3]We omit the individual subscript for the sake of simplicity.

[4]We suppose that each individual has an idiosyncratic propensity to achieve a high level of education. This propensity is partly affected by random factors, such as his/her own health status and unexpected changes in his/her family environment. These factors are assumed to be *ex ante* unknown by the individual when choosing his/her major, and then revealed when attending university. Hence, at end of stage 2, there are known by the student.

where

$$v_{1j} = \alpha \sum_{k \in \{0,1,\ldots,M\}} \Pr(K = k \mid J = j).E\left(V_e \mid J = j, K = k\right)$$

$E\left(V_e \mid J = j, K = k\right)$ denoting the expected flow of future earnings associated with education (j, k), and α being an unknown parameter to be estimated.[5]

The subcomponent v_{0j} can be interpreted as the non-pecuniary value of field j. It may correspond to the "social gratification" brought by studying in major j and to the individual's taste for this major. We assume that v_{0j} is a linear function of a set of observable individual covariates that affect the attractiveness of field j (e.g. gender, place of birth, parents' nationality and profession, past educational history of the student, including the cumulated delay when entering secondary school or when graduating from high school). It is also depending on a random term u_j representing the unobservable part of individual preference for studying in major j. Consequently, v_{0j} is specified as

$$v_{0j} = X_1' \beta_1^j + u_j$$

where β_1^j is a parameter vector associated with X_1 and specific to field j. The individual chooses the education field j^* that corresponds to the highest indirect utility level:

$$j^* = \arg \max_{j \in \{1,\ldots,M\}} V_j$$

2.2 Stage 2: Choice of the educational level

Once the individual has chosen his/her major j^*, he/she studies until he/she reaches a level k_j^* of education within field j. We assume that this level k_j^* is an element of a set of $L+1$ possible levels corresponding to the different degrees which may be obtained in each major; $k = 0$ corresponds to a dropout, which occurs when a student leaves university during the first year of college (without any post-secondary degree), $k = 1$ refers to the degree called "DEUG" in France which is generally obtained after two years of college, $k = 2$ corresponds to the BA degree (called "Licence" in France), $k = 3$ corresponds to the MA degree ("Maîtrise") and $k = L = 4$ refers to the Graduate level.

The optimal level k_j^* of education within major j is supposed to be determined by the individual propensity $\widetilde{k}_j$ to succeed in long post-secondary studies within

[5]The functional form of probabilities $\Pr(K = k \mid J = j)$ is specified in the section devoted to the econometric specification of the model.

this major.[6] More precisely, we assume that the optimal level k_j^* of education is defined by the following latent model:

$$k_j^* = \begin{cases} 0 & \text{if } \widetilde{k}_j \leq \alpha_1 \\ 1 & \text{if } \alpha_1 < \widetilde{k}_j \leq \alpha_2 \\ \vdots & \\ L & \text{if } \alpha_L < \widetilde{k}_j \end{cases}$$

where $\{\alpha_1, \ldots, \alpha_L\}$ are latent (unobservable) thresholds that correspond to the minimum ability levels required to obtain the different degrees. The latent propensity $\widetilde{k}_j$ is assumed to depend linearly on observable covariates X_2 (such as gender, place of birth, parents' profession, etc..), on the expected earnings increase associated with one more year of education within field j, equal to $E(V_e \mid J = j, K = k+1) - E(V_e \mid J = j, K = k)$, and on an individual unobservable term v which is unknown *ex ante* by the student when he /she decides to enter college. Thus the propensity $\widetilde{k}_j$ is defined as:

$$\begin{aligned} \widetilde{k}_j &= X_2'\beta_2 + \gamma_{20} E(V_e \mid J = j, K = 0) \\ &\quad + \Sigma_{k=0}^{L-1} \gamma_{2(k+1)} \left[E(V_e \mid J = j, K = k+1) - E(V_e \mid J = j, K = k) \right] + v \end{aligned}$$

where β_2, γ_{20}, $\gamma_{21}, \ldots, \gamma_{2L}$ are unknown scalars to be estimated.

Note that, throughout stage 2, discussion is conducted under the assumption that students cannot be forced, for instance after a maximum number of grade repetitions, to leave the educational system, so that the educational level that is effectively reached results from the individual's decision. It is important to note that we only account for self-selection through the post secondary educational path: this seems to be a quite sensible assumption for the French university system.

2.3 Stage 3: Labor market earnings

Having obtained the educational level (degree) k_j^* in major j^*, the student then enters the labor market. We assume that the labor market is an absorbing state: individuals do not resume studies after entering the labor force. When making his/her post-secondary schooling decisions (stages 1 and 2), the individual is assumed to

[6] This framework is consistent with the ordered probit model that will be detailed in the next section, devoted to the econometric specification.

anticipate their effects on future labor market earnings. In order to take both employment and nonemployment spells into account, we refer to average *earnings* as wages weighted by employment spell durations. Hence, the average annual log-earnings in a T years long labor market history for a worker with education (j, k) is given by :

$$\overline{\ln w}_{jk} = \frac{\sum_{s=1}^{N_e} \ln(w_{s,jk}) l_s^e}{T} \tag{1}$$

with

$$T = \sum_{s=1}^{N_e} l_s^e + \sum_{s=1}^{N_u} l_s^u$$

where N_e (respectively, N_u) is the number of observed employment (nonemployment) spells in the individual labor market history, $w_{s,jk}$ is the annual wage in the s-th employment spell, l_s^e (respectively, l_s^u) are durations of the s-th employment (respectively, nonemployment) spell, and T is the total length of the observed labor market history of the individual.[7]

Thereafter, we focus only on this aggregate notion of labor market earnings, without modeling separately wages and individual probabilities of employment (and nonemployment). This could be consistent with the students' behavior when tehy take their post-secondary schooling decisions: most individuals anticipate future labor market conditions as a whole, without separately taking into account the effects of their educational choices on wages and on employment probabilities.

Labor market earnings depend on the post-secondary educational field and level, namely on the pair (j^*, k_j^*). Note that our framework accounts for the earnings gaps, not only between schooling levels (within a given field of study), but also between fields of study (for a given educational level, or degree). Earnings are also supposed to be a function of exogenous and predetermined individual characteristics.[8]

If we assume that the discount factor is equal to 1, the value of the expected flow of future labor market (log-)earnings for a post-secondary education (j, k) is estimated by :

$$E\left(V_j \mid J = j, K = k\right) = T \times \overline{\ln w}_{jk}$$

[7]As there is no information in the data about the level of unemployment benefits received during the unemployment spells, we assume that this amount is equal to one.

[8]In a further version of the paper, we will also study the impact of earning risks on post-secondary schooling choices by allowing the variance of earnings to differ across post-secondary educational levels.

3 Econometric specification

3.1 The econometric model

The individual choice of the post-secondary field of study (*stage 1*) is generated by a random utility multinomial model:

$$j^* = j \Leftrightarrow V_j \geq V_{j'}, \quad \forall j' \neq j, \quad j' = 1, 2, ..., M \tag{2}$$

with

$$V_j = X_1' \beta_1^j + \alpha \sum_{k \in \{0,1,\ldots,M\}} \Pr(K = k \mid J = j).E\left(V_e \mid J = j, K = k\right) + u_j$$

where j^* is the chosen field among the choice set $\{1, ..., M\}$. Then, the highest education level k_j^* that is achieved within the chosen field is given by :

$$\forall k \in \{0, 1, ..., L\}, k_j^* = k \Leftrightarrow s_k < \widetilde{k}_j \leq s_{k+1} \tag{3}$$

with

$$\begin{aligned} \widetilde{k}_j \quad = \quad & X_2' \beta_2 + \gamma_{20} E\left(V_e \mid J = j, K = 0\right) \\ & + \Sigma_{k=0}^{L-1} \gamma_{2(k+1)} \left[E\left(V_e \mid J = j, K = k+1\right) - E\left(V_e \mid J = j, K = k\right)\right] + v \end{aligned}$$

Finally, the average log-earnings equation is assumed to be given by:

$$\overline{\ln w}_{jk} = X_{3(j,k)}^{'} \beta_3 + \epsilon \tag{4}$$

where $\overline{\ln w}_{jk}$ is the average log-earnings defined in equation (1), $X_{3(j,k)}$ is a vector of observed characteristics that may affect labor market earnings, including post-secondary education, and ϵ denotes the individual's unobserved ability that affects earnings.

We assume that, when making their post-secondary educational choices, students evaluate their expected flow of labor market earnings by observing the earnings of comparable workers (i.e. having the same educational level but currently working). This assumption concerning expectations was introduced by Freeman (1971) and developed by Manski (1993).[9] More precisely, students are supposed to predict their future labor market earnings by using the parameter estimates of earnings equation of comparable individuals who are already in the labor force.

[9] See also Boudarbat and Montmarquette (2007) for a more recent example.

Denoting by $\widehat{\beta}_{(3,p)}$ the estimated coefficients of the earnings equation (4) of these comparable workers, then the expected earnings for a student who chooses to reach an educational level (j,k) is:

$$\widehat{\ln w}_{jk} = X_{3(jk)}\widehat{\beta}_{(3,p)}$$

3.2 Stochastic assumptions

Residuals are supposed to be normally distributed. Potentially, we should allow for non-zero correlations between the unobserved individual-specific random terms that affect the three equations of our model. Nevertheless, for tractability reasons, we assume independence between the individual unobserved preference u_j $(j = 1, \ldots, M)$ for the j-th field of study and the residuals v and ϵ of the two other equations.[10] However, since only differences in utility levels matter in random utility models, it is sufficient to impose that the correlations between each of the residuals $u_j - u_1$ $(j = 2, ..., M)$ and v (or ϵ) are zero. Consequently, the whole vector of residuals is assumed to be distributed as:

$$\begin{pmatrix} v \\ u_2 - u_1 \\ u_3 - u_1 \\ \ldots \\ u_M - u_1 \\ \epsilon \end{pmatrix} \sim \mathcal{N}(0, \Sigma)$$

where Σ is the $(M+1) \times (M+1)$ covariance matrix of the model residuals, with $\Sigma[1,1] = 1$ and $\Sigma[2,2] = 1$ for identifiability reasons. Given the constraints we impose on correlations, the covariance matrix is:

$$\Sigma = \left(\begin{array}{c|ccccc|c} 1 & \ldots & \ldots & \ldots & \ldots & & \ldots \\ \hline 0 & 1 & \ldots & \ldots & \ldots & & \ldots \\ 0 & \Sigma_{32} & \Sigma_{33} & \ldots & \ldots & & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & & \ldots \\ 0 & \Sigma_{M2} & \ldots & \ldots & \Sigma_{MM} & & \ldots \\ \hline \Sigma_{(M+1),1} & 0 & \ldots & \ldots & 0 & & \Sigma_{(M+1),(M+1)} \end{array}\right) \quad (5)$$

The particular order of the residuals in this vector enables us both to use Cholesky decomposition and to verify our constraints. Thus, if Γ denotes the Cholesky factor for the covariance matrix Σ, we have:

[10] This simplifying assumption will be relaxed in a further version of the paper. Still, it does not seem absurd to assume that the idiosyncratic tastes for each major are very specific to the individual, and thus that they are correlated neither to the random propensity (ability) to reach high levels of education, nor to the unobserved labor market productivity.

$$\Sigma = \Gamma.\Gamma' \tag{6}$$

where

$$\Gamma = \left(\begin{array}{c|ccccc|c} 1 & 0 & 0 & ... & ... & & 0 \\ \hline 0 & 1 & 0 & 0 & ... & & 0 \\ 0 & \alpha_{32} & \exp(d_1) & 0 & ... & & 0 \\ ... & ... & ... & ... & ... & & .... \\ \hline \alpha_{(M+1),1} & 0 & ... & ... & 0 & & \exp(d_{M-1}) \end{array} \right) \tag{7}$$

Note that we impose the positivity of the diagonal terms of matrix Γ. Hence, the Cholesky decomposition of Σ is unique.

3.3 The likelihood function

Under our stochastic assumptions, the contribution to the likelihood function of an individual who chooses the educational level (j, k) and who gets the average labor market log-earnings $\overline{\ln w}_{jk}$ is:

$$\begin{aligned} l(j^*, k^*_{j^*}, \overline{\ln w}_{jk}) &= \Pr\left[\bigcap_{j' \neq j^*} \left(u_{j'} - u_{j^*} \leq f\left(j^*\right) - f\left(j'\right) \right) \right] \times g(\epsilon) \\ &\quad \times \Pr\left[s_{k^*_{j^*}} - \widetilde{h} < v \leq s_{k^*_{j^*}+1} - \widetilde{h} \mid \epsilon \right] \end{aligned} \tag{8}$$

where

$$\widetilde{h} = X_2'\beta_2 + \gamma_{20}\left[T X_{3(j0)} \widehat{\beta}_{(3,p)} \right] + \Sigma_{k=0}^{L-1} \gamma_{2(k+1)} T \left[X_{3(j,k+1)} \widehat{\beta}_{(3,p)} - X_{3(jk)} \widehat{\beta}_{(3,p)} \right],$$

$$f\left(j\right) = X_1'\beta_1^j + \alpha T \sum_{k \in \{0,1,\ldots,M\}} X_{3(jk)} \widehat{\beta}_{(3,p)} \times \left[\Phi\left(s_{k+1} - \widetilde{h} \right) - \Phi\left(s_k - \widetilde{h} \right) \right],$$

$$g(\epsilon) = \frac{1}{\sqrt{\Sigma[M+1, M+1]}} \times \varphi\left(\frac{\epsilon}{\sqrt{\Sigma[M+1, M+1]}} \right),$$

and

$$\Pr\left[s_k - \widetilde{h} < v \leq s_{k+1} - \widetilde{h} \mid \epsilon \right] = \Phi\left(\frac{s_{k+1} - \widetilde{h} - m_{cond}}{\sigma_{cond}} \right) - \Phi\left(\frac{s_k - \widetilde{h} - m_{cond}}{\sigma_{cond}} \right)$$

φ and Φ being respectively the density and cumulative density functions of the standard normal distribution $\mathcal{N}(0,1)$. In the last expression, m_{cond} and σ_{cond} denote respectively the conditional mean and the conditional standard deviation of the gaussian distribution of v given ϵ, given by:

$$m_{cond} = \frac{\Sigma[1, M+1]}{\Sigma[M+1, M+1]}\epsilon$$

and

$$\sigma_{cond}^2 = 1 - \frac{\Sigma[1, M+1]^2}{\Sigma[M+1, M+1]}$$

Finally, for estimating the probability:

$$\Pr\left[\bigcap_{j' \neq j^*} \left(u_{j'} - u_{j^*} \leq f\left(j^*\right) - f\left(j'\right)\right)\right],$$

we use a method proposed by Train (Train, 2003); first, we complete the Cholesky matrix Γ by adding a column and a row of zeros:

$$\Gamma^c = \left(\begin{array}{c|c|cccccc}
1 & \mathbf{0} & 0 & 0 & ... & ... & 0 \\
\hline
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & ... & ... & \mathbf{0} \\
\hline
0 & \mathbf{0} & 1 & 0 & 0 & ... & 0 \\
0 & \mathbf{0} & \alpha_{32} & \exp(d_1) & 0 & ... & 0 \\
... & ... & ... & ... & ... & .... & \\
\alpha_{(J+1),1} & \mathbf{0} & 0 & ... & ... & 0 & \exp(d_{J-1})
\end{array}\right)$$

Then

$$\Sigma^c = \Gamma^c.\Gamma^{c'}$$

and the covariance matrix of the vector:

$$(u_1 - u_{j^*}, u_2 - u_{j^*}, ..., u_{j^*-1} - u_{j^*}, u_{j^*}, u_{j^*+1} - u_{j^*}, ..., u_J - u_{j^*}, v, \epsilon)'$$

is equal to:

$$A_{j^*}.\Sigma^c.A_{j^*}^{'} \tag{9}$$

where A_{j^*} is a transformation matrix defined by:

$$A_{j^*} = \begin{pmatrix}
1 & 0 & 0 & ... & 0 & ... & ... & 0 \\
0 & 1 & 0 & ... & -1_{2,j^*+1} & 0 & ... & 0 \\
0 & 0 & 1 & ... & -1_{3,j^*+1} & 0 & ... & 0 \\
0 & 0 & 0 & ... & +1_{j^*+1,j^*+1} & 0 & ... & 0 \\
... & ... & ... & ... & ... & ... & ... & ... \\
0 & 0 & 0 & ... & 0 & 0 & ... & 1
\end{pmatrix}$$

Once we get the covariance matrix of the vector

$$(u_1 - u_{j^*}, u_2 - u_{j^*}, ..., u_{j^*-1} - u_{j^*}, u_{j^*+1} - u_{j^*}, ..., u_J - u_{j^*}, v, \epsilon)'$$

we can estimate

$$\Pr\left[\bigcap_{j' \neq j^*} \left(u_{j'} - u_{j^*} \leq f\left(j^*\right) - f\left(j'\right)\right)\right]$$

by using a GHK simulation procedure. Note that the first stage of the econometric model corresponds to the estimation of a multinomial probit model (MNP). Within the MNP framework, the choice probabilities $\Pr(j)$ do not have a closed-form expression.[11] Hence we use the GHK simulator which is usually considered to be the most accurate probit simulator to approximate the probabilities of choosing each field of study.[12] As it is detailed in the following section devoted to data, estimations for the current version of the paper are based on $J = 3$ aggregated majors. Thus, in stage 1, each choice probability is expressed as a double integral which can be evaluated using usual integration procedures (such as quadrature methods).

3.4 Identification

In order to identify the model without relying only on functional forms, we use exclusion restrictions. The covariates indicating the father's and mother's professions (in 1998) as well as the age of the student in 6^{th} and 12^{th} grades, are included in the list of regressors affecting the choice of the length of studies and excluded from the earnings equation. Furthermore, the dummy *Ile de France* variable is only used to explain earnings : it is excluded from the schooling choice equations.

4 Data

The model presented above is estimated using French data coming from the "Génération 98" survey collected by CEREQ (*Centre d'Etudes et de Recherches sur les Qualifications,* Marseille).[13] This survey consists of a large sample of 55,000 individuals who left the French educational system in 1998 and were interviewed three years later, in 2001. In the original sample, education levels range from the lowest to the highest, respectively referred to as "Level VI" and "Level I" in the French

[11] Each choice probability is a $J - 1$ dimensional integral which must be evaluated numerically.

[12] See Hajivassiliou et al. (1996) for a comparison of probit simulators.

[13] These data have been previously used by Brodaty, Gary-Bobo and Prieto (2006), who estimate a structural model of individual educational investments in presence of students' attitudes toward risk.

qualification nomenclature. The main advantage of this database for our approach lies in the fact that it includes information both on individuals' educational trajectory and on their labor market histories (over the three first years following the exit from the educational system). Furthermore, the survey provides us with a set of individual covariates which are used as controls in our estimation procedure, such as gender, place of birth, nationality, parents' profession, and residence when leaving the educational system. We also use data from the "Génération 92" survey (*CEREQ*, Marseille) to compute students' earnings expectations. Most of the individual covariates observed in the "Génération 98" survey are also provided by this sample of 26,359 individuals who left the French educational system in 1992, six years before the individuals we are interested in. Hence, we compute the students' earnings expectations using predicted earnings estimated with the "Génération 92" data, as exposed in the preceding section.

Our subsample of interest is constituted of respondents having at least passed the national high school final examination successfully:[14] it is then restricted to 14,365 individuals. Furthermore, within this selected sample, we focused on the 4,213 individuals having attended university except medicine faculties and IUT ("*Institut Universitaires de Technologie*", which are two-years vocational colleges). This sample selection was made in order to keep an homogeneous set of post-secondary tracks, both in terms of selection and possible length of studies.

University studies are aggregated into three broad fields: "Sciences", "Humanities and Social Sciences" (including art studies) and "Management, Economics and Law". We then consider five different educational levels (i.e. degrees) that may be reached within each major. They are respectively denoted by "dropout" (less than two years of college), "two years of college", "BA degree" ("Licence" in French), "MA degree" ("Maîtrise") and "Graduate" (more than four years after High School). Tables 1 and 2 below provide basic descriptive statistics for the selected subsample.

We cross our main variables of interest (post-secondary track, length of studies, and labor marker wages) with several individual characteristics. We also study the associations between the variables of interest which are endogenous variables in the structural model exposed above. Tables 10 to 13 (reported in the Appendix) provide a descriptive outlook for the determinants of universitary schooling choices in France.

We first focus on the choice of the study field. Tables 10 and 11 show that this choice is related with gender, age in 6^{th} grade and age in 12^{th} grade,[15] parents' nationality and profession.

[14] In France, this national exam is called "*baccalauréat*".

[15] These variables can be seen as proxies for the individual schooling ability.

Noteworthy, male students are more likely to attend majors in Sciences while female students are more likely to attend majors in Humanities and Social Sciences. There is also a high statistical association between students' age in 6^{th} grade and the chosen field: individuals who were above the "normal" age in 6^{th} grade are less likely to attend a major in Science, while they are more likely to attend a major in Law, Economics and Management. The age in 12^{th} grade is also, to a lower extent, correlated with the choice of the major : individuals who are above 18 when getting their *Baccalaureat* are less likely to attend a major in Sciences, while they are more likely to study in Humanities and Social Sciences.

Parental characteristics also seem to play in important role on the choice of the major. Noteworthy, students whose at least one parent is not French, are less likely to study Sciences. Parents' professions are also correlated with the choice of the major: students whose father is a farmer are more likely to study in Sciences, while they are less likely to study in Humanities and Social Sciences, or in Law, Economics and Management. Also note that individuals whose father is a blue-collar worker are more likely to attend a major in Human and Social Sciences, and less likely to attend a major in Sciences.[16] Table 10 also shows a strong correlation between the chosen field and the length of studies. While students studying in Sciences are more likely to complete high level studies ("Graduate" level), those studying in Humanities and Social Sciences are much more likely to drop out during the first two years of college.

Besides, all individual characteristics considered here are correlated with the length of studies (see tables 12 and 13).Noteworthy, the individual age in 12^{th} grade is negatively correlated with the length of studies.

Finally, the higher the level, the larger the mean of log earnings (table 3 reported below). There are significant differences in average earnings associated with different majors. However, it is less pronounced than the difference between long and short studies. We find that Sciences ranks first, followed by Law, Economics and Management, and finally Humanities and Social Sciences. There is no significant difference between these fields either in terms of the number of months after the first job, or in terms of the contract of the first job (see Tables 14 and 15 reported in the Appendix). Contrary to the field of studies, the level of the degree individuals get seems to have a crucial effect on earnings. And as expected, the mean log-earning is greater in the Paris region, as well as for men (see Table 16 in the Appendix).

[16]Mother's profession is associated with the field of study in a similar way.

4.1 Descriptive Sample Statistics

Table 1: Descriptive statistics: majors and levels of education

	Number	Per cent
University fields		
Sciences	1,093	25.94
Humanities and Social Sciences	1,719	40.80
Law, Economics and Management	1,401	33.25
Post Secondary Education Level		
Dropout	1,359	32.26
Two years of college	479	11.37
Licence (BA degree)	693	16.45
Maîtrise (MA degree)	741	17.59
Post Maîtrise (Graduates)	941	22.34
Baccalaureat		
General	3,421	81.28
Technological	655	15.56
Vocational	133	3.16
Secondary schooling track		
L	1,013	24.89
ES	963	23.66
S	1,439	35.36
ST,SMS	655	16.09

Table 2: Descriptive statistics: covariates

	Number	Per cent
Gender		
Male	1,746	41.44
Female	2,467	58.56
Born abroad		
No	4,123	97.86
Yes	90	2.14
Age in 6^{th} grade		
≤ 10	402	9.54
11	3,542	84.07
≥ 12	269	6.38
Age in 12^{th} grade		
≤ 17	537	12.75
18	2,011	47.73
19	1,078	25.59
≥ 20	587	13.93
Parents' nationality		
Mother or father is not french	338	8.02
Both parents are french	3,875	91.98
Father's profession (in 1998)		
Farmer	157	4.13
Tradesman	457	12.02
Executive	1,153	30.33
Technician	435	11.44
White-collar	926	24.36
Blue-collar	674	17.73
Mother's profession (in 1998)		
Farmer	84	2.11
Tradesman	178	4.47
Executive	738	18.54
Technician	233	5.85
White-collar	2,012	50.54
Blue-collar	268	6.73
Housewife	468	11.76

Table 3: Mean of log earnings according to the length and the field of studies

Field	Length	Mean of monthly log earnings
	Dropout	5.97
	Two years of college	6.18
	Licence (BA degree)	6.27
	Maitrise (MA degree)	6.36
	Post Maitrise (Graduates)	6.75
Sciences		6.54
Humanities and Social Sciences		6.08
Law. Economics and Management		6.34
Sciences	Dropout	6.16
	Two years of college	6.39
	Licence (BA degree)	6.43
	Maitrise (MA degree)	6.52
	Post Maitrise (Graduates)	6.85
Humanities and Social Sciences	Dropout	5.91
	Two years of college	6.04
	Licence (BA degree)	6.20
	Maitrise (MA degree)	6.10
	Post Maitrise (Graduates)	6.41
Law. Economics and Management	Dropout	5.94
	Two years of college	6.22
	Licence (BA degree)	6.29
	Maitrise (MA degree)	6.44
	Post Maitrise (Graduates)	6.84
		6.29

5 Results

Tables 17 to 21 (reported in the Appendix) give the parameter estimates of the model. To get a more precise view of the effect of expected wages on the choice of the field and the length of studies, we run some simulation exercises. Several shocks have been imposed to expected earnings. We use the parameter estimates to simulate how these changes in the expected earnings affect post-secondary schooling choices, simultaneously in terms of the field and the length of studies. The first set of simulations relies on a 10% increase or decrease in the expected earnings associated with a given field of study j (tables 4 to 6 below).[17] Each length of study is equally affected by this type of shock.

The second set of simulations consists in a 10% increase in the expected earnings associated with a given length of study, each field of study being equally affected. Then, we examine whether the propensity to continue studies is affected by such shocks. We concentrate on the levels $k = 1$ and $k = 4$ (tables 8 and 9 below).

We tend to believe that a rather large increase in the expected earnings associated with a given field will not induce students to change significantly their preferred field of studies: people have specific abilities and tastes for, let's say, humanities or sciences, which is difficult to change. On the contrary, the propensity to conduct long studies in a given field may be counterbalanced by a shock on expected earnings, and it seems interesting to compare students' choice sensitivity to a shock on expected earnings associated with short post-secondary studies, with their choice sensitivity to a shock on expected earnings associated with long studies.

First, comparing for each field and each length of studies the observed proportion with the predicted probability allows us to evaluate the fit of the model. We find that both the actual distribution of fields and lengths of studies are very similar. Thus, we can be quite confident with the simulations results that will now be discussed.

The effect of an increase or a decrease in expected earnings associated with a given major is different according to the major considered. The most important impact concerns Humanities and Social Sciences. A 10% increase leads to a significant change in the sample distribution: the proportion of students attending a major in this field increases by 7.4 points. the decreases in the other proportions are equally distributed: -3.72 points in Sciences and -3.67 points in Law, Economics and Management. On the contrary, a 10% increase in expected earnings of students graduated in Sciences only affects the proportion of students attending a major in Sciences or in Humanities: the proportion of students in Sciences increases by 3.64 points, while the proportion in Humanities decreases by 3.23 points. Law is roughly not affected. A 10% increase in expected earnings associated with a degree in Law has the same effect: a 3.92 point increase in the proportion of students attending a major in Law, a 3.46 points decrease in the proportion of students attending a major in Humanities, while Sciences are not affected. It is important to note that these estimated increases have to be compared with the initial sample proportion. Absolute increases of 7.4 points in Humanities, 3.64 points in Sciences

[17] Simulating both an increase and a decrease enables us to see whether the quantitative effect is different when earnings come up or down.

and 3.92 points in Law correspond to relative increases equal to 18% in Humanities, 14.2% in Sciences and 11.7% in Law.[18]

Table 7 reports the simulated proportions of students in the different majors and levels after simultaneously decreasing by 18% and increasing by 30% respectively sciences and humanities and social sciences expected earnings. This simulation is tantamount to cancelling, on average, the earning gaps between the three broad university fields. Results show that the simulation has a very significant and quantitatively large impact on the proportions of students accross the different fields. Cancelling the earning gaps between the university fields leads to a strong 24.3 points increase in the proportion of students choosing to study Humanities and Social Sciences, and simultaneously to 14.8 and 9.5 points decreases in the proportions of students respectively studying Sciences and Law, Economics and Management.

We find effects with the same magnitude when simulating a 10% increase in the expected earnings of two years of college (whatever the field of studies). This simulated shock increases by 3.94 points the proportion of individuals who decide to complete two years or less of college (see Table **??**). On the contrary, the same increase applied to the expected earnings of Graduates leads to no significant change in the proportions of students conducting long studies (see Table 9). For some individuals, the cost associated with long studies may be too high to make them change their choice concerning their educational level. But on the contrary, people who have the ability to do long studies may decide to drop earlier if the wage incentive is strong enough.

[18] Namely $\frac{p^s - p^p}{p^p}$.

Table 4: Simulation of a 10% variation in expected earnings of the Sciences major

	Observed Probability	Predicted Probability	$(p^S - p^P)$	$\widehat{\sigma}_{(p^S - p^P)}$ Standard error
Sciences				
	10% *increase*			
Sample distribution				
Sciences	26.07	25.60	3.64	1.48
Humanities and Social Sciences	40.55	40.98	-3.23	1.39
Law. Economics and Management	33.38	33.42	-0.41	0.16
Dropout	31.91	39.54	-1.04	0.14
Two years of college	11.28	11.03	-0.23	0.02
Licence (BA degree)	16.51	15.13	-0.19	0.04
Maitrise (MA degree)	17.77	15.55	0.09	0.06
Post Maitrise (Graduates)	22.54	18.76	1.36	0.14
	10% *decrease*			
Sample distribution				
Sciences	26.07	25.60	-3.85	1.44
Humanities and Social Sciences	40.55	40.98	3.50	1.41
Law. Economics and Management	33.38	33.42	0.35	0.14
Dropout	31.91	39.54	0.90	0.08
Two years of college	11.28	11.03	0.21	0.02
Licence (BA degree)	16.51	15.13	0.16	0.04
Maitrise (MA degree)	17.77	15.55	-0.10	0.05
Post Maitrise (Graduates)	22.54	18.76	-1.16	0.14

Source: Génération 1992 and Génération 1998 (CEREQ)

Note: p^S and p^P denote the predicted probabilities after and before the simulation, respectively.

Table 5: Simulation of a 10% variation in expected earnings of the Humanities and Social Sciences major

	Observed Probability	Predicted Probability	$(p^S - p^P)$	$\widehat{\sigma}_{(p^S - p^P)}$ Standard error
Humanities and Social Sciences				
		10% *increase*		
Sample distribution				
Sciences	26.07	25.60	-3.72	1.50
Humanities and Social Sciences	40.55	40.98	7.39	2.78
Law. Economics and Management	33.38	33.42	-3.67	1.33
Dropout	31.91	39.54	-2.75	0.38
Two years of college	11.28	11.03	-0.13	0.04
Licence (BA degree)	16.51	15.13	0.24	0.06
Maitrise (MA degree)	17.77	15.55	0.77	0.11
Post Maitrise (Graduates)	22.54	18.76	1.87	0.26
		10% *decrease*		
Sample distribution				
Sciences	26.07	25.60	4.06	1.64
Humanities and Social Sciences	40.55	40.98	-8.01	2.96
Law. Economics and Management	33.38	33.42	3.95	1.40
Dropout	31.91	39.54	2.45	0.27
Two years of college	11.28	11.03	0.02	0.03
Licence (BA degree)	16.51	15.13	-0.34	0.06
Maitrise (MA degree)	17.77	15.55	-0.74	0.09
Post Maitrise (Graduates)	22.54	18.76	-1.39	0.16

Source: Génération 1992 and Génération 1998 (CEREQ)

Note: p^S and p^P denote the predicted probabilities after and before the simulation, respectively.

Table 6: Simulation of a 10% variation in expected earnings of the Law, Economics and Management major

	Observed Probability	Predicted Probability	$(p^S - p^P)$	$\widehat{\sigma}_{(p^S-p^P)}$ Standard error
Law. Economics and Management				
		10% *increase*		
Sample distribution				
Sciences	26.07	25.60	-0.47	0.18
Humanities and Social Sciences	40.55	40.98	-3.46	1.28
Law. Economics and Management	33.38	33.42	3.92	1.37
Dropout	31.91	39.54	-0.57	0.28
Two years of college	11.28	11.03	0.13	0.03
Licence (BA degree)	16.51	15.13	0.21	0.04
Maitrise (MA degree)	17.77	15.55	0.19	0.08
Post Maitrise (Graduates)	22.54	18.76	0.04	0.18
		10% *decrease*		
Sample distribution				
Sciences	26.07	25.60	0.42	0.16
Humanities and Social Sciences	40.55	40.98	3.72	1.35
Law. Economics and Management	33.38	33.42	-4.14	1.40
Dropout	31.91	39.54	0.27	0.38
Two years of college	11.28	11.03	-0.16	0.03
Licence (BA degree)	16.51	15.13	-0.18	0.07
Maitrise (MA degree)	17.77	15.55	-0.09	0.12
Post Maitrise (Graduates)	22.54	18.76	0.16	0.20

Source: Génération 1992 and Génération 1998 (CEREQ)

Note: p^S and p^P denote the predicted probabilities after and before the simulation, respectively.

Table 7: Simulation: on average, no earning gaps between the majors

	Observed Probability	Predicted Probability	$(p^S - p^P)$	$\widehat{\sigma}_{(p^S - p^P)}$ Standard error
Sample distribution				
Sciences	26.07	25.60	-14.85	4.29
Humanities and Social Sciences	40.55	40.98	24.33	7.54
Law. Economics and Management	33.38	33.42	-9.48	3.35
Dropout	31.91	39.54	-7.85	1.59
Two years of college	11.28	11.03	-0.55	0.23
Licence (BA degree)	16.51	15.13	0.40	0.15
Maitrise (MA degree)	17.77	15.55	2.01	0.33
Post Maitrise (Graduates)	22.54	18.76	5.99	1.57

Source: Génération 1992 and Génération 1998 (CEREQ)

Note: p^S and p^P denote the predicted probabilities after and before the simulation, respectively.

Table 8: Simulation of a 10% increase in expected earnings associated with two years of college

	Observed Probability	Predicted Probability	$(p^S - p^P)$	$\widehat{\sigma}_{(p^S-p^P)}$ Standard error
	10% increase			
Sample distribution				
Sciences	26.07	25.60	0.53	0.19
Humanities and Social Sciences	40.55	40.98	-0.72	0.25
Law. Economics and Management	33.38	33.42	0.19	0.07
Dropout	31.91	39.54	3.86	0.68
Two years of college	11.28	11.03	0.08	0.04
Licence (BA degree)	16.51	15.13	-0.36	0.10
Maitrise (MA degree)	17.77	15.55	-1.00	0.19
Post Maitrise (Graduates)	22.54	18.76	-2.58	0.40

Source: Génération 1992 and Génération 1998 (CEREQ)

Note: p^S and p^P denote the predicted probabilities after and before the simulation, respectively.

Table 9: Simulation of a 10% increase in expected earnings of Graduates

	Observed Probability	Predicted Probability	$(p^S - p^P)$	$\widehat{\sigma}_{(p^S - p^P)}$ Standard error
		10% *increase*		
Sample distribution				
Sciences	26.07	25.60	0.23	0.17
Humanities and Social Sciences	40.55	40.98	0.11	0.19
Law. Economics and Management	33.38	33.42	-0.34	0.10
Dropout	31.91	39.54	0.46	0.36
Two years of college	11.28	11.03	0.01	0.01
Licence (BA degree)	16.51	15.13	-0.04	0.03
Maitrise (MA degree)	17.77	15.55	-0.11	0.09
Post Maitrise (Graduates)	22.54	18.76	-0.32	0.25

Source: Génération 1992 and Génération 1998 (CEREQ)

Note: p^S and p^P denote the predicted probabilities after and before the simulation, respectively.

References

AGHION P. and COHEN E. (2004) : *Education and Growth*, French Council of Economic Analysis Report n 46, La Documentation Française, Paris.

ALTONJI J. (1993) : “The Demand for and Return to Education when Education Outcomes Are Uncertain”, *Journal of Labor Economics*, vol. 11, 48-83.

ARCIDIACONO P. (2004) : “Ability Sorting and the Returns to College Major”,*Journal of Econometrics*, vol. 121, 343-375.

ARCIDIACONO P. (2005) : “Affirmative Action in Higher Education: How do Admission and Financial Aid Rules Affect Future Earnings?”,*Econometrica*, vol. 73, 1477-1524.

BELZIL C. and HANSEN J. (2002) : “Unobserved Ability and the Return to Schooling”,*Econometrica*, vol. 70, 2075-2091.

BERGER M. (1988) : “Predicted Future Earnings and Choice of College Major”, *Industrial and Labor Relations Review*, vol. 41, 418-29.

BOUDARBAT B. and MONTMARQUETTE C. (2007) : “Choice of Fields of Study of Canadian University Graduates: The Role of Gender and their Parents’ Education”, *IZA Discussion Paper* No.2552.

BRODATY T., GARY-BOBO R. and PRIETO A. (2006): “Risk Aversion and Human Capital Investment: A Structural Econometric Model", *CEPR Discussion Paper* 5694, London.

BUCHINSKY M. and LESLIE P. (2000) : “Educational Attainment and the Changing U.S. Wage Structure : Dynamic Implications Without Rational Expectations", Brown University, mimeo.

CAMERON S.V. and HECKMAN J.J. (1998) : “Life Cycle Schooling and Dynamic Selection Bias : Models and Evidence for Five Cohorts of American Males”, *Journal of Political Economy*, vol. 106, 262-333.

CAMERON S.V. and HECKMAN J.J. (2001) : “The Dynamics of Educational Attainment for Black, Hispanic, and White Males”, *Journal of Political Economy*, vol. 109, 455-499.

ECKSTEIN Z. and WOLPIN K.I. (1999) : “Why Youths Drop Out of High School: The Impact of Preferences, Opportunities and Abilities”,*Econometrica*, vol. 67, 1295-1339.

FREEMAN R. (1971) : *The Market for College-Trained Manpower*. Cambridge, MA: Harvard University Press.

FREEMAN R. (1975) : “Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948-1973”,*American Economic Review*, vol. 65, 27-39.

HAJIVASSILIOU V., McFADDEN D. and RUUD P. (1996) : “Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives: Theoretical and Computational Results”, *Journal of Econometrics*, vol. 72, 85-134.

KEANE M.P. and WOLPIN K.I. (1997) : “The Career Decisions of Young Men”, *Journal of Political Economy*, vol. 105, 473-522.

KEANE M.P. and WOLPIN K.I. (2001) : “The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment”, *International Economic Review*, vol. 42, 1051-1103.

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

TRAIN K. (2003) : *Discrete Choice Methods with Simulation*. Cambridge University Press.

WILLIS R. and ROSEN S. (1979) : “Education and Self Selection”, *Journal of Political Economy*, vol. 87, S7-36.

A Other descriptive statistics

Table 10: Distribution of majors according to given covariates (in %, beginning)

	Sciences	Humanities and Social Sciences	Law, Economics and Management
Gender			
Male	39.40	29.32	31.27
Female	16.42	48.93	34.66
Born Abroad			
No	25.93	40.84	33.23
Yes	26.67	38.89	34.44
Age in 6^{th} grade			
≤ 10	29.60	38.81	31.59
11	26.23	41.22	32.55
≥ 12	16.73	38.29	44.98
Age in 12^{th} grade			
≤ 17	27	41.53	31.47
18	29.39	37.69	32.92
19	21.99	43.78	34.23
≥ 20	20.44	45.32	34.24
Parents' nationality			
Mother or father not french	16.27	44.08	39.64
Both parents french	26.79	40.52	32.70
Father's profession (in 1998)			
Farmer	33.76	36.31	29.94
Tradesman	27.35	38.29	34.35
Executive	29.66	38.94	31.40
Technician	27.13	40.23	32.64
White-collar	24.84	40.82	34.34
Blue-collar	20.62	44.96	34.42
Mother's profession (in 1998)			
Farmer	34.52	39.29	26.19
Tradesman	28.09	37.64	34.27
Executive	28.86	40.24	30.89
Technician	25.32	43.35	31.33
White-collar	25.05	42.15	32.80
Blue-collar	22.39	41.42	36.19
Housewife	25.43	36.75	37.82
Education Level			
Dropout	23.33	40.61	28.98
Two years of college	10.80	12.16	10.85
Licence (BA degree)	11.07	22.45	13.28
Maîtrise (MA degree)	16.19	12.39	25.05
Post Maîtrise (Graduates)	38.61	12.39	21.84

Table 11: Distribution of majors according to given covariates(in %,end)

	Sciences	Humanities and Social Sciences	Law, Business and Management
Baccalauréat			
General	28.00	41.33	30.66
Technological	18.02	40.15	41.83
Vocational	12.78	29.32	57.89
Secondary schooling track			
L	1.78	77.59	20.63
ES	3.74	40.71	55.56
S	62.68	16.26	21.06
ST,SMS	18.02	40.15	41.83

Source: Génération 1998 (CEREQ, Marseille)

Table 12: Distribution of levels according to given covariates(in %,beginning)

	Dropout	Two years of college	BA	MA	Graduates
Gender					
Male	32.02	12.20	13.63	16.27	25.89
Female	32.43	10.78	18.44	18.52	19.82
Born Abroad					
No	32.16	11.47	16.61	17.66	22.10
Yes	36.67	6.67	8.89	14.44	33.33
Age in 6^{th} *grade*					
≤ 10	21.64	9.20	12.19	20.90	36.07
11	31.96	11.55	17.00	17.73	21.77
≥ 12	52.04	12.27	15.61	10.78	9.29
Age in 12^{th} *grade*					
≤ 17	20.11	9.87	17.32	20.48	32.22
18	24.47	10.64	16.96	19.34	28.59
19	40.35	12.15	16.79	17.07	13.64
≥ 20	55.20	13.80	13.29	9.88	7.84
Parents' nationality					
Mother or father is not french	51.48	12.43	8.28	14.20	13.61
Both parents are french	30.58	11.28	17.16	17.88	23.10
Father's profession (in 1998)					
Farmer	23.57	10.19	24.20	22.93	19.11
Tradesman	28.45	14	17.72	18.38	21.44
Executive	23.16	8.59	15.52	18.21	34.52
Technician	26.67	11.95	18.39	20.69	22.30
White-collar	38.44	12.31	15.44	17.28	16.52
Blue-collar	45.55	12.91	14.84	14.69	12.02
Mother's profession (in 1998)					
Farmer	26.19	9.52	22.62	22.62	19.05
Tradesman	25.84	14.61	18.54	15.73	25.28
Executive	21.82	10.43	17.89	17.89	31.98
Technician	34.33	11.16	18.45	12.45	23.61
White-collar	35.39	11.68	15.26	18.14	19.53
Blue-collar	43.66	8.58	17.54	17.16	13.06
Housewife	33.55	11.97	14.74	16.24	23.50

Table 13: Distribution of levels according to given covariates(in %,end)

	Dropout	Two years of college	BA	MA	Graduates
Baccalauréat					
General	25.29	11.28	17.68	19.56	26.19
Technological	58.93	12.67	11.91	9.77	6.72
Vocational	80.45	6.02	7.52	5.26	0.75
Secondary schooling track					
L	34.75	12.24	23.79	16.19	13.03
ES	26.27	11.01	18.28	25.75	18.69
S	18.00	10.84	13.00	17.79	40.38
ST,SMS	58.93	12.67	11.91	9.77	6.72

Source: Génération 1998 (CEREQ, Marseille)

Table 14: Contract of the first job (in %)

Major	Type of contract	Dropout	Two years of college	BA	MA	Graduates
HSS	Short term contract	35.46	37.13	40.94	37.32	44.50
	Long term contract	21.07	18.32	17.32	17.70	20.10
	Emploi jeune	10.39	15.35	9.97	11.96	5.74
	Temping	14.09	9.41	-	-	-
	Civil servant	-	-	13.39	14.83	19.62
LEM	Short term contract	34.68	35.33	34.07	33.81	42.24
	Long term contract	18.48	26.67	29.67	33.81	39.60
	Emploi jeune	-	10.00	-	-	-
	Temping	18.73	7.33	8.79	11.17	4.29
	Civil servant	-	-	9.34	6.59	7.26
	Contrat de qualification	7.85	-	-	-	-
S	Short term contract	33.33	33.33	40.34	39.20	42.72
	Long term contract	15.66	26.32	21.01	32.39	41.29
	Temping	22.09	21.05	12.61	10.80	2.39
	Emploi jeune	8.84	8.77	-	-	-
	Civil servant	-	-	10.08	6.82	6.44

Table 15: Number of months before the first job

		Mean of the number of months before the first job
Sciences	Dropout	6.23
	2 years	5.55
	BA degree	4.79
	MA degree	5.47
	Graduates	3.55
Humanities and Social Sciences	Dropout	5.79
	2 years	5.05
	BA degree	6.49
	MA degree	6.36
	Graduates	3.53
Law. Economics and Management	Dropout	5.61
	2 years	5.01
	BA degree	4.41
	MA degree	5.42
	Graduates	3.56

Source: Génération 1998 (CEREQ, Marseille)

Table 16: Mean of monthly log earnings according to residence, parents' nationality and gender

	Mean of monthly log earnings
Out of the Paris region	6.22
In the Paris region	6.68
At least one of the parents born abroad	6.21
Both parents born in France	6.30
Man	6.50
Woman	6.13

Source: Génération 1998 (CEREQ, Marseille)

B Parameter estimates

Table 17: Equation for the choice of major (beginning)

	Estimate	Standard Error
Sciences	*Ref*	*Ref*
Humanities and Social Sciences		
Father's profession (in 1998)		
Farmer or Tradesman	-0.1998	0.1290
Executive	*Ref*	*Ref*
Technician	-0.2678	0.2056
White-collar	-0.0861	0.1345
Blue-collar	-0.2033	0.2533
Unknown	-0.0623	0.1399
Mother's profession (in 1998)		
Farmer or Tradesman	-0.0948	0.2279
Executive	*Ref*	*Ref*
Technician	-0.0934	0.1489
White-collar	-0.2961	0.2514
Blue-collar	0.1898	0.1570
Unknown	0.2248	0.2885
Born abroad	-0.1906	0.2162
Man	*Ref*	*Ref*
Woman	-0.3083	0.3120
Mother and father born in France	-0.0435	0.1607
Both parents are french	-0.2874	0.3327
Age in 6^{th} *grade*		
Less than 10 years old	-0.1626	0.1279
11 years old	*Ref*	*Ref*
12 years old or more	-0.3525	0.1862
Age in 12^{th} *grade*		
≤ 17	-0.0520	0.1534
18	*Ref*	*Ref*
19	-0.0800	0.2614
≥ 20	-0.2385	0.1177
Baccauláuréat		
General, sciences	*Ref*	*Ref*
General, humanities	-0.1737	0.2184
General, economics	-0.3609	0.1993
Vocational or technological	-0.5718	0.2356

Table 18: Equation for the choice of major (end)

	Estimate	Standard Error
Law. Economics and Management		
Father's profession (in 1998)		
Farmer or Tradesman	0.4458	0.0949
Executive	*Ref*	*Ref*
Technician	0.4322	0.1766
White-collar	-0.1659	0.1410
Blue-collar	-0.1532	0.2661
Unknown	-0.5339	0.2005
Mother's profession (in 1998)		
Farmer or Tradesman	-0.8193	0.2842
Executive	*Ref*	*Ref*
Technician	-0.0737	0.1653
White-collar	-0.0077	0.2813
Blue-collar	-0.1153	0.1555
Unknown	0.3949	0.2246
Born abroad	0.0120	0.1290
Man	*Ref*	*Ref*
Woman	-0.0773	0.2608
Mother and father born in France	0.1070	0.1014
Both parents are french	-0.2022	0.1950
Age in 6^{th} grade		
Less than 10 years old	0.2038	0.1276
11 years old	*Ref*	*Ref*
12 years old or more	-0.3683	0.2090
Age in 12^{th} grade		
≤ 17	2.6661	0.1425
18	*Ref*	*Ref*
19	2.3638	0.3234
≥ 20	2.2853	0.1382
Baccaulauréat		
General, sciences	*Ref*	*Ref*
General, humanities	3.9255	0.1622
General, economics	1.5372	0.1254
Vocational or technological	2.5208	0.2053

Source: Génération 1992 and Génération 1998 (CEREQ, Marseille)

Table 19: Equation for the choice of the educational level

	Estimate	Standard Error
Father's profession (in 1998)		
Farmer or Tradesman	-0.1236	0.0817
Executive	*Ref*	*Ref*
Technician	-0.0856	0.0957
White-collar	-0.3003	0.0763
Blue-collar	-0.3392	0.1007
Unknown	-0.1608	0.1108
Mother's profession (in 1998)		
Farmer or Tradesman	0.0441	0.1329
Executive	*Ref*	*Ref*
Technician	-0.1763	0.1115
White-collar	-0.0480	0.0788
Blue-collar	-0.1656	0.1134
Unknown	0.0801	0.1016
Born abroad	0.0356	0.1878
Man	*Ref*	*Ref*
Woman	0.2082	0.1274
Mother and father born in France	-0.3241	0.1202
Both parents are french	0.2603	0.1520
Age in 6^{th} grade		
Less than 10 years old	0.3303	0.1083
11 years old	*Ref*	*Ref*
12 years old or more	0.4009	0.1034
Age in 12^{th} grade		
≤ 17	0.0253	0.0880
18	*Ref*	*Ref*
19	-0.3447	0.0569
≥ 20	-0.6359	0.0783
Baccaulauréat		
General, sciences	*Ref*	*Ref*
General, humanities	-0.6364	0.0818
General, economics	-0.3556	0.0717
Vocational or technological	-1.0182	0.0828
$\overline{E}^{(j,k=0)}$	1.3199	0.0235
$\overline{E}^{(j,k=1)} - \overline{E}^{(j,k=0)}$	0.0239	0.0877
$\overline{E}^{(j,k=2)} - \overline{E}^{(j,k=1)}$	1.3329	0.1843
$\overline{E}^{(j,k=3)} - \overline{E}^{(j,k=2)}$	0.5193	0.1281
$\overline{E}^{(j,k=4)} - \overline{E}^{(j,k=3)}$	-0.1916	0.1258

Source: Génération 1992 and Génération 1998 (CEREQ, Marseille)

Table 20: Earnings equation

	Estimate	Standard Error
Constant	5.9863	0.1141
Mother and father born in France	-0.0179	0.0858
Man	*Ref*	*Ref*
Woman	-0.3455	0.1058
Born abroad	-0.0074	0.2203
Field of studies		
Sciences	*Ref*	*Ref*
Humanities and Social Sciences	-0.0975	0.1032
Law. Economics and Management	-0.0398	0.1173
Level of studies		
Dropout	*Ref*	*Ref*
Two years of college	0.3332	0.1497
Licence (BA degree)	0.4382	0.1467
Maitrise (MA degree)	0.6227	0.2143
Post Maitrise (Graduates)	1.2447	0.2403
Interactions between field and level		
Humanities and Social Sciences		
Dropout	*Ref*	*Ref*
Two years of college	-0.1126	0.2761
Licence (BA degree)	0.0654	0.2086
Maitrise (MA degree)	-0.2487	0.2276
Post Maitrise (Graduates)	-0.1707	0.2033
Law. Economics and Management		
Dropout	*Ref*	*Ref*
Two years of college	-0.0023	0.2201
Licence (BA degree)	0.0603	0.2169
Maitrise (MA degree)	0.0252	0.1952
Post Maitrise (Graduates)	0.1416	0.1712
Interactions between Female gender and level		
Dropout	*Ref*	*Ref*
Two years of college	0.0946	0.2314
Licence (BA degree)	0.0962	0.2228
Maitrise (MA degree)	0.2731	0.1650
Post Maitrise (Graduates)	-0.0985	0.1851

Source: Génération 1992 and Génération 1998 (CEREQ, Marseille)

Table 21: Other parameters

	Parameter	Standard Error
Covariance matrix of residuals	$\begin{pmatrix} 1 & 0 & 0 & -0.25 \\ 0 & 1 & 2.27 & 0 \\ 0 & 2.27 & 6.93 & 0 \\ -0.25 & 0 & 0 & 1.77 \end{pmatrix}$	
Thresholds		
s_2	8.2471	0.0268
s_3	8.6002	0.0244
s_4	9.0882	0.0198
s_5	9.6843	0.0279
α	0.1659	0.0583

Source: Génération 1992 and Génération 1998 (CEREQ)