Choosing the Field of Study in Post-Secondary Education: Do Expected Earnings Matter?*

Magali Beffy† Denis Fougère‡ Arnaud Maurel§

2nd March 2009

Abstract

This paper examines the determinants of the choice of the major when the length of studies is uncertain, by using a framework in which students entering post-secondary education are assumed to anticipate their future earnings. For that purpose, we use French data coming from the 1992 and 1998 *Generation* surveys collected by the Centre for Study and Research on Qualifications. Our econometric approach is based on a semi-structural three-equations model, which is identified thanks to some exclusion restrictions. We exploit in particular exogenous variations across the business cycle in the returns to each major in order to identify the causal effect of expected earnings on the probability to choose each major. Once graduated from high school, individuals are supposed to choose their post-secondary major in which they reach a certain level of education. Finally they enter the labor market. Relying on a three-component mixture distribution, we account for correlation between the unobserved individual-specific preferences that affect the values of each post secondary field of study, the unobserved individual-specific factors that affect the equation determining the level reached within the major and the labor market earnings equation. Following Arcidiacono and Jones (2003), we use the EM algorithm with a sequential maximization step to produce consistent parameter estimates. Simulating for each

*We would like to thank Christian Belzil, Moshe Buchinsky, Nicolas Chopin, Xavier d'Haultfoeuille, Francis Kramarz, Guy Laroque, Robert Miller, Jean-Marc Robin, Gerard J. van den Berg, participants at the 2009 North American Winter Meeting of the Econometric Society, at the EALE Annual Conference (Amsterdam, September 2008), at the sixty-third European Meeting of the Econometric Society (Milan, August 2008), at the IZA eleventh European Summer School in Labor Economics (Buch, May 2008), in seminars at Paris 1 (Paris, June 2008), CREST-INSEE (Paris, January 2008) and at Université du Mans (January 2008), at the CEPR-EEEPE conference (Madrid, October 2007) and at the IZA Workshop on "Heterogeneity in Micro Econometric Models" (Bonn, June 2007) for very helpful discussions and comments.

†DEPP, CREST-INSEE, Paris, and IZA, Bonn. E-mail: magali.beffy@education.gouv.fr

‡CNRS, CREST-INSEE, Paris, CEPR, London, and IZA, Bonn. E-mail: fougere@ensae.fr

§CREST-ENSAE and PSE, Paris, and IZA, Bonn. E-mail: maurel@ensae.fr

given field of study a 10 percent increase in the expected earnings results in a statistically significant but quantitatively small impact on the allocation between fields.

JEL Classification: J24, C35, D84

1 Introduction

Over recent years, the French higher education system has been the object of much debate and sharp criticism. In a report for the French *Council of Economic Analysis*, Aghion and Cohen (2004) emphasize the main difficulties that the French post-secondary education system, and especially the French university, has to cope with. Pointing out, among others, the high dropout rate in French universities, they argue that the French post-secondary education system needs urgently to be reformed. In this context, it seems crucial to understand students' educational choices.

In this paper, we focus on the effect of expected labor market income on individual post-secondary major choices. In particular, we assess the sensitivity of students' major choices to expected earnings by estimating a semi-structural model of post-secondary educational choices. More precisely, we try to disentangle the simultaneous effects of, on the one hand, preferences and abilities, and on the other hand, expected returns, on the choice of major. In the existing applied literature, several papers explicitly consider the impact of expected labor market earnings on schooling choices. A first set of papers study these issues by using a rational expectations framework. In a seminal paper, Willis and Rosen (1979) allow the demand for college education to depend on expected future earnings[1]. 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 expected flow of future earnings than by their expected initial earnings.[2] Then, following Keane and Wolpin (1997), several econometricians have estimated structural dynamic models of schooling decisions (Cameron and Heckman, 1998, 2001; Eckstein and Wolpin

[1]On a related ground, Altonji (1993) estimates a sequential model in which schooling decisions depend on expected returns to education, without explicitly considering the choice of major.

[2]Several other articles have shown that there exists some large differences in earnings across majors in the U.S. (see, for instance, James et al., 1989; Loury and Garman, 1995; Brewer, Eide and Ehrenberg, 1999). However, none of these papers model the choice of the major itself as a function of expected earnings.

1999; Keane and Wolpin, 2001; Belzil and Hansen, 2002; Lee, 2005[3]). Their papers assume that students form rational earnings expectations conditional on schooling decisions, and that the expected earnings affect in turn their educational choices. Recently, Arcidiacono (2004, 2005) has considered sequential models of college attendance, accounting both for the demand as well as the supply side of schooling, in which the value of each major depends on the corresponding expected flow of earnings. Noteworthy is that in the literature quoted above, papers by Berger (1988) and Arcidiacono (2004, 2005) are the only ones focusing on the effect of expected earnings on the choice of major and not on the level of education. This paper builds on this literature by assuming that students face an uncertain length of studies when choosing their post-secondary major. As we will see further, including uncertainty in terms of level of education seems to be necessary to correctly account for the observed educational paths.

A second set of papers examines the validity of the rational expectations assumption in the context of educational choices. More precisely, these papers consider the specification and the estimation of schooling decision models in which the rational expectations assumption is relaxed. In particular, 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 the field of studies 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; assuming myopic expectations, the predicted earnings are computed from the wages of young individuals who have the same education level and who are currently working.

This paper contributes to the literature on the effects of expected earnings on schooling choices in several ways. First, unlike the previous papers, our approach concentrates on the effects of expected earnings on the choice of the field, in a framework in which the length of post-secondary studies is uncertain to the individual when choosing her major. Noteworthy, stylized facts appear to be consistent with such a framework.[4]. Then, after estimating our model, we obtain the elastic-

[3]Unlike the preceding papers which rely on partial equilibrium settings, Lee(2005) specifies and estimates a general equilibrium model of work, schooling and occupational choices.

[4]Indeed, descriptive statistics from the *Panel 1989* database (DEPP, French Ministry of Education) show that most students complete a final level of education which is different from the level they aimed at when entering college (see Appendix B, Table 12).

ities of major choices to expected earnings by simulating exogenous variations of earnings distribution. Another interesting feature of our paper lies in the fact that we exploit the arguably exogenous variation across the business cycle in the relative returns to each major in order to identify these elasticity parameters. It is finally the first microeconometric study devoted to these issues in France. Our study has two main limitations. First, in the absence of appropriate information allowing identification of risk-aversion coefficients, we do not consider individual attitudes towards risk.[5] We also ignore the possibility for the student to switch major during her post-secondary studies. Such a switch is potentially an endogenous event whose treatment would make the model much more complicated, and stylized facts show that this is a sensible assumption, given the broad majors we consider in the paper (see Table 11 in Appendix B).

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 presents the identification strategy and Section 6 contains the estimation and simulation results. Finally, Section 7 summarizes and concludes.

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 their field of study in which they will complete a certain (partly random) level of education. Note that we restrict our analysis to individuals who attend university. [6] 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 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.

[5] Among recent studies addressing this issue, the reader can consult papers by Belzil and Hansen (2004), Saks and Shore (2005), Brodaty, Gary-Bobo and Prieto (2006).

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

Following Heckman and Singer (1984), we assume that there are R types of individuals, with Π_r denoting the proportion of type r in the population of students.[7] Individuals are supposed to know their type. Within this framework, unobserved heterogeneity (i.e. unobserved preferences for each major, unobserved schooling ability and unobserved labor market productivity) is type-specific.

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^*.[8] We assume that this choice is made among a set of M majors. Furthermore, we assume that the chosen field (j^*) depends on the individual's expectations concerning both the education level that the student will achieve within this major (see stage 2) and his/her future labor market earnings, which are assumed to depend on his/her educational level (see stage 3). An important underlying assumption is that future earnings as well as the highest level of education reached in field j^* are partly uncertain.[9]

For a student of type r, let us denote by V_j^r the value function associated with the choice of field j $(j = 1, \ldots, M)$. This value function is assumed to be composed of two additive elements, respectively denoted by v_{0j} and v_{1j}^r. The first term v_{0j}^r represents the *intrinsic* value (i.e. the consumption value) of the major, while v_{1j}^r 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 future average (monthly) 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, for a student of type r, the value V_j^r of major j can be written as :

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

[7]Examples of econometric models of schooling decisions relying on a similar assumption can be found in Keane and Wolpin (1997, 2001), Eckstein and Wolpin (1999), Cameron and Heckman (1998, 2001), Belzil and Hansen (2002, 2004), Arcidiacono (2004, 2005) and Lee (2005).

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

[9]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 her own health status and unexpected changes in her family environment. These factors are *ex ante* unknown by the individual when choosing her major, and then revealed when attending university. Hence, at the end of stage 2, there are known by the student.

where

$$v_{1j}^r = \alpha \sum_{k\in\{0,1,\ldots,L\}} \Pr(K=k \mid r, J=j).E\left(V_{e(j,k)}^r \mid r, J=j, K=k\right)$$

$E\left(V_{e(j,k)}^r \mid r, J=j, K=k\right)$ denoting the expected flow of earnings associated with education (j,k), for a student of type r, and α being an unknown sensibility parameter to be estimated.[10]

The subcomponent v_{0j}^r can be interpreted as the non-pecuniary value of field j for a student of type r. 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}^r 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). It is also depending on a type-specific intercept $\alpha_{(1,j)}^r$ and on a random term u_j independent of $\alpha_{(1,j)}^r$. Consequently, v_{0j}^r is specified as

$$v_{0j}^r = \alpha_{(1,j)}^r + X_1'\beta_1^j + u_j$$

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 value function:

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

2.2 Stage 2: Determination of the length of studies

Once a student of type r 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 length of studies k_j^* within major j is supposed to be determined by the individual propensity $\widetilde{k}_j$ to succeed in long post-secondary studies within this

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

major.[11] More precisely, we assume that the length of studies k_j^* is generated by the following latent model:

$$k_j^* = \begin{cases} 0 & \text{if } \widetilde{k}_j^r \leq s_1 \\ \\ 1 & \text{if } s_1 < \widetilde{k}_j^r \leq s_2 \\ \vdots & \\ L & \text{if } s_L < \widetilde{k}_j^r \end{cases}$$

where $\{s_1, \ldots, s_L\}$ are latent (unknown) thresholds that correspond to the minimum ability levels required to obtain the different degrees. The latent propensity $\widetilde{k}_j^r$ is assumed to depend linearly on observable covariates X_2 (such as gender, nationality, parents' profession, etc..). It also depends on a type-specific intercept α_2^r and on an independent term v which is unknown *ex ante* by the student when he /she decides to enter college. Thus the propensity $\widetilde{k}_j^r$ is defined as:

$$\widetilde{k}_j^r \quad = \quad \alpha_2^r + X'_{2,j}\beta_2 + v$$

where α_2^r and β_2 are unknown parameters to be estimated. In this expression, $X_{2,j}$ is a vector of exogenous regressors including individual characteristics but also covariates that are specific to the major j. Namely, we allow the average proportion of college students in the same major and in the same university to affect the length of studies[12]. In the absence of variables plausibly affecting the choice of major and not the length of studies, we choose not to include major-specific dummies in $X_{2,j}$ since the related coefficients would only be identified through nonlinearities.

Note that, in our framework, the length of studies is not the number of years spent effectively in post-secondary education, but the terminal level of education that is reached by the student, whatever the time spent at the university. We should also remark that we do not account for selection of applicants by the university administration at the entry of college: 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:

[11] This framework is consistent with an ordered probit model.

[12] This variable is calculated using information coming from the *SISE* database provided by the French Ministry of Education.

individuals do not resume studies after entering the labor force. When making his/her post-secondary schooling decision in stage 1, the individual is assumed to anticipate the impact of the major and of the length of the studies on his/her future labor market earnings. In order to take both employment and nonemployment spells into account, we refer to average *earnings* as the sum of wages weighted by employment spell durations, and unemployment benefits[13] weighted by unemployment spell durations. Hence, the logarithm of the average monthly earnings in a T_{obs} years long labor market history for a worker with education (j, k) and of type r, is given by :

$$\overline{\ln w}^r_{jk} = \ln \frac{\sum_{s=1}^{N_e} w_{s,jk} l^e_s + \sum_{s'=1}^{N_u} b_{s',jk} l^u_{s'}}{T_{obs}} \tag{1}$$

with

$$T_{obs} = \sum_{s=1}^{N_e} l^e_s + \sum_{s'=1}^{N_u} l^u_{s'}$$

where N_e (respectively, N_u) is the number of observed employment (unemployment) spells in the individual labor market history, $w_{s,jk}$ is the monthly wage in the s-th employment spell, l^e_s (respectively, $b_{s',jk}$ is the monthly unemployment benefit in the s'-th unemployment spell, l^u_s) are durations of the s-th employment (respectively, unemployment) spell, and T_{obs} is the total length of the observed labor market history of the individual. By definition, we set:

$$V^r_{e(j,k)} \quad = \quad \overline{\ln w}^r_{jk} \tag{2}$$

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 they 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. For a student of type r, the average log-earnings equation is assumed to be given by:

$$\overline{\ln w}^r_{jk} = \alpha^r_3 + X'_{3(j,k)} \beta_3 + \epsilon \tag{3}$$

[13] Unemployment benefits are assumed to be equal to a constant times the former wage received when employed. This constant is taken equal to 0.7 as often done in the literature.

where $X_{3(j,k)}$ is a vector of observed characteristics that may affect labor market earnings, including post-secondary education, α_3^r represents the type-specific intercept, and ϵ denotes an independent random factor that affects the individual's earnings.

3 Econometric specification

Let us recall that the type-specific intercepts are mass points of a discrete distribution with probabilities $(\Pi_1, ..., \Pi_R)$ verifying $\sum_{r=1}^{R} \Pi_r = 1$, and that the residuals of the three equations are stochastically independent of these type-specific intercepts[14].

3.1 The econometric model

For a student of type r, the choice of the post-secondary field of study (*stage 1*) is assumed to be generated by a random utility multinomial model:

$$j^* = j \Leftrightarrow V_j^r \geq V_{j'}^r, \quad \forall j' \neq j, \quad j' = 1, 2, ..., M \tag{4}$$

with

$$\begin{aligned} V_j^r &= \alpha_{(1,j)}^r + X_1' \beta_1^j + \alpha \sum_{k \in \{0,1,\ldots,M\}} \Pr(K = k \mid r, J = j).E\left(V_{e(j,k)}^r \mid r, j, k\right) \\ &+ u_j \end{aligned}$$

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 :

$$\forall k \in \{0, 1, ..., L\}, k_j^* = k \Leftrightarrow s_k < \widetilde{k}_j^r \leq s_{k+1} \tag{5}$$

with

$$\widetilde{k}_j^r = \alpha_2^r + X_{2,j}' \beta_2 + v$$

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

$$\overline{\ln w}_{jk}^r = \alpha_3^r + X_{3(j,k)}' \beta_3 + \epsilon \tag{6}$$

where $\overline{\ln w}_{jk}^r$ is the average log-earnings defined in equation (1).

[14] Some covariates introduced in the equations may not be independent of the individual's type. It applies especially to the high school graduation track, which may be in particular related to unobserved preferences for each major. Nevertheless, conditioning the type probabilities on the high school graduation track did not change significantly our results, which are reported in the paper relying on unconditional heterogeneity distribution.

3.2 Stochastic assumptions

Residuals are supposed to be normally distributed. We assume that the random vector $(u_1, \ldots, u_M)$ affecting the choice equation, and the residuals v and ϵ entering the two other equations are independently distributed.[15] Consequently, the whole vector of residuals[16] 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:

$$\Sigma = \left(\begin{array}{c|ccccc|c} 1 & \ldots & \ldots & \ldots & \ldots & & \ldots \\ \hline 0 & 1 & \ldots & \ldots & \ldots & & \ldots \\ 0 & \Sigma_{32} & \Sigma_{33} & \ldots & \ldots & & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & & \ldots \\ 0 & \Sigma_{M2} & \ldots & \ldots & \Sigma_{MM} & & \ldots \\ \hline 0 & 0 & \ldots & \ldots & 0 & & \Sigma_{(M+1),(M+1)} \end{array}\right) \tag{7}$$

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:

$$\Sigma = \Gamma . \Gamma' \tag{8}$$

where

$$\Gamma = \left(\begin{array}{c|ccccc|c} 1 & 0 & 0 & \ldots & \ldots & & 0 \\ \hline 0 & 1 & 0 & 0 & \ldots & & 0 \\ 0 & \alpha_{32} & \exp(d_1) & 0 & \ldots & & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & & \ldots. \\ \hline 0 & 0 & \ldots & \ldots & 0 & & \exp(d_{M-1}) \end{array}\right) \tag{9}$$

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

[15] Correlated unobserved heterogeneity across equations is captured by type-specific random intercepts $(\alpha^{\tau}_{(1,j)})_{j=1,\ldots,M}$, α^{τ}_2, and α^{τ}_3.

[16] Only differences in utility levels matter in random utility models.

3.3 The likelihood function

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

$$\begin{aligned} l(j^*, k_{j^*}^*, \overline{\ln w}_{jk}^r | r) &= \Pr\left[\bigcap_{j' \neq j^*} \left(u_{j'} - u_{j^*} \leq f_r\left(j^*\right) - f_r\left(j'\right)\right)\right] \times g(\epsilon) \\ &\quad \times \Pr\left[s_{k_{j^*}^*} - \widetilde{h}_r < v \leq s_{k_{j^*}^*+1} - \widetilde{h}_r\right] \end{aligned} \tag{10}$$

where

$$\widetilde{h}_r = \alpha_2^r + X'_{2,j}\beta_2,$$

$$f_r\left(j\right) = \alpha_{(1,j)}^r + X'_1\beta_1^j + \alpha\sum_{k=0}^{M} X_{3(j,k)}\beta_3 \times \left[\Phi\left(s_{k+1} - \widetilde{h}_r\right) - \Phi\left(s_k - \widetilde{h}_r\right)\right],$$

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

with

$$\epsilon = \overline{\ln w}_{j,k}^r - \alpha_3^r - X_{3(j,k)}\beta_3,$$

and

$$\Pr\left[s_k - \widetilde{h}_r < v \leq s_{k+1} - \widetilde{h}_r\right] = \Phi\left(s_{k+1} - \widetilde{h}_r\right) - \Phi\left(s_k - \widetilde{h}_r\right)$$

φ and Φ being respectively the density and cumulative density functions of the standard normal distribution $\mathcal{N}(0,1)$. Finally, for estimating the probability:

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

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

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

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:

$$A_{j^*} . \Sigma^c . A_{j^*}^{'} \tag{11}$$

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 compute

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

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 $\Pr(j|r)$ do not have a closed-form expression.[17] As it is detailed in the following section devoted to data, estimations 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), without the need to rely on GHK probit simulator.

Unconditional on the type, the contribution to the likelihood function of a student who chooses the field j^*, who reaches the educational level $k_{j^*}^*$ and who gets the average labor market log-earnings $\overline{\ln w}_{j^*,k_{j^*}^*}$ follows a finite mixture distribution:

$$l(j^*, k_{j^*}^*, \overline{\ln w}_{j^*,k_{j^*}^*}) = \sum_{r=1}^{R} \Pi_r l(j^*, k_{j^*}^*, \overline{\ln w}^{r}_{j^*,k_{j^*}^*} | r) \tag{12}$$

where $l(j^*, k_{j^*}^*, \overline{\ln w}^{r}_{j^*,k_{j^*}^*} | r)$ denotes the individual contribution to the likelihood given the type r.

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

3.4 Estimation

In order to explain our estimation strategy, let us introduce some further notations: θ_F denotes the whole parameters of the choice equation, θ_L those of the equation for the length of studies, and finally θ_W those of the wage equation. These vectors do not include type-specific intercepts.

As it is usual for a finite mixture of gaussian distributions, we rely on the Expectation-Maximization (EM) algorithm (Dempster, Laird and Rubin, 1977) to estimate our model. This algorithm works by iterating the two following steps until the stability of the log-likelihood function is reached.

At each iteration n of this algorithm, we use the values $\left(\theta_F^{(n)}, \theta_L^{(n)}, \theta_W^{(n)}\right)$ of the parameter vector, the values $(\pi_r^{(n)})_{r=1...R}$ of the mixture distribution and the values $(\alpha_r^{(n)})_{(r)}$ of the type-specific intercepts, which are all obtained from the previous iteration of the algorithm. More precisely, the two steps are the following:

▷ *E-step*

For each type $r = 1, ..., R$ and for each individual i, the posterior probability for the individual i to be of type r is:

$$Pr\left(T_i = r | j_i^*, k_i^*, w_i, X_i\right) = \frac{\pi_r^{(n)} Pr(j_i^*, k_i^*, w_i | T_i = r, X_i)}{\sum_{r=1}^R \pi_r^{(n)} Pr(j_i^*, k_i^*, w_i | T_i = r, X_i)}$$

where T_i is the random variable representing the individual type. In the following, $\pi_{i,r}^{(n)}$ denote these posterior probabilities. Then, we compute the expected completed log-likelihood :

$$\sum_{i=1}^N \sum_{r=1}^R \pi_{i,r}^{(n)} \ln l\left(j_i^*, k_i^*, w_i | T_i = r, (\Pi_r)_r, (\alpha_r)_r, \theta_F, \theta_L, \theta_W\right) \quad (13)$$

▷ *M-step*

We maximize the expected completed log-likelihood function in terms of $((\Pi_r)_r, (\alpha_r)_r, \theta_F, \theta_L, \theta_W)$.

This maximization can be done in two successive steps.

First we update $\pi_k^{(n)}$ such as:

$$\pi_r^{(n+1)} = \frac{\sum_{i=1}^N \pi_{ir}^{(n)}}{\sum_{l=1}^R \sum_{i=1}^N \pi_{il}^{(n)}} \quad (14)$$

Then, due to the partial separability of the conditional completed log-likelihood function (Arcidiacono and Jones, 2003), we get three sequential optimization problems since residuals are assumed to be independent across the three equations. Henceforth:

$$\begin{aligned}
& \sum_{i=1}^{N}\sum_{r=1}^{R}\pi_{i,r}^{(n)}\ln l\left(f_i, l_i, w_i | T_i = r, (\Pi_r)_r, (\alpha_r)_r, \theta_F, \theta_L, \theta_W\right) \\
= & \sum_{i=1}^{N}\sum_{r=1}^{R}\pi_{i,r}^{(n)}\ln l\left(w_i | T_i = r, (\Pi_r)_r, (\alpha_r^W)_r, \theta_W\right) \\
+ & \sum_{i=1}^{N}\sum_{r=1}^{R}\pi_{i,r}^{(n)}\ln l\left(l_i | T_i = r, (\Pi_r)_r, (\alpha_r^W)_r, (\alpha_r^L)_r, \theta_W, \theta_L\right) \\
+ & \sum_{i=1}^{N}\sum_{r=1}^{R}\pi_{i,r}^{(n)}\ln l\left(f_i | T_i = r, (\Pi_r)_r, (\alpha_r^W)_r, (\alpha_r^L)_r, (\alpha_r^F)_r, \theta_W, \theta_L, \theta_F\right)
\end{aligned}$$

It implies that first, we maximize the log-wage equation. Given the estimates of this equation, we estimate the parameters of the equation for the length of studies. Finally, given the previous estimates, we maximize the choice equation. Although this procedure does not yield Full Information Maximum Likelihood estimates, Arcidiacono and Jones (2003) show that this method produces consistent estimates of the parameters, with large computational savings.[18]

In order to get standard errors estimates, we rely on a parametric bootstrap procedure, instead of a non parametric one, since this last method is unstable when applied to the EM algorithm. The parametric bootstrap consists first in obtaining reliable parameter estimates denoted $\widehat{\theta}$. We get $\widehat{\theta}$ by replicating the previously described EM algorithm with different random initial values for the parameters. The iteration process is necessary to ensure we obtain a global maximum. Then, given X and $\widehat{\theta}$, we draw H vectors of the endogenous variables $\left(j_i^h, k_i^h, w_i^h\right)_{h=1\ldots H}$. For each newly generated data set, we estimate θ_h^*. Final parameters and standard errors estimates are calculated

[18]Allowing type probabilities to depend on high school graduation track does not affect the sequential EM algorithm.

as:

$$\overline{\theta^*} = \frac{1}{H}\sum_{h=1}^{H}\theta_h^* \qquad (15)$$

$$\sigma_{\theta^*} = \frac{1}{H-1}\sum_{h=1}^{H}(\theta_h^* - \overline{\theta^*})^2 \qquad (16)$$

4 Data

The model presented above is estimated using French data coming from the "Génération 92" and "Génération 98" surveys, which are collected by the Centre for Study and Research on Qualifications (CEREQ, Marseille).[19] The "Génération 92" survey consists of a large sample of 26,359 individuals who left the French educational system in 1992 and were interviewed five years later, in 1997. 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 qualification nomenclature. This database has the main advantage to contain information on both educational and labor market histories (over the first five years following the exit from the educational system). Furthermore, the survey provides 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. Most of the individual covariates observed in the "Génération 92" survey are also provided by the "Génération 98" survey, which consists of a sample of 14,365 individuals who left the French educational system six years later, in 1998, and were interviewed in 2003[20]. In this paper, we exploit the pooled dataset which contains informations on a total of 40,724 individuals entering the labor market either in 1992 or in 1998.

Our subsample of interest is constituted of respondents from both of these surveys having at least passed the national high school final examination:[21] it is then restricted to 27,389 individuals. Furthermore, within this selected sample, we restrict our analysis to the individuals having attended university except medicine

[19]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.

[20]Although a longer observation window is available for each *Generation* dataset, the average log-earnings are computed using only the observations from 1992 to 1995 for *Generation 1992* (resp. 1998 to 2001 for *Generation 1998*). In particular, restricting to a 4-years window allow us to limit the number of individuals that have to be dropped because of missing earnings values, in addition to the fact that it allows us to work with two periods of virtually opposed economic conditions and helps identifying the earnings elasticities of major choices.

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

faculties and IUT ("*Institut Universitaire de Technologie*", which are two-year 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. Missing covariates values finally leaves us with a sample of 7,346 individuals[22].

Post-secondary 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. Table 7 (reported in Appendix B) provide a descriptive outlook for the determinants of post-secondary schooling choices in France.

We first focus on the choice of the study field. Tables 7 show that this choice is related with gender, age in 6^{th} grade,[23] and parents' profession.

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.

Parental characteristics also seem to play in important role on the choice of the major. The higher the parental qualification, the higher the probability to study sciences. For instance, 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.[24] Table 7 also shows a strong correlation between the chosen field and the length of studies. Only one quarter of graduates complete their degree in Humanities and Social Sciences. Unlike graduates, half of dropouts during the first two years of college studied Humanities.

[22] In order to prevent our estimates to be driven by outliers, we also drop individuals with average log-earnings below the 2.5 percentile (respectively above the 97.5 percentile) of the log-earnings distribution.

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

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

Finally, the higher the level, the larger the mean of log earnings (table 3 reported below): graduates earn 1.7 times more than dropouts. There are significant differences in average earnings associated with different majors: Sciences ranks first, followed by Law, Economics and Management, and finally Humanities and Social Sciences. The discrepancy between majors is greater in 1998 than in 1992. Sciences and Law, Economics and Management benefited from a dynamic and sustained growth, whatever the length considered.

Table 1: Descriptive statistics: majors and levels of post-secondary schooling

	Number	Percent
Major		
Sciences	2,106	28.67
Humanities and Social Sciences	2,761	37.59
Law, Economics and Management	2,479	33.75
Post-secondary education level		
Dropout	1,762	23.99
Two years of college	732	9.97
Licence (BA degree)	1,400	19.06
Maîtrise (MA degree)	1,486	20.23
Post Maîtrise (Graduates)	1,966	26.76

Source: Générations 1992 and 1998(CEREQ, Marseille)

Table 2: Descriptive statistics: covariates

	Number	Percent
Year of entry into the labor market		
1992	3,436	46.77
1998	3,910	53.23
Gender		
Male	3,197	43.52
Female	4,149	56.48
Born abroad		
No	7,164	97.52
Yes	182	2.48
Age in 6^{th} *grade*		
≤ 10	858	11.68
11	6,109	83.16
≥ 12	379	5.16
Secondary Schooling Track		
Humanities (L)	1,712	23.31
Economics and Social Sciences (ES)	1,733	23.59
Sciences (S)	2,523	34.35
Vocational or Technological (ST,SMS)	1,378	18.76
Father's profession (in 1992 and 1998)		
Farmer or Tradesman	1131	15.40
Executive	2213	30.13
Technician	898	12.22
White-Collar	1468	19.98
Blue-collar	1237	16.84
Housewife	399	5.43
Mother's profession (in 1992 and 1998)		
Farmer or Tradesman	527	7.17
Executive	1226	16.69
Technician	508	6.92
White-Collar	3269	44.50
Blue-collar	508	6.92
Housewife	1308	17.81

Source: Générations 1992 and 1998 (CEREQ, Marseille)

Table 3: Average monthly earnings (constant 1992 Francs) according to the length and the field of studies

Field	Length	Average monthly earnings
	Dropout	4,920
	Two years of college	5,983
	Licence (BA degree)	6,181
	Maitrise (MA degree)	6,739
	Post Maitrise (Graduates)	8,414
Sciences		7,277
Humanities and Social Sciences		5,942
Law, Economics and Management		6,666
Generation 1992		
	Dropout	4,205
	Two years of college	6,057
	Licence (BA degree)	6,082
	Maitrise (MA degree)	6,556
	Post Maitrise (Graduates)	7,621
Sciences		6,833
Humanities and Social Sciences		6,088
Law, Economics and Management		6,318
Generation 1998		
	Dropout	5,219
	Two years of college	5,938
	Licence (BA degree)	6,292
	Maitrise (MA degree)	6,942
	Post Maitrise (Graduates)	9,450
Sciences		7,758
Humanities and Social Sciences		5,835
Law, Economics and Management		6,976

Source: Générations 1992 and 1998 (CEREQ, Marseille)

5 Identification strategy

For identifiability reasons, we impose usual restrictions on the type-specific heterogeneity terms of stage 1 and stage 2. Namely, in the MNP model corresponding to the choice equation, we set $\forall r \in \{1, \ldots, R\}, \alpha^r_{(1,1)} = 0$ and, in the ordered probit model corresponding to the second equation, $\alpha^1_2 = 0$.

In order to identify our model, and in particular the effect of expected earnings on the probability to choose each major, without relying on distributional assumptions, we exploit variations in the relative monetary returns to each major induced by the year of entry into the labor market[25]. Noteworthy is that descriptive statistics reported below (see table 3 in data section) suggest a significant change in the relative returns to each major between 1992 and 1998, which correspond respectively to a period of very weak economic growth and strong growth[26]. Namely, after controlling for the change in the distribution of educational levels between 1992 and 1998 as well as for inflation, we find a relative increase of respectively 13.5% and 10.4% in the average earnings of sciences and law, economics and management between the two periods, while the average earnings of humanities and social sciences majors decreased by 4.2% in the meantime[27]. Besides, it seems reasonable to assume that whether the individual will enter the labor market in 1992 or in 1998 has no direct effect her choice of major, in other words that every observed characteristics being equal, preferences for each major are stable between 1992 and 1998[28]. We exploit the fact that the returns to the different university majors are unequally affected by the business cycle in order to identify the elasticity of the choice of major to expected earnings[29]. Hence, we introduce into the earnings equation interaction terms between the chosen major and an entry year dummy. This dummy variable is equal to zero if the individual enters the labor market in

[25]Berger(1988) also relies on exogenous variations in the returns to each major according to time of entry into the labor market in order to identify the effect of expected earnings on college major choice. Unlike ours, his framework does not take into account the determination of the length of studies. Besides, his results rely on the *Independence from Irrelevant Alternative* assumption for the choice of the major which is unlikely to hold in such a context.

[26]See figure 1 in appendix.

[27]These relative variations between 1992 and 1998 are obtained by computing for each major the average of mean monthly earnings conditional on each educational level, weighted by the frequency of each level.

[28]In particular, noteworthy is that no major university reform was implemented between 1992 and 1998. The progressive application of the Bologna process to the French post-secondary educational system began in 1999, thus not affecting the individuals in our sample who had already entered the labor market at that time.

[29]On a related ground, in a recent paper considering the career effects of graduating in a recession, Oreopoulos et Al. (2008) show that Canadian college graduates are unequally affected by the recession according to their major of study.

1992, and one otherwise if she enters the labor market six years later in 1998. Its interaction with the chosen major is assumed to affect only the earnings and not the two other outcomes. This exclusion restriction (over)identifies the parameter α associated with the expected returns in the choice equation without relying on a distributional assumption on the error terms. Besides, the covariates indicating the father's and mother's professions (respectively in 1992 and 1998), the age of the student in 6^{th} grade, and the high-school major are included in the list of regressors affecting the choice of the major and the determination of the length of studies, but they are excluded from the earnings equation. Similarly to Arcidiacono (2005, section 4), these exclusion restrictions, in addition to the assumed functional forms, allow to identify the unobserved heterogeneity types. These covariates may be correlated with the individual's preferences and ability, represented respectively by $\alpha^r_{(1,j)}$ and α^r_2. We finally also assume, considering that overcrowding may affect educational attainment, that the proportion of college students who are registered in the same major and in the same university than the individual may only affect the length of studies.[30]

6 Results

Tables 12 to 16 (reported in Appendix C) give the parameter estimates of the model. Tables 12 and 13 report the parameter estimates of the equations generating the major choice.

Students whose mother is a white-collar choose less frequently majors in Humanities and Social Sciences, compared to Sciences, than a student whose mother is an executive. Noteworthy, students whose mother is a farmer or a tradeswoman, a technician, or a white-collar worker also choose less frequently majors in Law, Economics and Management compared to Sciences. In all other cases, parental, and in particular father's profession has generally no effect on the major choice.

The nationality of his/her parents has a significant and quantitatively large impact on the choice of a major in Law, Economics and Management as well as in Humanities and Social Sciences, compared to Sciences. Besides, students born abroad are significantly less likely to study law, economics or management. Noteworthy, female students are very significantly less likely to study sciences. As expected, students who obtained a *Baccalauréat* (i.e. the terminal high-school diploma in France) in sciences are significantly more likely to choose a post-secondary major in sciences. Students who were older than expected (i.e. 12 years old or above) at the entry into junior high school (sixth grade) choose less frequently a major in

[30]This variable is calculated using information coming from the *SISE* database provided by the French Ministry of Education.

sciences. Finally, the expected wage returns in a given post-secondary major has a statistically significant but rather small effect on the choice of the major (see the value for the estimate of parameter α in Table 16).

Most covariates have a significant impact on the length of post-secondary studies (see Table 14). For instance, students whose parents are white-collar or blue-collar workers leave the post-secondary educational system at a lower level. Students whose both parents are French reach generally a higher level of post-secondary education. Students who were younger than expected (i.e. 10 years old or below) at the entry into junior high-school reach a higher level of education. Those who obtained their *Baccalauréat* in sciences are also more likely to reach a higher level of post-secondary education. When the proportion of college students who are studying in the same major and in the same university increases, which implies that the proportion of students preparing a BA or MA degree is lower in this major and in this university, the individual probability of reaching a high level of education (B.A. and above) in this major is lower, other things being equal. This may result from the selection imposed by the university after the end of college (i.e. at the entry in the third year of post-secondary schooling in the major), or from peers effects; this second interpretation is the one set forth by Arcidiacono (2004, 2005). Finally, women are less likely to pursue long studies. This is a common result in France: nowadays on average French women are more educated than men, but graduated men are more numerous than women.

Table 15 gives the parameter estimates of the (log-)earnings equation. On average, earnings are lower for females and they are higher in the region Ile-de-France (including Paris). Mean (log-)earnings increase with the length of studies in post-secondary education. However, this increase is lower from the BA degree in the majors in humanities and social sciences. Noteworthy, the marginal returns to each additional year of post secondary education are also lower, up to graduate level, for the individuals entering the labor market in 1998 than for those leaving university six years before. Besides, consistently with the fact that the individuals entering the French labor force in 1998 benefit from positive economic conditions, in particular as compared to 1992, mean (log-)earnings are substantially higher for those leaving university in 1998. Finally, while controlling for selection on observables and unobservables turns into insignificance the differences in returns to each major for the individuals leaving university in 1992, individuals from the *Generation 1998* sample experience negative relative returns to studying humanities and social sciences.

Tables 16 and 17 report the parameter estimates of the distribution of unobserved individual heterogeneity terms. The first group of individuals represents 38 percent of the population of students. Individuals in this group are characterized by the lowest unobserved type-specific preference for studies in sciences as well as the highest highest type-specific earnings intercept α_3. The second group represents approximatively 34 percent of the population of students. Individuals in this group are characterized by the lowest type-specific preference $\alpha_{(1.3)}$ for studies in law, economics and management. They also have the lowest type-specific propensity (or ability) α_2 to undertake long post-secondary studies. Finally, the third group represents about 28 percent of the population; it is both characterized by the lowest type-specific earnings intercept term α_3 and the highest propensity to pursue long post-secondary studies.[31].

The model fit is quite good. Table 4 shows that the model slightly overestimates (resp. underestimates) the proportion of students in humanities and social sciences (resp. law, economics and management).

To get a more precise view of the effect of expected earnings on the choice of the post-secondary major, we run simulation exercises that consider a 10% increase or decrease in the expected earnings associated with a given major (tables 4 to 6 below).[32]

In general, the impacts are quantitatively small even though they are statistically significant. The lowest impacts concern the majors in sciences. A 10% increase in the expected earnings associated with majors in sciences leads to an increase of 0.25 percentage points in the proportion of students in this major. This increase is mainly compensated by a decrease of 0.19 percentage points in the proportion of students in humanities and social sciences (see Table 4).A 10% decrease in the expected earnings associated with majors in sciences results in almost symmetric variations in allocations across majors.

Impacts resulting from a 10% increase or decrease in the expected earnings associated with majors in humanities and social sciences are substantially higher although still quantitatively small (see Table 5). For instance, a 10% increase in the expected earnings associated with a post-secondary in these majors results in an increase of about 0.53 percentage points in the proportion of students in these majors, this increase being mainly compensated by a decrease of about 0.34 percentage points in the proportion of students in law, economics and management and to a lesser extent by a 0.19 points decrease in the proportion of students in sciences. Once again, a 10% decrease in expected earnings has almost symmetric impacts on allocations.

Finally, a 10% increase in the expected earnings associated with a post-secondary

[31]Tables 18 and 19 report the sample distributions of majors, length of studies and monthly log-earnings by heterogeneity type. This is done, after the estimation, by affecting to each individual in the sample the type maximizing the posterior membership probability. The resulting distributions are consistent with the preceding characterization of each type.

[32]Simulating both types of variation enables us to see whether the impacts on allocations across majors are symmetric or not.

education in law, economics and management majors result in an increase of 0.4 percentage points in the proportion of students in these majors, this increase being mainly compensated by a decrease of 0.34 percentage points in the proportion of students in humanities and social sciences (see Table 6). The effects are still symmetric for a 10% decrease in the expected earnings associated with this major[33].

The preceding simulation exercices allow us to compute the earnings elasticities of major choice (at the observed sample distribution), which present the advantage of being easily interpreted. Namely, simulating a 10% increase in the expected earnings for each major yields low elasticities of respectively 0.09 for sciences, 0.14 for humanities and social sciences and finally 0.12 for law, economics and management.

Table 4: Simulation of a 10% variation in expected earnings of the majors in sciences

	Observed proportions	Predicted proportions	$(p^S - p^P)$	$\widehat{\sigma}_{(p^S - p^P)}$ Standard error
Sciences				
	10% *increase*			
Sample distribution				
Sciences	28.67	27.97	0.251	0.019
Humanities and Social Sciences	37.59	41.17	-0.189	0.013
Law, Economics and Management	33.75	30.86	-0.062	0.009
	10% *decrease*			
Sample distribution				
Sciences	28.67	27.97	-0.276	0.021
Humanities and Social Sciences	37.59	41.17	0.209	0.014
Law, Economics and Management	33.75	30.86	0.068	0.009

Source: Générations 1992 and 1998 (CEREQ)

Remark: p^S and p^P denote the predicted proportions after and before the simulation, respectively.

[33]Given that the model we estimate *a priori* yields non linear effects of expected earnings on the probability to choose each major, we also simulated 20% increases in the expected earnings associated with each field of study. The resulting effects were about twice (namely 1.9) larger. We therefore provide the earnings elasticities of major choice relying only on the first set of simulations.

Table 5: Simulation of a 10% variation in expected earnings of the majors in humanities and social sciences

	Observed proportion	Predicted proportion	$(p^S - p^P)$	$\widehat{\sigma}_{(p^S - p^P)}$ Standard error
Humanities and Social Sciences				
10% *increase*				
Sample distribution				
Sciences	28.67	27.97	-0.189	0.013
Humanities and Social Sciences	37.59	41.17	0.526	0.048
Law, Economics and Management	33.75	30.86	-0.336	0.038
10% *decrease*				
Sample distribution				
Sciences	28.67	27.97	0.209	0.014
Humanities and Social Sciences	37.59	41.17	-0.580	0.053
Law, Economics and Management	33.75	30.86	0.371	0.042

Source: Générations 1992 and 1998 (CEREQ).

Remark: p^S and p^P denote the predicted proportions after and before the simulation, respectively.

Table 6: Simulation of a 10% variation in expected earnings of majors in law, economics and management

	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	28.67	27.97	-0.062	0.009
Humanities and Social Sciences	37.59	41.17	-0.337	0.038
Law, Economics and Management	33.75	30.86	0.399	0.042
	10% decrease			
Sample distribution				
Sciences	28.67	27.97	0.068	0.009
Humanities and Social Sciences	37.59	41.17	0.371	0.042
Law, Economics and Management	33.75	30.86	-0.439	0.046

Source: Générations 1992 and 1998 (CEREQ).

Remark: p^S and p^P denote the predicted proportions after and before the simulation, respectively.

7 Conclusion

Our results suggest a low elasticity of post-secondary major choices to expected earnings. Thus it appears that the choice of a major of study which is made when entering university is mainly driven by the *consumption value* of schooling which is related both to schooling preferences and abilities, rather than by its *investment value*. Our paper provide strong evidence, in line with Carneiro, Hansen and Heckman (2003), that, at least for the French university context, nonpecuniary factors are a key determinant of schooling choices.

From a policy point of view, this paper suggest that the solution to the shortage existing for certain skills, mainly scientific in the European context, does not lie in financial incentives. Providing incentives, as often advocated, to implement gain and profit-sharing schemes appears to be unlikely to overcome skill shortages. The solution probably lies upstream, within preferences formation at school.

References

AGHION P. and COHEN E. (2004) : *Education and Growth*, Conseil d'Analyse Economique (French Council of Economic Analysis), Report No. 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.

ARCIDIACONO P. and JONES J.B. (2003) : "Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm", *Econometrica*, vol. 71, 933-946.

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

BELZIL C. and HANSEN J. (2004) : "Earnings Dispersion, Risk Aversion and Education", *Research in Labor Economics*, vol. 23, 335-358.

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.

BREWER D. J., EIDE E. R. and EHRENBERG R. G. (1999) : "Does it Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings", *Journal of Human Resources*, vol. 34, 104-123.

BRODATY T., GARY-BOBO R. and PRIETO A. (2006): "Risk Aversion and Human Capital Investment: A Structural Econometric Model", *CEPR Discussion Paper* No. 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.

CARNEIRO P., HANSEN K. and HECKMAN J.J. (2003) : "Estimating Distributions of Counterfactuals with an Application to the Returns to Schooling and Measurement of the Effect of Uncertainty on Schooling Choice", *International Economic Review*, vol. 44, 361-422.

DEMPSTER A.P, LAIRD M. and RUBIN D.B. (1977) : "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society*, Series B, vol. 39, 1-38.

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.

HECKMAN J. and SINGER B. (1984) : "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, vol. 52, 271-320.

JAMES E., ALSALAM N., CONATY J. C. and TO D. L. (1989) : "College Quality and Future Earnings: Where Should You Send Your Child to College", *American Economic Review*, vol. 79, 247-252.

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.

LEE D. (2005) : "An Estimable Dynamic General Equilibrium Model of Work, Schooling, and Occupational Choice", *International Economic Review*, vol. 46, 1-34.

LOURY L. D. and GARMAN D. (1998) : "College Selectivity and Earnings", *Journal of Labor Economics*, vol. 13, 289-308.

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.

McLACHLAN G.J. and PEEL D. (2000) : *Finite Mixture Models*. Wiley, New York.

OREOPOULOS P., von VACHTER T. and HEISZ A., "The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates", *IZA Discussion Paper* No. 3578.

SAKS R. E. and SHORE S. H. (2005) : "Risk and Career Choice", *Advances in Economic Analysis and Policy*, vol. 5, 1-43.

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

Figure 1: French real GDP growth, 1990-2002

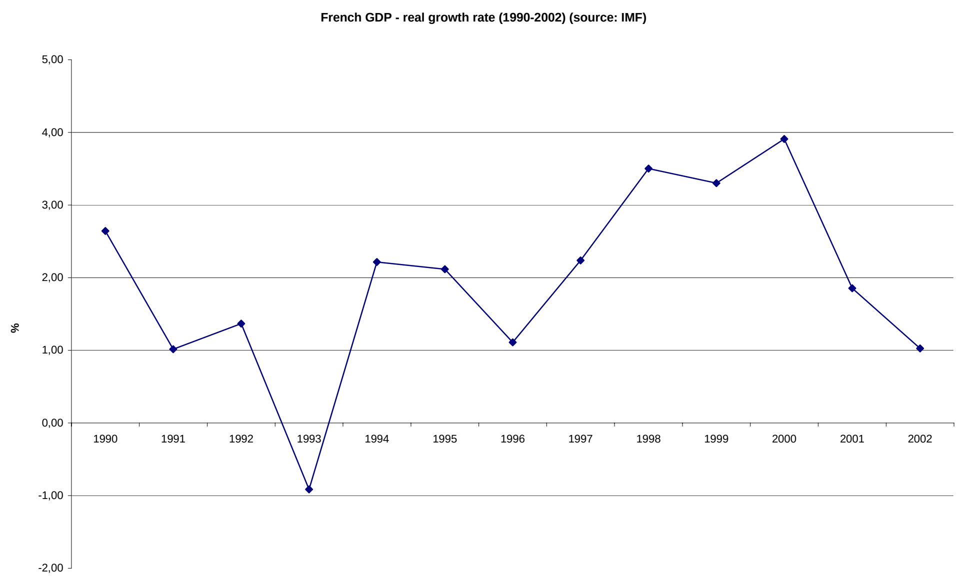

Table 7: Distribution of various subgroups across majors (in percent, beginning)

	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
Father's profession			
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			
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
Educational 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
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: Générations 1992 and 1998 (CEREQ, Marseille)

Remarks: Lines sum up to 100%, except for educational levels, for which columns sum up to 100%.

Table 8: Majors switching after one year of college (in percent)

Major (first year of college)	LEM	HSS	S
Major (second year of college)			
LEM	94.95	1.45	0.69
HSS	4.89	97.78	3.70
S	0.16	0.77	95.60

Source: Panel 1989 (DEPP, French Ministry of Education)

Remarks: Lines sum up to 100%.

Abbreviations: HSS for Humanities and Social Sciences, LEM for Law, Economics and Management, S for Sciences.

Table 9: Aspiration levels and effective level of studies (in percent)

Level of studies	Less than college	College	BA	MA or more
Aspiration (first year of college)				
Less than college	33.71	12.36	28.09	25.84
College	45	20.50	17	17.50
BA	32.49	16.40	24.61	26.50
MA or more	23.06	13.97	25.40	37.57

Source: Panel 1989 (DEPP, French Ministry of Education)

Remarks: Lines sum up to 100%.

Table 10: Distribution of majors and education levels (Generation 92 subsample)

	Number	Percent
University fields		
Sciences	1,094	31.84
Humanities and Social Sciences	1,174	34.17
Law, Economics and Management	1,168	33.99
Post-secondary education level		
Dropout	518	15.08
Two years of college	281	8.18
Licence (BA degree)	742	21.59
Maîtrise (MA degree)	781	22.73
Post Maîtrise (Graduates)	1,114	32.42
Total	3,436	100

Table 11: Distribution of majors and education levels (Generation 98 subsample)

	Number	Percent
University fields		
Sciences	1,012	25.88
Humanities and Social Sciences	1,587	40.59
Law, Economics and Management	1,311	33.53
Post-secondary education level		
Dropout	1,244	31.82
Two years of college	451	11.53
Licence (BA degree)	658	16.83
Maîtrise (MA degree)	705	18.03
Post Maîtrise (Graduates)	852	21.79
Total	3,910	100

B Parameter estimates

Table 12: Choice of the major (beginning)

Covariates	Estimate	Standard Error
Sciences	*Ref*	*Ref*
Humanities and Social Sciences		
Father's profession		
Executive	*Ref*	*Ref*
Farmer or tradesman	-0.103	0.095
Technician	-0.083	0.090
White-collar	-0.053	0.068
Blue-collar	0.073	0.089
Unknown	0.438	0.129
Mother's profession		
Executive	*Ref*	*Ref*
Farmer or tradesman	-0.210	0.109
Technician	-0.139	0.101
White-collar	-0.134	0.057
Blue-collar	-0.175	0.107
Unknown	-0.237	0.082
Born abroad	-0.190	0.123
Woman	0.920	0.051
Both parents are French	-0.303	0.062
Age in 6^{th} *grade*		
≤ 10	-0.021	0.072
11	*Ref*	*Ref*
≥ 12	0.391	0.102
Baccalauréat		
General, sciences	*Ref*	*Ref*
General, humanities	2.200	0.075
General, economics	2.287	0.082
Vocational or technological	1.164	0.064

Table 13: Choice of the major (end)

Covariates	Estimate	Standard Error
Law, Economics and Management		
Father's profession		
Executive	*Ref*	*Ref*
Farmer or tradesman	-0.030	0.111
Technician	-0.094	0.110
White-collar	-0.026	0.097
Blue-collar	0.004	0.117
Unknown	0.477	0.145
Mother's profession		
Executive	*Ref*	*Ref*
Farmer or tradesman	-0.335	0.143
Technician	-0.261	0.135
White-collar	-0.179	0.073
Blue-collar	-0.046	0.162
Unknown	-0.165	0.088
Born abroad	-0.335	0.180
Woman	0.900	0.072
Both parents are French	-0.343	0.084
Age in 6^{th} *grade*		
≤ 10	-0.031	0.092
11	*Ref*	*Ref*
≥ 12 years	0.528	0.150
Baccalauréat		
General, sciences	*Ref*	*Ref*
General, humanities	1.888	0.117
General, economics	3.065	0.150
Vocational or technological	1.587	0.105

Source: Générations 1992 and 1998 (CEREQ. Marseille)

Table 14: Equation for the length of studies

Covariates	Estimate	Standard Error
Father's profession		
Farmer or tradesman	-0.232	0.043
Executive	*Ref*	*Ref*
Technician	-0.214	0.046
White-collar	-0.424	0.040
Blue-collar	-0.391	0.042
Unknown	-0.238	0.060
Mother's profession		
Farmer or tradesman	0.010	0.053
Executive	*Ref*	*Ref*
Technician	-0.143	0.064
White-collar	-0.118	0.038
Blue-collar	-0.236	0.049
Unknown	0.070	0.046
Born abroad	0.319	0.079
Woman	-0.063	0.031
Both parents are French	0.165	0.044
Age in 6^{th} *grade*		
≤ 10	0.192	0.047
11	*Ref*	*Ref*
≥ 12	-0.313	0.084
Baccalauréat		
General, sciences	*Ref*	*Ref*
General, humanities	-0.484	0.036
General, economics	-0.267	0.031
Vocational or technological	-1.051	0.045
Proportion of students in college	-1.306	0.063
Generation 1998	-0.446	0.035

Source: Générations 1992 and 1998 (CEREQ, Marseille)

Table 15: Earnings equation

Covariates	Estimate	St. Error
Both parents are French	0.004	0.016
Region Ile de France	0.118	0.015
Female	-0.074	0.020
Born abroad	0.009	0.039
Generation 1998	0.366	0.030
Field of studies		
Sciences	*Ref*	*Ref*
Humanities and Social Sciences	0.056	0.039
Law. Economics and Management	-0.047	0.040
Level of studies		
Dropout	*Ref*	*Ref*
Two years of college	0.397	0.050
Licence (BA degree)	0.496	0.037
Maitrise (MA degree)	0.534	0.046
Post Maitrise (Graduates)	0.760	0.039
Interactions between field and level		
Humanities and Social Sciences		
Dropout	*Ref*	*Ref*
Two years of college	-0.087	0.053
Licence (BA degree)	-0.155	0.043
Maitrise (MA degree)	-0.208	0.040
Post Maitrise (Graduates)	-0.194	0.048
Law, Economics and Management		
Dropout	*Ref*	*Ref*
Two years of college	0.044	0.049
Licence (BA degree)	-0.003	0.047
Maitrise (MA degree)	-0.008	0.041
Post Maitrise (Graduates)	-0.076	0.053
Interactions between female gender and level		
Dropout	*Ref*	*Ref*
Two years of college	0.058	0.037
Licence (BA degree)	0.090	0.036
Maitrise (MA degree)	0.136	0.037
Post Maitrise (Graduates)	0.018	0.036
Interactions between a dummy Generation 1998 and level		
Dropout	*Ref*	*Ref*
Two years of college	-0.267	0.042
Licence (BA degree)	-0.259	0.031
Maitrise (MA degree)	-0.231	0.034
Post Maitrise (Graduates)	-0.054	0.051
Interactions between a dummy Generation 1998 and field		
Sciences	*Ref*	*Ref*
Humanities and Social Sciences	-0.121	0.031
Law, Business and Management	0.016	0.035

Table 16: Other parameters

Covariance matrix of residuals

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1.053 & 0 \\ (-) & (-) & \mathit{(0.129)} & (-) \\ 0 & 1.053 & 2.379 & 0 \\ (-) & \mathit{(0.129)} & \mathit{(0.318)} & (-) \\ 0 & 0 & 0 & 0.516 \\ (-) & (-) & (-) & \mathit{(0.005)} \end{pmatrix}$$

	Estimate	St. Error
Thresholds		
s_2	-2.556	0.067
s_3	-2.154	0.070
s_4	-1.472	0.067
s_5	-0.710	0.069
α	0.019	0.001
Type probabilities		
Type 1	0.380	0.004
Type 2	0.337	0.004
Type 3	0.283	0.004

Source: Générations 1992 and 1998 (CEREQ)

Table 17: Type-specific heterogeneity parameters

	Estimate	St. Error
Type 1		
$\alpha_{(1.1)}$	0.000	-
$\alpha_{(1.2)}$	1.244	0.091
$\alpha_{(1.3)}$	1.037	0.093
α_2	0.000	-
α_3	8.192	0.038
Type 2		
$\alpha_{(1.1)}$	0.000	-
$\alpha_{(1.2)}$	-1.312	0.091
$\alpha_{(1.3)}$	-2.828	0.141
α_2	-0.363	0.043
α_3	8.111	0.032
Type 3		
$\alpha_{(1.1)}$	0.000	-
$\alpha_{(1.2)}$	-1.363	0.082
$\alpha_{(1.3)}$	-1.571	0.119
α_2	0.502	0.051
α_3	8.089	0.036

Source: Générations 1992 and 1998 (CEREQ)

Table 18: Distribution of majors and length of studies, by type

Major	Type 1	Type 2	Type 3
Sciences	0.00	44.66	50.34
Humanities and Social Sciences	41.02	55.34	17.49
Law, Business and Management	58.98	0.00	32.16
Length	Type 1	Type 2	Type 3
Dropout	26.40	47.55	0.00
Two years of college	12.43	16.73	0.86
Licence (BA degree)	26.81	24.42	4.58
Maitrise (MA degree)	19.77	10.53	29.43
Post Maitrise (Graduates)	14.58	0.77	65.14

Source: Générations 1992 and 1998 (CEREQ)

Table 19: Distribution of log-earnings, by type

	Mean	St.Error
Whole sample	8.65	0.57
Type 1	8.62	0.32
Type 2	8.48	0.31
Type 3	8.84	0.29

Source: Générations 1992 and 1998 (CEREQ)