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

Magali Beffy† Denis Fougère‡ Arnaud Maurel§

6th May 2008
Preliminary version

Abstract

This paper examines the determinants of the choice of the major when the length of post-secondary studies and subsequent labor market earnings are uncertain. For that purpose, we use French data coming from the 1998 *Generation* survey collected by CEREQ (*Centre d'Etudes et de Recherches sur les Qualifications*, France). Our econometric approach is based on a structural dynamic programming model of schooling and employment choices. Once graduated from high school, individuals are supposed to compute the optimal value function corresponding to each major and then to choose the major associated with the highest value function. Relying on a two-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 probabilities to reach the different educational levels and those that affect the labor market earnings equation. Following Arcidiacono and Jones (2003), we rely on the EM algorithm with a sequential maximization step to produce consistent parameter estimates. Simulating for each 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. We also show that the overall finding is robust to an alternative semi-structural specification of the schooling decision process, which yields significant but even smaller effects.

*We would like to thank Nicolas Chopin, Xavier d'Haultfoeuille, Guy Laroque, Robert Miller, Gerard J. van den Berg, participants in seminars at 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. E-mail: maurel@ensae.fr

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 *Conseil d'Analyse Economique*, 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 things, 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 both a dynamic structural model and a semi-structural model of post-secondary educational choices. More precisely, we try to disentangle the simultaneous effects of preferences, abilities and 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 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 expected earnings affect in turn schooling choices. More recently, Arcidiacono (2004, 2005) has considered sequential models of college attendance in which the value of each major depends on the corresponding expected

[1] Several 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. Altonji (1993) estimates a sequential model in which schooling decisions depend on expected returns to education, but he does not explicitly consider the choice of major.

flow of earnings. Our model is close to Arcidiacono's models. Nevertheless, unlike Arcidiacono, we assume 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 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. For instance, Buchinsky and Leslie (2000) use a dynamic schooling decision framework in which they compare the predictions of 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 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; following a suggestion by Manski (1993), the predicted earnings are computed from the wages of young individuals who have the same education level and who are currently working.

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. Noteworthy, stylized facts seem to be consistent with such a framework.[2] It is also the first microeconometric study devoted to these issues in France. Our study has at least two main limitations. First, in the absence of appropriate information allowing identification of risk-aversion coefficients, we do not consider individual attitudes towards risk.[3] We also ignore the possibility for the student to switch major during his/her

[2]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 A, Table 15).

[3]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).

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 14 in Appendix A).

The remainder of the paper is organized as follows. Section 2 describes the structural dynamic model of educational choices. The econometric counterpart of this model and the likelihood function are discussed in Section 3. Section 4 presents an alternative semi-structural model of post-secondary schooling choices. Section 5 describes the data and presents some preliminary statistics, while Section 6 contains the estimation and simulation results.

2 A structural dynamic model

In this section, we present a structural dynamic model of post-secondary educational choices. After graduating from high-school, individuals are assumed to choose their field of study in which they will complete a certain level of education. Note that we restrict our analysis to individuals who attend university. [4] Once they leave the post-secondary education system, they are supposed to enter the labor market which is supposed to be an absorbing state.

The structural approach requires that we explicitly specify the post-secondary education path in terms of successive decisions: the individual goes on studying at the end of the year considered or leaves the post-secondary education system to enter the labor market.

Thus we consider the following sequence of individual decisions:

- Stage 1: When entering college, each student chooses his/her post-secondary major j^* among a set of M majors. For each $j \in \{1, \ldots, M\}$, we denote by d_j^1 a dummy variable which is equal to one if major j is chosen, zero otherwise.

- Stage 2: At the end of each completed level of education $k \in \{1, \ldots, K-1\}$, he/she decides either to keep on studying in the field chosen in stage 1 ($d_{k+1} = 1$) or to enter the labor market ($d_{k+1} = 0$).

Let us consider a post-secondary student maximizing his/her expected discounted lifetime utility. Lifetime utility is assumed to be time separable and the

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

per-period utility of choosing option d, school or work, U_t^d, depends on the choice made by the individual at time t. In the following, we will denote by β the discount factor.

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.[5] Individuals are supposed to know their type. Within this framework, unobserved heterogeneity (i.e. unobserved preferences and abilities for each major and each level of education, and unobserved labor market productivity) is type-specific. From now on we will consider a type-r individual and for the sake of simplicity the subscript r will be omitted.

The utility associated to schooling in the post-secondary field j at the level k is assumed to be given by:

$$U_{j,k}^s = \alpha_{s,j,k} + X_s\beta_{s,j} + \epsilon_{s,j,k} \tag{1}$$

It depends on a type-specific intercept $\alpha_{s,j,k}$, which captures unobserved preferences as well as abilities for major j and level of study k, and on a random term $\epsilon_{s,j,k}$ independent of $\alpha_{s,j,k}$. X_s is a set of observable individual covariates that affect the attractiveness of studying in 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). $\beta_{s,j}$ is a parameter vector associated with X_s and specific to field j.

Having obtained the educational level (degree) k in major j^*, the student may enter the labor market. We assume that the labor market is an absorbing state: individuals do not resume studies after entering the labor force. In order to take both employment and nonemployment spells into account in the utility associated to work, we refer to average *earnings* as wages weighted by employment spell durations. Hence, the average monthly log-earnings in a T_{obs} years long labor market history for a worker with education (j, k), is given by :

$$\overline{\ln w}_{jk} = \frac{\sum_{s=1}^{N_e} \ln(w_{s,jk}) l_s^e}{T_{obs}} \tag{2}$$

with

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

[5] Econometric models of schooling decisions estimated by Keane and Wolpin (1997, 2001), Eckstein and Wolpin (1999), Cameron and Heckman (1998, 2001), and Arcidiacono (2004, 2005) rely on a similar assumption.

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 monthly 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_{obs} is the total length of the observed labor market history of the individual.[6]Denoting by $U_{j,k}^w$ the utility of being in the work force given that k years of schooling have been completed in the post-secondary field j, we set:

$$U_{j,k}^w = \overline{\ln w}_{jk}^r = \alpha_w + X_{w,j,k}\beta_w + \epsilon_w \tag{3}$$

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. $X_{w,j,k}$ is a vector of observed characteristics that may affect labor market earnings, including post-secondary education, α_w represents the type-specific intercept, and ϵ_w denotes an independent random factor that affects the individual's earnings.

Let us now write the value functions associated with each schooling decision.

- The value of choosing major j is given by (Bellman,1957):

$$V_j^s = V_{j,1}^s = U_{j,1}^s + \beta \text{Emax}\left(V_{j,1}^w, V_{j,2}^s \mid X_s, X_w\right)$$

With

$$V_{j,1}^w = U_{j,1}^w$$

- The value of studying one more year to reach the level of education $k = 2$ is given by:

$$V_{j,2}^s = U_{j,2}^s + \beta \text{Emax}\left(V_{j,2}^w, V_{j,3}^s \mid X_s, X_w\right)$$

[6]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.

With

$$V_{j,2}^{w} = U_{j,2}^{w}$$

- The value of studying one more year to reach a level of education k, $k \in \{3, ..., K-1\}$,once level $k-1$ has been reached, is given by:

$$V_{j,k}^{s} = U_{j,k}^{s} + \beta \text{Emax}\left(V_{j,k}^{w}, V_{j,k+1}^{s} \mid X_s, X_w, d_2 = 1, ..., d_{k-1} = 1\right)$$

With

$$V_{j,k}^{w} = U_{j,k}^{w}$$

- Finally, the value of studying one more year to reach the last level of studying K, once $K-1$ has been reached, is given by:

$$V_{j,K}^{s} = U_{j,K}^{s} + \beta \text{Emax}\left(V_{j,K}^{w} \mid X_s, X_w, d_2 = 1, ..., d_{K-1} = 1\right)$$

With

$$V_{j,K}^{w} = U_{j,K}^{w}$$

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 equations are stochastically independent of these type-specific intercepts.

3.1 The econometric model

3.2 Stochastic assumptions

Residuals are supposed to be normally distributed. We assume that the random vector $(\epsilon_{s,1,1}, \ldots, \epsilon_{s,M,1})$ affecting the choice of the major, and the residuals $(\epsilon_{s,j,k})_{k>1}$ and $\epsilon_w$are independently distributed.[7] Consequently, the whole vector of residuals[8]

[7]Correlated unobserved heterogeneity across equations is captured by type-specific random intercepts $(\alpha_{s,j,k}^{r})_{j=1,\ldots,M;k=1,\ldots,K}$ and α_w^r.

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

is assumed to be distributed as:

$$\begin{pmatrix} \epsilon_{s,1,2} \\ ... \\ \epsilon_{s,1,K} \\ \epsilon_{s,2,2} \\ ... \\ \epsilon_{s,M,K} \\ \hline \epsilon_{s,2,1} - \epsilon_{s,1,1} \\ \epsilon_{s,3,1} - \epsilon_{s,1,1} \\ ... \\ \epsilon_{s,M,1} - \epsilon_{s,1,1} \\ \hline \epsilon_{w} \end{pmatrix} \sim \mathcal{N}(0, \Sigma)$$

where Σ is the $(MK) \times (MK)$ covariance matrix of the model residuals.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:

$$\Sigma = \Gamma.\Gamma' \tag{4}$$

where

$$\Gamma = \left(\begin{array}{c|ccccc|c} I_{M(K-1)} & 0 & 0 & ... & ... & & 0 \\ \hline 0_{1,M(K-1)} & 1 & 0 & 0 & ... & & 0 \\ 0_{1,M(K-1)} & \alpha_{11} & \exp(d_1) & 0 & ... & & 0 \\ ... & ... & ... & ... & ... & & .... \\ \hline 0_{1,M(K-1)} & 0 & ... & ... & 0 & & \exp(d_{M-1}) \end{array} \right) \tag{5}$$

Note that we impose the positivity of the diagonal terms of matrix Γ. Hence, the Cholesky decomposition of Σ is unique. $I_{M(K-1)}$ denotes the identity matrix with $M(K-1)$ rows and columns, $0_{1,M(K-1)}$ denotes a row of $M(K-1)$ zeros.

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 $k^* \in \{2, \ldots, K-1\}$, and who gets the average labor market log-earnings $\overline{\ln w}_{jk}^{r}$ is:

$$
\begin{aligned}
& \Pr\left(\forall j \neq j^*, V_{s,j^*} > V_{s,j}\right) \times \Pr\left(d_2 = 1, ..., d_{k^*} = 1, d_{k^*+1} = 0\right) \times f\left(V_{w,j^*,k^*}\right) \\
= \; & \Pr\left(\forall j \neq j^*, V_{s,j^*} > V_{s,j}\right) \times \prod_{k=1}^{k^*-1} \Pr\left(V_{s,j^*,k+1} > V_{w,j^*,k}\right) \\
& \qquad \times \Pr\left(V_{s,j^*,k^*+1} < V_{w,j^*,k^*}\right) \times f\left(V_{w,j^*,k^*}\right)
\end{aligned}
$$

For an individual entering the labor market after reaching only the educational level $k^* = 1$, the contribution to the likelihood is expressed as:

$$
\Pr\left(\forall j \neq j^*, V_{s,j^*} > V_{s,j}\right) \times \Pr\left(V_{s,j^*,2} < V_{w,j^*,1}\right) \times f\left(V_{w,j^*,1}\right)
$$

Finally, for an individual reaching the last level of education $k^* = K$, the contribution to the likelihood is expressed as:

$$
\Pr\left(\forall j \neq j^*, V_{s,j^*} > V_{s,j}\right) \times \prod_{k=1}^{K-1} \Pr\left(V_{s,j^*,k+1} > V_{w,j^*,k}\right) \times f\left(V_{w,j^*,K}\right)
$$

where

$$
\begin{aligned}
& \Pr\left(\forall j \neq j^*, V_{s,j*} > V_{s,j}\right) \\
= \; & \Pr\left[\forall j \neq j^*, \epsilon_{s,j^*,1} - \epsilon_{s,j,1} > \alpha_{s,j,1} - \alpha_{s,j^*,1} + X_s\left(\beta_{s,j} - \beta_{s,j^*}\right)\right. \\
& \left. + \beta\left(\text{Emax}\left(V_{w,j,1}, V_{s,j,2}\right) - \text{Emax}\left(V_{w,j^*,1}, V_{s,j^*,2}\right)\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
& \Pr\left(V_{s,j,k} > V_{w,j,k-1}\right) \\
= \; & \Pr\left[\epsilon_w - \epsilon_{s,j,k} < \alpha_{s,j,k} + X_s\beta_{s,j} + \beta\text{Emax}\left(V_{w,j,k}, V_{s,j,k+1}\right) - \left(\alpha_w + X_{w,j,k-1}\beta_w\right)\right] \\
= \; & \Phi\left(\frac{\alpha_{s,j} + X_s\beta_{s,j} + \beta\text{Emax}\left(V_{w,j,k}, V_{s,j,k+1}\right) - \left(\alpha_w + X_{w,j,k-1}\beta_w\right)}{\sqrt{\sigma_w^2 + \sigma_{s,j,k}^2}}\right),
\end{aligned}
$$

and

$$
f\left(V_{w,j,k}\right) = \frac{1}{\sigma_w}\varphi\left(\frac{V_{w,j,k} - \alpha_w - X_{w,j,k}\beta_w}{\sigma_w}\right)
$$

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:

$$\Pr\left(\forall j \neq j^*, V_{s,j^*} > V_{s,j}\right),$$

we use a method proposed by Train (2003).[9]

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.[10] As it is detailed in the 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^* and who gets the average labor market log-earnings $\overline{\ln w}_{j^*,k^*}$ follows a finite mixture distribution:

$$l(j^*, k^*, \overline{\ln w}_{j^*,k}) = \sum_{r=1}^{R} \Pi_r l(j^*, k^*, \overline{\ln w}^r_{j^*,k^*}|r) \tag{6}$$

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

3.4 Estimation

In order to explain our estimation strategy, let us introduce some further notations: θ_S denotes the schooling choices parameters, both in terms of major and level of education, and 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

[9]This method basically consists in completing the Cholesky matrix Γ by adding columns and rows of zeros:

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

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

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 $\left(\theta_S^{(n)}, \theta_W^{(n)}\right)$ of the parameter vector, the values $(\pi_r^{(n)})_{r=1...R}$ of the mixture distribution and the set of values $\Omega_r^{(n)} = \left(\alpha_{s,j,k}^{r,(n)}, \alpha_w^{r,(n)}\right)_{(r)}$ 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 :

$$\sum_{i=1}^{N} \sum_{r=1}^{R} \pi_{i,r}^{(n)} \ln l\left(j_i^*, k_i^*, w_i | T_i = r, (\Pi_r)_r, (\Omega_r^{(n)})_r, \theta_S, \theta_W\right) \tag{7}$$

▷ *M-step*

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

This maximization can be done in two successive steps.

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

$$\pi_r^{(n+1)} = \frac{\sum_{i=1}^{N} \pi_{ir}^{(n)}}{\sum_{l=1}^{R} \sum_{i=1}^{N} \pi_{il}^{(n)}} \tag{8}$$

Then, due to the partial separability of the conditional completed log-likelihood function (Arcidiacono and Jones, 2003), we get two sequential optimization problems since the residuals associated with schooling choices are assumed to be independent from the residuals associated with earnings. Henceforth:

$$\begin{aligned}
& \sum_{i=1}^{N}\sum_{r=1}^{R}\pi_{i,r}^{(n)}\ln l\left(f_i,l_i,w_i|T_i=r,(\Pi_r)_r,(\Omega_r)_r,\theta_S,\theta_W\right) \\
= & \sum_{i=1}^{N}\sum_{r=1}^{R}\pi_{i,r}^{(n)}\ln l\left(w_i|T_i=r,(\Pi_r)_r,(\alpha_w^r)_r,\theta_W\right) \\
+ & \sum_{i=1}^{N}\sum_{r=1}^{R}\pi_{i,r}^{(n)}\ln l\left(f_i,l_i|T_i=r,(\Pi_r)_r,(\alpha_w^r)_r,(\alpha_{s,j,k}^r)_r,\theta_S,\theta_W\right)
\end{aligned}$$

It implies that first, we maximize the log-wage equation. Then, given the estimates of this equation, we estimate the parameters of the post-secondary schooling choices. 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.

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 as:

$$\overline{\theta^*} = \frac{1}{H}\sum_{h=1}^{H}\theta_h^* \tag{9}$$

$$\sigma_{\theta^*} = \frac{1}{H-1}\sum_{h=1}^{H}(\theta_h^*-\overline{\theta^*})^2 \tag{10}$$

3.5 Identification

In order to ensure identification, we impose restrictions on the parameters[11]. Namely, as the probit model is only identified up to a scaling parameter, we set :

$$\forall j \in \{1, \ldots, J\}, \sigma_{s,j,2} = \sigma_{s,j,3} = \ldots = \sigma_{s,j,5} = 1 \tag{11}$$

We also impose the following restrictions on type-specific unobserved heterogeneity parameters:

$$\alpha_{s,1,5}^{r=1} = 0 \tag{12}$$

and finally:

$$\forall r \in \{1, \ldots, R\}, \alpha_{s,1,1}^{r} = 0 \tag{13}$$

In order to identify our model without relying on distributional assumptions, we use some exclusion restrictions. The most crucial one concerns the introduction into the wage equation of a dummy exogenous variable, which is assumed to affect directly only the wage and not the two other outcomes. In our application, it is chosen as an indicator of the residence in Paris (*Region Ile de France*)[12].Besides, the covariates indicating the father's and mother's professions (in 1998), the age of the student in 6^{th} and 12^{th} grades, 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.

4 A semi-structural approach

Let us now consider an alternative semi-structural model of post-secondary educational choices relying on a somewhat less extreme form of rationality. It will then be interesting to check the robustness of the results obtained with the structural dynamic model to this alternative specification.

The main difference with the structural specification comes from the underlying process that determines the educational level. Unlike the preceding model, this approach does not explicitly assume that the individual chooses when to leave university by computing and comparing the expected value of entering the labor market with the expected value of completing a supplementary year of schooling. Instead, we suppose that the length of studies k_j^* is generated by the following

[11]The need to impose those identifying constraints stems from the expression of the likelihood function.

[12]Arcidiacono(2004) uses a similar type of exclusion. In order to identify the effects of expected earnings on the choice of major, he incorporates average state earnings into the earnings equation.

reduced-form latent model:

$$k_j^* = \begin{cases} 1 & \text{if } \widetilde{k}_j^r \leq s_2 \\ \\ 2 & \text{if } s_2 < \widetilde{k}_j^r \leq s_3 \\ \vdots & \\ K & \text{if } s_{K+1} < \widetilde{k}_j^r \end{cases}$$

where $\widetilde{k}_j$ denotes the individual propensity to succeed in long post-secondary studies within major j, and $\{s_2, \ldots, s_{K+1}\}$ 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:

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

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. For instance, we assume that the average proportion of college students in the same major and in the same university may affect the length of studies.[13]

As for the choice of the major, 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 K 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 within field j $(j = 1, ..., M)$. Then, for a student of type r, the value V_j^r of major j can be written as :

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

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

where

$$v_{1j}^r = \alpha \sum_{k \in \{1,\dots,K\}} \Pr(k^* = k \mid r, j^* = j).E\left(V_{w,j,k}^r \mid r, j^* = j, k^* = k\right)$$

$E\left(V_{w,j,k}^r \mid r, j^* = j, k^* = k\right)$ denoting the expected earnings associated with education (j, k), for a student of type r, and α being an unknown sensibility parameter to be estimated. Note that $V_{w,j,k}$ refers to the value of being in the work force given that k years of schooling have been completed in the post-secondary field j and has the same expression as in the structural dynamic model.

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 or when graduating from high 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:

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

Finally, the stochastic assumptions as well as the estimation strategy are similar to those of the structural model. Residuals are supposed to be normally distributed, and the random vector affecting the choice of the major, the residuals v and ϵ entering the two other equations are supposed to be independently distributed.[14]We finally also rely on the EM algorithm with a sequential maximization step to estimate the model.

5 Data

The models presented above are estimated using French data coming from the "Génération 98" survey collected by CEREQ (*Centre d'Etudes et de Recherches*

[14]Correlated unobserved heterogeneity across equations is captured by type-specific random intercepts $(\alpha_{(1,j)}^r)_{j=1,\dots,M}$, α_2^r, and α_3^r.

sur les Qualifications, Marseille).[15] 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 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.

Our subsample of interest is constituted of respondents having at least passed the national high school final examination successfully:[16] 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 Appendix A) 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,[17] parents' nationality and profession.

Noteworthy, male students are more likely to attend majors in Sciences while

[15] 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.

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

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

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.[18] 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 16 and 17 reported in Appendix A). Unsurprisingly, 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 18 in Appendix A).

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

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

	Number	Percent
Major		
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

Source: Génération 1998 (CEREQ, Marseille)

Table 2: Descriptive statistics: covariates

	Number	Percent
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

Source: Génération 1998 (CEREQ, Marseille)

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

Field	Length	Average 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
Total		6.29

Source: Génération 1998 (CEREQ, Marseille)

6 Results

Tables 19 to 24 (reported in Appendix B.1) give the parameter estimates of the structural dynamic model. The estimates of the semi-structural model are reported in Tables 25 to 30 (Appendix B.2).

6.1 Structural dynamic model

Tables 19 to 21 report the parameter estimates associated with the individual covariates (X_s) affecting the propensity to pursue studies within each major[19]. Students whose father is a farmer, tradesman or a technician have a higher propensity to pursue studies within sciences majors. Besides, students whose father is a technician have a higher propensity to pursue studies within humanities and social sciences majors. Having a technician or blue-collar father is positively related to the probability of pursuing studies within law, economics and management majors. Similar overall effects are found for mother's profession. The gender of the student has no statistically significant effect on the probability to pursue studies within each of the three majors. Students whose both parents are French have a higher propensity to pursue studies in each major. Students who obtained a *Baccalauréat* (i.e. the terminal high-school diploma in France) from a vocational or a technological track as well as in humanities have a higher propensity to pursue studies within sciences major. Those who obtained a *Baccalauréat* in humanities also have a higher propensity to pursue studies within humanities and social sciences major. Students who were older than expected (i.e. 12 years old or above) at the entry into junior high-school (sixth grade) have a lower propensity to pursue studies within sciences as well as law, economics and management majors. Those who were younger than expected when graduating from high school (i.e. 17 years old or below) have a higher probability of pursuing studies within sciences and humanities and social sciences majors.

Table 22 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[20].

Tables 23 and 24 report the parameter estimates of the distribution of unobserved individual heterogeneity terms[21]. The first group of individuals represents 90,5 percent of the population of students, while the second group represents only 9,5 percent of the population. Comparing with the first group, individuals in the second group are characterized by a lower type-specific preference $\alpha_{2,1}$ for studies

[19]Note that it stems from the specification of the structural model that the covariates X_s affect both the choice of the major and the sequential decisions to pursue studies or to enter the labor market. Interpreting these estimates is therefore not straightforward.

[20]Nevertheless this increase is not significant. Estimating the earnings equation without the interactions between female gender and level as well as between *Ile-de-France* and level would probably provide more significant parameter estimates.

[21]Two types of heterogeneity are considered in this version. We are currently working on the estimation of the structural model with more types which will be more satisfying to control for unobserved heterogeneity.

in humanities and social sciences, and a higher preference $\alpha_{3,1}$ for studies in law, economics and management. Noteworthy, they are also characterized by a lower type-specific earnings intercept α_w. Finally, they also have a lower propensity to pursue MA level studies within sciences majors, graduate level studies within law,economics and management majors and a higher propensity to pursue graduate level studies when studying sciences.

The model fit is rather good. Table 4 shows nevertheless that the model overestimates the proportion of students in humanities and social sciences.

To get a more precise view of the effect of expected wages 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).[22]

In general, the impacts are quantitatively rather small and statistically significant. The lowest impacts concern the majors in law, economics and management. A 10% increase in the expected earnings associated with majors in sciences leads to an increase of 2.4 percentage points in the proportion of students in this major. This increase is mainly compensated by a decrease of 2.1 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, although very slightly higher, 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 slightly higher (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 2.8 percentage points in the proportion of students in these majors, this increase being mainly compensated by a decrease of about 2 percentage points in the proportion of students in sciences. Once again, a 10% decrease in expected earnings has slightly higher impacts on allocations.

A variation in the level of expected earnings associated with a post-secondary education in law, economics and management has quantitatively lower impacts. For instance, a 10% increase in the expected earnings associated with a post-secondary education in these majors results in an increase of 1.1 percentage points in the proportion of students in these majors, this increase being compensated by a decrease of 0.30 percentage points in the proportion of students in sciences, and by a decrease of 0.79 percentage points in the proportion of students in humanities and social sciences (see Table 6).

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

6.2 Semi-structural model

Let us now turn to the estimations of the semi-structural model. Under this alternative specification, we find even smaller but still significant effects of expected earnings on the allocation between majors. Hence, the main finding of a significant but quantitatively small impact of expected earnings on post-secondary major choices is robust to this alternative semi-structural specification.

Tables 25 and 26 report the parameter estimates of the equations generating the major choice. Students whose father is a farmer or a tradesman choose less frequently majors in humanities and social sciences. Those whose father is a technician or a white-collar worker choose more frequently majors in law, economics and management, while students whose father is blue-collar worker choose less frequently this major. Noteworthy, students whose mother is blue-collar worker choose more frequently majors in law, economics and management. Other professions of the parents have generally no effect on the major choice. The place of birth of the student and the nationality of his/her parents, as well as his/her gender, have no statistically significant effect on this choice. Students who obtained a *Baccalauréat* (i.e. the terminal high-school diploma in France) in sciences are less likely to choose a post-secondary major in law, economics and management. 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 post-secondary major in law, economics and management. Those who were younger than expected when graduating from high school (i.e. 17 years old or below) choose less frequently a major in humanities and social sciences, while those who were on time when graduating from high school (i.e. 18 years old) choose less frequently a major in law, economics and management. 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 29).

Most covariates have a significant impact on the length of post-secondary studies (see Table 27). For instance, students whose parents are white-collar or blue-collar workers leave more rapidly (i.e. at a lower level) the post-secondary educational system. Students whose both parents are French reach generally a higher level of post-secondary education. Students who were older than expected (19 years old or above) when leaving high-school (i.e. in twelfth grade) are more likely to drop earlier, while those 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 implemented by the university administration 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, there is no gender difference in the length of post-secondary studies.

Table 28 gives the parameter estimates of the (log-)earnings equation. On average, earnings are lower for females (especially before the BA degree) and they are higher in the region Ile-de-France (including Paris). Note that the latter positive effect is lower for those holding a BA degree. Mean (log-)earnings increase with the length of studies in post-secondary education. However, this increase is lower from the MA degree in the majors in humanities and social sciences.

Tables 29 and 30 report the parameter estimates of the distribution of unobserved individual heterogeneity terms. The first group of individuals represents approximately 59 percent of the population of students. Individuals in this group are characterized by the lowest type-specific preference $\alpha_{(1.2)}$ for studies in humanities and social sciences, and the highest type-specific propensity (or ability) α_2 to undertake long post-secondary studies. The second group represents 30 percent of the population of students. Individuals in this group are characterized by the highest type-specific preference $\alpha_{(1.2)}$ for studies in humanities and social sciences and the lowest preference $\alpha_{(1.3)}$ for studies in law, economics and management. They also have the highest type-specific earnings intercept α_3. The third group represents 11 percent of the population; it is characterized by the lowest type-specific propensity (or ability) α_2 to undertake long post-secondary studies as well as by the lowest type-specific productivity term α_3.

The model fit is very good. Table 7 shows that the model very slightly underestimates the proportion of students in sciences.

To get a more precise view of the effect of expected wages 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 7 to 9 below).[23]

In general, the impacts are quantitatively small even though they are statistically significant. The lowest impacts concern the majors in law, economics and management. A 10% increase in the expected earnings associated with majors in sciences leads to an increase of 0.4 percentage points in the proportion of students in this major. This increase is mainly compensated by a decrease of 0.31

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

percentage points in the proportion of students in humanities and social sciences (see Table 7).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 slightly higher (see Table 8). For instance, a 10% increase in the expected earnings associated with a post-secondary in these majors results in an increase of 0.46 percentage points in the proportion of students in these majors, this increase being mainly compensated by a decrease of about 0.31 percentage points in the proportion of students in sciences. Once again, a 10% decrease in expected earnings has almost symmetric impacts on allocations.

A variation in the level of expected earnings associated with a post-secondary education in law, economics and management has quantitatively lower impacts. For instance, a 10% increase in the expected earnings associated with a post-secondary education in these majors results in an increase of 0.24 percentage points in the proportion of students in these majors, this increase being compensated by a decrease of 0.09 percentage points in the proportion of students in sciences, and by a decrease of 0.15 percentage points in the proportion of students in humanities and social sciences (see Table 9).

Table 4: Simulation of a 10% variation in expected earnings of the majors in sciences (structural dynamic model)

	Observed proportions	Predicted proportions	$(p^S - p^P)$	$\widehat{\sigma}_{(p^S - p^P)}$ Standard error
Sciences				
	10% *increase*			
Sample distribution				
Sciences	25.94	22.88	2.392	0.047
Humanities and Social Sciences	40.80	47.04	-2.077	0.028
Law, Economics and Management	33.25	30.07	-0.315	0.024
	10% *decrease*			
Sample distribution				
Sciences	25.94	22.88	-2.507	0.060
Humanities and Social Sciences	40.80	47.04	2.191	0.039
Law, Economics and Management	33.25	30.07	0.316	0.026

Source: Génération 1998 (CEREQ)

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

Table 5: Simulation of a 10% variation in expected earnings of the majors in humanities and social sciences (structural dynamic model)

	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	25.94	22.88	-2.004	0.036
Humanities and Social Sciences	40.80	47.04	2.797	0.046
Law, Economics and Management	33.25	30.07	-0.793	0.038
10% *decrease*				
Sample distribution				
Sciences	25.94	22.88	2.280	0.031
Humanities and Social Sciences	40.80	47.04	-3.130	0.048
Law, Economics and Management	33.25	30.07	0.850	0.042

Source: Génération 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 (structural dynamic model)

	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	25.94	22.88	-0.303	0.024
Humanities and Social Sciences	40.80	47.04	-0.788	0.039
Law, Economics and Management	33.25	30.07	1.092	0.049
10% *decrease*				
Sample distribution				
Sciences	25.94	22.88	0.328	0.026
Humanities and Social Sciences	40.80	47.04	0.858	0.041
Law, Economics and Management	33.25	30.07	-1.186	0.051

Source: Génération 1998 (CEREQ).

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

Table 7: Simulation of a 10% variation in expected earnings of the majors in sciences (semi-structural model)

	Observed proportions	Predicted proportions	$(p^S - p^P)$	$\widehat{\sigma}_{(p^S-p^P)}$ Standard error
Sciences				
	10% *increase*			
Sample distribution				
Sciences	25.94	25.82	0.404	0.016
Humanities and Social Sciences	40.80	40.67	-0.312	0.014
Law, Economics and Management	33.25	33.51	-0.091	0.005
	10% *decrease*			
Sample distribution				
Sciences	25.94	25.82	-0.444	0.017
Humanities and Social Sciences	40.80	40.67	0.344	0.015
Law, Economics and Management	33.25	33.51	0.100	0.005

Source: Génération 1998 (CEREQ)

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

Table 8: Simulation of a 10% variation in expected earnings of the majors in humanities and social sciences (semi-structural model)

	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	25.94	25.82	-0.312	0.014
Humanities and Social Sciences	40.80	40.67	0.461	0.018
Law, Economics and Management	33.25	33.51	-0.149	0.009
	10% *decrease*			
Sample distribution				
Sciences	25.94	25.82	0.344	0.016
Humanities and Social Sciences	40.80	40.67	-0.507	0.020
Law, Economics and Management	33.25	33.51	0.163	0.009

Source: Génération 1998 (CEREQ).

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

Table 9: Simulation of a 10% variation in expected earnings of majors in law, economics and management (semi-structural model)

	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	25.94	25.82	-0.091	0.005
Humanities and Social Sciences	40.80	40.67	-0.148	0.009
Law, Economics and Management	33.25	33.51	0.239	0.013
	10% *decrease*			
Sample distribution				
Sciences	25.94	25.82	0.100	0.005
Humanities and Social Sciences	40.80	40.67	0.164	0.009
Law, Economics and Management	33.25	33.51	-0.264	0.014

Source: Génération 1998 (CEREQ).

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

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.

BELLMAN R. (1957) : *Dynamic Programming*. Princeton: Princeton University Press.

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.

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.

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.

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

Table 10: 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
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 is not French	16.27	44.08	39.64
Both parents are 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
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

Table 11: Distribution of various subgroups across majors (in percent, 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)

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

Table 12: Distribution of various subgroups across educational levels (in percent, 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 various subgroups across educational levels (in percent, 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)

Remarks: Lines sum up to 100%.

Table 14: 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 15: 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 16: Type of the labor contract in the first job (in percent)

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
	Interim	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	-	-	-
	Interim	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
	Interim	22.09	21.05	12.61	10.80	2.39
	Emploi jeune	8.84	8.77	-	-	-
	Civil servant	-	-	10.08	6.82	6.44

Remarks: Columns sum up to 100%. *Emplois jeunes* are publicly subsidized jobs, *Contrats de qualification* are workplace employment programs.

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

Table 17: Average number of months in nonemployment before the first job

		Average number of months
Sciences	Dropout	6.23
	2 years of college	5.55
	BA degree	4.79
	MA degree	5.47
	Graduates	3.55
Humanities and Social Sciences	Dropout	5.79
	2 years of college	5.05
	BA degree	6.49
	MA degree	6.36
	Graduates	3.53
Law, Economics and Management	Dropout	5.61
	2 years of college	5.01
	BA degree	4.41
	MA degree	5.42
	Graduates	3.56

Source: Génération 1998 (CEREQ, Marseille)

Table 18: Average monthly log-earnings

	Average monthly log-earnings
Out of the region Ile-de-France	6.22
In the region Ile-de-France	6.68
At least one of the parents born abroad	6.21
Both parents born in France	6.30
Male	6.50
Female	6.13

Source: Génération 1998 (CEREQ, Marseille)

B Parameter estimates

B.1 Structural dynamic model

Table 19: Schooling parameters

Covariates	Estimate	Standard Error
Sciences		
Father's profession (in 1998)		
Executive	*Ref*	*Ref*
Farmer or tradesman	0.051	0.018
Technician	0.070	0.018
White-collar	0.017	0.020
Blue-collar	0.038	0.020
Unknown	0.009	0.024
Mother's profession (in 1998)		
Executive	*Ref*	*Ref*
Farmer or tradesman	0.103	0.031
Technician	0.092	0.030
White-collar	0.058	0.017
Blue-collar	0.074	0.025
Unknown	0.055	0.020
Born abroad	0.039	0.052
Woman	0.012	0.022
Both parents are French	0.108	0.029
Age in 6^{th} *grade*		
≤ 10	-0.017	0.019
11	*Ref*	*Ref*
≥ 12	-0.062	0.030
Age in 12^{th} *grade*		
≤ 17	0.039	0.018
18	*Ref*	*Ref*
19	-0.005	0.018
≥ 20	0.008	0.020
Baccalauréat		
General, sciences	*Ref*	*Ref*
General, humanities	0.036	0.015
General, economics	0.024	0.014
Vocational or technological	0.036	0.017

Table 20: Schooling parameters

Covariates	Estimate	Standard Error
Humanities and Social Sciences		
Father's profession (in 1998)		
Executive	*Ref*	*Ref*
Farmer or tradesman	0.031	0.017
Technician	0.051	0.019
White-collar	-0.011	0.017
Blue-collar	0.018	0.017
Unknown	-0.017	0.023
Mother's profession (in 1998)		
Executive	*Ref*	*Ref*
Farmer or tradesman	0.114	0.021
Technician	0.108	0.028
White-collar	0.061	0.016
Blue-collar	0.065	0.024
Unknown	0.060	0.020
Born abroad	0.059	0.030
Woman	0.004	0.021
Both parents are French	0.087	0.022
Age in 6^{th} *grade*		
≤ 10	-0.013	0.021
11	*Ref*	*Ref*
≥ 12	-0.028	0.021
Age in 12^{th} *grade*		
≤ 17	0.041	0.019
18	*Ref*	*Ref*
19	0.005	0.012
≥ 20	0.018	0.016
Baccalauréat		
General, sciences	*Ref*	*Ref*
General, humanities	0.031	0.015
General, economics	0.009	0.015
Vocational or technological	0.018	0.015

Table 21: Schooling parameters (end)

Covariates	Estimate	Standard Error
Law, Economics and Management		
Father's profession (in 1998)		
Executive	*Ref*	*Ref*
Farmer or tradesman	0.042	0.022
Technician	0.054	0.026
White-collar	-0.019	0.016
Blue-collar	0.066	0.021
Unknown	-0.006	0.029
Mother's profession (in 1998)		
Executive	*Ref*	*Ref*
Farmer or tradesman	0.017	0.040
Technician	0.042	0.036
White-collar	0.016	0.024
Blue-collar	-0.005	0.039
Unknown	0.030	0.029
Born abroad	-0.050	0.056
Woman	0.035	0.025
Both parents are French	0.075	0.025
Age in 6^{th} *grade*		
≤ 10	0.009	0.024
11	*Ref*	*Ref*
≥ 12 years	-0.046	0.020
Age in 12^{th} *grade*		
≤ 17	-0.014	0.023
18	*Ref*	*Ref*
19	-0.030	0.019
≥ 20	0.013	0.020
Baccalauréat		
General, sciences	*Ref*	*Ref*
General, humanities	0.020	0.024
General, economics	0.019	0.019
Vocational or technological	0.020	0.017

Source: Génération 1998 (CEREQ. Marseille)

Table 22: Earnings equation

Covariates	Estimate	St. Error
Intercept	3.798	0.052
Both parents are French	0.083	0.045
Region Ile de France	0.326	0.069
Female	-0.257	0.052
Born abroad	0.023	0.078
Field of studies		
Sciences	*Ref*	*Ref*
Humanities and Social Sciences	0.023	0.049
Law. Economics and Management	0.018	0.056
Level of studies		
Dropout	*Ref*	*Ref*
Two years of college	0.064	0.104
Licence (BA degree)	0.036	0.092
Maitrise (MA degree)	0.011	0.084
Post Maitrise (Graduates)	0.040	0.078
Interactions between field and level		
Humanities and Social Sciences		
Dropout	*Ref*	*Ref*
Two years of college	-0.067	0.124
Licence (BA degree)	-0.048	0.090
Maitrise (MA degree)	-0.028	0.088
Post Maitrise (Graduates)	-0.050	0.082
Law, Economics and Management		
Dropout	*Ref*	*Ref*
Two years of college	-0.063	0.120
Licence (BA degree)	-0.021	0.109
Maitrise (MA degree)	-0.014	0.088
Post Maitrise (Graduates)	-0.027	0.087
Interactions between female gender and level		
Dropout	*Ref*	*Ref*
Two years of college	0.001	0.134
Licence (BA degree)	0.009	0.090
Maitrise (MA degree)	0.017	0.086
Post Maitrise (Graduates)	-0.003	0.085
Interactions between region Ile-de-France and level		
Dropout	*Ref*	*Ref*
Two years of college	-0.031	0.149
Licence (BA degree)	-0.040	0.133
Maitrise (MA degree)	0.023	0.104
Post Maitrise (Graduates)	-0.003	0.127

Source: Génération 1998 (CEREQ. Marseille)

Table 23: Other parameters

Covariance matrix of residuals for major choices

$$\begin{pmatrix} 1 & 0.376 \\ (-) & \textit{(0.029)} \\ 0.376 & 9.997 \\ \textit{(0.029)} & \textit{(0.962)} \end{pmatrix}$$

	Estimate	St. Error
σ (standard error of earnings equation)	0.802	0.011
β (discount parameter)	0.992	0.001
Type probabilities		
Type 1	0.905	0.005
Type 2	0.095	0.005

Source: Génération 1998 (CEREQ)

Table 24: Type-specific heterogeneity parameters (schooling and earnings specific)

	Estimate	St. Error
Schooling heterogeneity parameters		
Type 1		
$\alpha_{s,1,1}$	0.000	-
$\alpha_{s,1,2}$	-1.289	0.097
$\alpha_{s,1,3}$	0.318	0.117
$\alpha_{s,1,4}$	0.311	0.092
$\alpha_{s,1,5}$	0.000	-
$\alpha_{s,2,1}$	0.733	0.098
$\alpha_{s,2,2}$	-1.350	0.084
$\alpha_{s,2,3}$	0.267	0.095
$\alpha_{s,2,4}$	0.377	0.089
$\alpha_{s,2,5}$	0.304	0.066
$\alpha_{s,3,1}$	-1.063	0.065
$\alpha_{s,3,2}$	-1.151	0.090
$\alpha_{s,3,3}$	0.496	0.095
$\alpha_{s,3,4}$	0.019	0.099
$\alpha_{s,3,5}$	0.508	0.103
Type 2		
$\alpha_{s,1,1}$	0.000	-
$\alpha_{s,1,2}$	-1.385	0.089
$\alpha_{s,1,3}$	0.397	0.107
$\alpha_{s,1,4}$	0.164	0.098
$\alpha_{s,1,5}$	0.275	0.094
$\alpha_{s,2,1}$	0.488	0.073
$\alpha_{s,2,2}$	-1.146	0.086
$\alpha_{s,2,3}$	0.177	0.096
$\alpha_{s,2,4}$	0.243	0.086
$\alpha_{s,2,5}$	0.338	0.065
$\alpha_{s,3,1}$	-0.350	0.102
$\alpha_{s,3,2}$	-1.210	0.096
$\alpha_{s,3,3}$	0.383	0.098
$\alpha_{s,3,4}$	0.123	0.101
$\alpha_{s,3,5}$	0.265	0.107
Earnings heterogeneity parameters		
Type 1		
α_w	2.568	0.036
Type 2		
α_w	-1.005	0.044

Source: Génération 1998 (CEREQ)

B.2 Semi-structural model

Table 25: Choice of the major (beginning)

Covariates	Estimate	Standard Error
Sciences	*Ref*	*Ref*
Humanities and Social Sciences		
Father's profession (in 1998)		
Executive	*Ref*	*Ref*
Farmer or tradesman	-0.244	0.099
Technician	-0.403	0.274
White-collar	-0.113	0.130
Blue-collar	-0.332	0.303
Unknown	-0.223	0.094
Mother's profession (in 1998)		
Executive	*Ref*	*Ref*
Farmer or tradesman	-0.111	0.217
Technician	-0.130	0.109
White-collar	-0.422	0.270
Blue-collar	0.178	0.140
Unknown	0.213	0.288
Born abroad	-0.154	0.182
Woman	-0.363	0.348
Both parents are French	-0.028	0.172
Age in 6^{th} grade		
≤ 10	-0.412	0.269
11	*Ref*	*Ref*
≥ 12	-0.182	0.106
Age in 12^{th} grade		
≤ 17	-0.373	0.180
18	*Ref*	*Ref*
19	-0.153	0.162
≥ 20	-0.002	0.299
Baccalauréat		
General, sciences	*Ref*	*Ref*
General, humanities	-0.200	0.142
General, economics	0.090	0.252
Vocational or technological	-0.038	0.214

Table 26: Choice of the major (end)

Covariates	Estimate	Standard Error
Law, Economics and Management		
Father's profession (in 1998)		
Executive	*Ref*	*Ref*
Farmer or tradesman	-0.375	0.370
Technician	0.596	0.070
White-collar	0.544	0.194
Blue-collar	-0.299	0.135
Unknown	-0.706	0.249
Mother's profession (in 1998)		
Executive	*Ref*	*Ref*
Farmer or tradesman	-0.017	0.129
Technician	0.090	0.328
White-collar	-0.180	0.159
Blue-collar	0.711	0.330
Unknown	0.105	0.135
Born abroad	0.048	0.266
Woman	0.110	0.084
Both parents are French	-0.278	0.216
Age in 6^{th} *grade*		
≤ 10	0.183	0.105
11	*Ref*	*Ref*
≥ 12 years	-0.622	0.242
Age in 12^{th} *grade*		
≤ 17	2.881	0.108
18	*Ref*	*Ref*
19	1.628	0.233
≥ 20	2.158	0.095
Baccalauréat		
General, sciences	*Ref*	*Ref*
General, humanities	4.697	0.196
General, economics	1.329	0.097
Vocational or technological	3.108	0.194

Source: Génération 1998 (CEREQ. Marseille)

Table 27: Equation for the length of studies

Covariates	Estimate	Standard Error
Father's profession (in 1998)		
Farmer or tradesman	-0.162	0.049
Executive	*Ref*	*Ref*
Technician	-0.165	0.063
White-collar	-0.387	0.055
Blue-collar	-0.425	0.052
Unknown	-0.168	0.065
Mother's profession (in 1998)		
Farmer or tradesman	0.086	0.105
Executive	*Ref*	*Ref*
Technician	-0.161	0.094
White-collar	-0.054	0.057
Blue-collar	-0.131	0.084
Unknown	0.126	0.080
Born abroad	0.252	0.135
Woman	0.029	0.043
Both parents are French	0.382	0.069
Age in 6^{th} *grade*		
≤ 10	0.221	0.063
11	*Ref*	*Ref*
≥ 12	0.086	0.083
Age in 12^{th} *grade*		
≤ 17	0.029	0.068
18	*Ref*	*Ref*
19	-0.387	0.041
≥ 20	-0.688	0.056
Baccalauréat		
General, sciences	*Ref*	*Ref*
General, humanities	-0.778	0.048
General, economics	-0.403	0.051
Vocational or technological	-1.192	0.056
Proportion of students in college	-1.287	0.135

Source: Génération 1998 (CEREQ, Marseille)

Table 28: Earnings equation

Covariates	Estimate	St. Error
Intercept	3.611	0.046
Both parents are French	0.072	0.047
Region Ile de France	0.267	0.080
Female	-0.251	0.051
Born abroad	0.102	0.097
Field of studies		
Sciences	*Ref*	*Ref*
Humanities and Social Sciences	-0.076	0.051
Law. Economics and Management	-0.044	0.056
Level of studies		
Dropout	*Ref*	*Ref*
Two years of college	0.219	0.092
Licence (BA degree)	0.418	0.074
Maitrise (MA degree)	0.231	0.078
Post Maitrise (Graduates)	0.803	0.055
Interactions between field and level		
Humanities and Social Sciences		
Dropout	*Ref*	*Ref*
Two years of college	-0.175	0.112
Licence (BA degree)	-0.059	0.083
Maitrise (MA degree)	-0.095	0.091
Post Maitrise (Graduates)	-0.173	0.077
Law, Economics and Management		
Dropout	*Ref*	*Ref*
Two years of college	0.004	0.105
Licence (BA degree)	0.047	0.087
Maitrise (MA degree)	0.128	0.082
Post Maitrise (Graduates)	0.069	0.075
Interactions between female gender and level		
Dropout	*Ref*	*Ref*
Two years of college	0.152	0.100
Licence (BA degree)	0.061	0.078
Maitrise (MA degree)	0.201	0.074
Post Maitrise (Graduates)	-0.077	0.064
Interactions between region Ile-de-France and level		
Dropout	*Ref*	*Ref*
Two years of college	-0.007	0.165
Licence (BA degree)	-0.254	0.120
Maitrise (MA degree)	0.088	0.101
Post Maitrise (Graduates)	0.012	0.084

Source: Génération 1998 (CEREQ. Marseille)

Table 29: Other parameters

Covariance matrix of residuals

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -0.105 & 0 \\ (-) & (-) & \mathit{(0.115)} & (-) \\ 0 & -0.105 & 13.075 & 0 \\ (-) & \mathit{(0.115)} & \mathit{(1.337)} & (-) \\ 0 & 0 & 0 & 0.772 \\ (-) & (-) & (-) & \mathit{(0.008)} \end{pmatrix}$$

	Estimate	St. Error
Thresholds		
s_2	-3.187	0.153
s_3	-2.775	0.151
s_4	-2.198	0.157
s_5	-1.491	0.151
α	0.287	0.011
Type probabilities		
Type 1	0.587	0.005
Type 2	0.301	0.004
Type 3	0.111	0.006

Source: Génération 1998 (CEREQ)

Table 30: Type-specific heterogeneity parameters

	Estimate	St. Error
Type 1		
$\alpha_{(1.1)}$	0.000	-
$\alpha_{(1.2)}$	-1.057	0.145
$\alpha_{(1.3)}$	-1.666	0.199
α_2	0.000	-
α_3	2.751	0.022
Type 2		
$\alpha_{(1.1)}$	0.000	-
$\alpha_{(1.2)}$	-0.158	0.158
$\alpha_{(1.3)}$	-2.705	0.170
α_2	-0.822	0.023
α_3	2.891	0.020
Type 3		
$\alpha_{(1.1)}$	0.000	-
$\alpha_{(1.2)}$	-0.472	0.191
$\alpha_{(1.3)}$	-1.796	0.257
α_2	-1.360	0.059
α_3	-0.589	0.031

Source: Génération 1998 (CEREQ)