CENTRE POUR LA RECHERCHE ECONOMIQUE ET SES APPLICATIONS

Document de travail (Docweb) n°2304

Is there a devaluation of degrees ? Unobserved heterogeneity in returns to education and early experience

Damiano ARGAN, Robert GARY-BOBO, Marion GOUSSÉ

Mai 2023

Observe-t-on une dévaluation des diplômes ? Hétérogènéité inobservée des rendements de l'éducation et de l'expérience en début de carrière[1]

Damiano Argan[2], Robert Gary-Bobo[3], Marion Goussé[4]

Résumé : Nous montrons que les salaires réels espérés associés à certains diplômes de l'enseignement supérieur ont diminué en valeur absolue en France, au cours des deux décennies passées, et que cette baisse n'est pas due à l'anti-sélection. Pour étudier le rendement salarial des diplômes et de l'expérience, nous supposons l'existence d'un nombre fini de types latents et estimons un modèle de mélange fini de lois. Chaque type a sa propre équation de log-salaires, d'accumulation de l'expérience effective et de choix d'éducation. Cela permet la décomposition des effets de traitement de l'éducation comme une moyenne d'effets conditionnels au type. On montre alors que certains types ont connu une baisse de salaire réel tandis que d'autres ont profité d'une hausse, avec le même niveau de diplôme. L'aplatissement observé des rendements de l'expérience est elle-même hétérogène. Dans le cas des diplômes de master, la distribution des types latents montre que la sélection des étudiants s'est améliorée avec le temps, en dépit du fait que le nombre de diplômés a fortement augmenté. Un excès d'offre de diplômés semble donc être l'hypothèse la plus vraisemblable pour expliquer la dévaluation du master.

Mots-clés : Salaire ; Rendements de l'éducation ; Rendements de l'expérience ; Capital Humain ; Sélection : hétérogénéité inobservée ; Types latents

Moving Opportunity Local Connectivity and Spatial Inequality

Abstract : We show that the expected real wages commanded by some higher-education degrees decreased in absolute terms in France, in the past two decades, and that this drop is not due to adverse selection. To study the returns to degrees and experience, we assume the existence of a finite number of latent types and estimate a finite-mixture model. Each type has its own log-wage equation, experience-accumulation and education-choice equation. This allows us to decompose the treatment effects of education as an average of type-dependent effects. We then show that some unobserved types experienced a real-wage drop while others benefited from an increase, with the same degree. The observed "flattening" of returns to experience is also heterogeneous. In the case of Master degrees, the estimated distribution of

[1]This work has received financial support fromCEPREMAP Axe 2.

[2]European University Institute, Florence, Italy, E-mail: damiano.argan@eui.eu

[3]Université Paris 1 Panthéon-Sorbonne, Centre d'Economie de la Sorbonne, 106 boulevard de l'Hopital, 75013 Paris, France, CREST, 5 avenue Henry Le Châtelier, 91120 Palaiseau, France, CES Ifo and CEPR, E-mail: garybobo@univ-paris1.fr

[4]CREST, 5 avenue Henry Le Châtelier, 91120 Palaiseau, France, ENSAI, 51 Rue Blaise Pascal, BP 37203, 35172 Bruz cedex, France., IZA, E-mail: marion.gousse@ensai.fr

latent types indicates that student selection improved with time, in spite of the fact that the number of graduates increased substantially. An excess supply of graduates might therefore be a likely explanation for the devaluation of Master's degrees.

Keywords : Wages; Returns to Education; Returns to Experience; Human Capital; Selection; Unobserved Heterogeneity; Finite-Mixture Models; Latent Types.

1 Introduction

In the past decades, the enrollment of Universities and Colleges has grown substantially in many countries.[1] The growth in the number of persons who reached tertiary education between 2010 and 2020 is impressive.[2] Various authors have argued that this growth has the potential to cause an excess supply of graduates, and hence a decrease in the real wages of College graduates relative to high-school graduates (*i.e.*, a drop in *college premia*).[3]

In the following, we study the returns to education and experience of young French men. We first show that some higher-education degrees, mainly the Master degrees, command on average a smaller real wage in the period 2010-2017, as compared to the period 1998-2005, and we propose a study of the causes of this drop. We model the unobserved heterogeneity with a system of latent types in order to address the important endogeneity problems of education and experience. We find that the observed drop in the return to Master's degrees is not due to a deterioration of the average quality of students (*i.e.*, adverse selection), in spite of enrollment growth.

For convenience, we define the *devaluation* of a given degree as an absolute decrease in the average real-wage of the holders of this degree (all other things being equal). Devaluation is commonly measured in relative terms, taking the form of a drop in the College wage premium, which we also observe, but for some categories of degrees, the drop can be absolute, as we will see below. When a devaluation of degrees is observed, it is an open question to disentangle the possible effects of an excess supply of graduates, the change due to a lesser quality of teaching, and finally, the variation caused by a less favorable selection of students. In the present article, using an econometric model with unobserved individual types, allowing for unobserved student heterogeneity, we conclude that, in the past 25 years, in France, the observed devaluation of Master's degrees does not seem to be due to a deterioration of the selection of students. On the contrary, it seems that the quality of selected students has improved, while at the same time, the number of graduates did increase substantially. We also show that the decrease in average real wages, conditional on some degree or level of education, is an average of heterogeneous variations through time: some unobserved types of students do not suffer from devaluation, while the real wage of others decreases. These variations are in turn due to changes in returns to education and returns to experience that are themselves heterogeneous.

In essence, the present article builds and estimates a model explaining individual wages, individual employment rates and education choices simultaneously with the help of panel data. We assume

[1] A "big push" occurred. According to OECD figures, in the United States, 7.7% of the population aged 25 or more had graduated from College in 1960, as compared to 37.5% in 2020.

[2] *Education at a Glance*, oecd-ilibrary.org, 2022.

[3] See our discussion of the literature below.

the existence of a finite number of latent individual types. Each type has a specific (*i.e.*, type-dependent) log-wage equation, a specific employment-rate equation and a specific discrete-choice model describing educational investment. In other words, the model is the product of three finite-mixture models for respectively, wages, employment and education, describing the accumulation of effective experience and the returns to experience of each latent type, as well as type-dependent returns to degrees.

We then assume that error terms are normally distributed in the log-wage and employment equations (resp., extreme-value distributed in the education-choice equation) *conditional* on observable and unobservable characteristics. The model is flexible: there are no cross-equation or cross-type restrictions and it is well-known that any smooth distribution of wages can be approximated, to any desired degree of precision, by a mixture of normal distributions. Thus, we assume that the endogeneity problems, typically arising in standard econometric regressions such as the Mincer equation, are entirely driven by the unobserved types: error terms are assumed independent of controls and types. In a nutshell, in this type of structure, identification of type-dependent parameters and of the distribution of types essentially relies on the panel structure, since each individual is typically observed more than twice, but we do not use instruments for identification.[4] We discuss the conditions for *nonparametric* identification of the model below, relying on results giving conditions for the nonparametric identification of finite latent structures.

The model is estimated by straightforward likelihood maximization. Good preliminary estimates are generated by means of a sequential EM algorithm. Given that almost all model coefficients are type-dependent, the number of parameters increases quickly with the number of latent types, but we estimated a model with 3 types by Maximum Likelihood. We discuss the choice of the number of types and show that three types is a reasonable choice. In particular, the three types provide a surprisingly good *classification* of individuals (with low entropy). The gains of an additional type are small.

An important output of finite mixture models is the probability of belonging to a given type, conditional on the individual's observed characteristics (hereafter the *individual posterior probabilities of types*). These probabilities show the most likely type of each individual. Heckman and his co-authors (see, *e.g.*, Heckman and Vytlacil (2005), and more recently, Heckman, Humphries, and Veramendi (2018)), propose to derive and compute *policy-relevant parameters* from the output of a structural model, in the presence of treatment-response heterogeneity.[5] Our model's estimated coefficients and the posterior probabilities of types allow us to compute a number of policy-relevant

[4]Instruments, could be added without difficulty, but we would then allow for a type-dependent impact of the instruments.

[5]We do not use the same model as Heckman and his co-authors, but our philosophy is similar, and the influence is direct.

parameters very easily. Posterior probabilities can be used as a system of weights to evaluate treatment effects. In particular, considering education as a treatment, when wages are the outcome, we can compute the ATT and ATE[6] of a certain level of education (or of a certain category of degree) and compare the two. The estimated model also allows us to compute ATEs conditional on the latent type and to uncover the heterogeneity of effects across types.

We estimate the model on a rich panel of young French workers: the *Generation surveys*. In these data, we follow the first seven years of career of three cohorts of workers. Each cohort is defined by the year during which the worker left the educational system, namely, 1998, 2004 and 2010. We show the existence of heterogeneity in the returns to experience across types. We find an erosion of returns to effective experience on average (and a corresponding flattening of wage curves), but we also find that the returns to experience of one of the types increased while that of the other types decreased.[7] We believe that the observed absolute devaluation of University Master's degrees, in France, is most likely due to an excess supply of graduates, because we find that the selection of students has improved with time. In contrast, we find that in the French business schools, the enrollment of which has also grown substantially, the quality of student selection has decreased with time. Then, we use simulations of the model to generate fictitious careers and compute discounted expected earnings over the first 7 years of career, type by type. Simulation results typically confirm the findings and give a synthetic view of degree devaluation.

Our model is relatively simple and easy to estimate. It can be called semi-structural: we do not explicitly model the sequential choice process of individuals (Heckman, Humphries, and Veramendi (2018) describe their work as developing "a methodological middle ground between the reduced-form treatment approach and the fully structural dynamic discrete choice approach"). Our description of education choices is essentially static and our employment equation (experience-accumulation model) is a kind of reduced form.[8] The types that we find are easily interpretable: there is an obvious ranking of types in terms of returns to education.

Literature. Empirical work confirms that returns to higher education have recently decreased. See, for instance, Valletta (2018), Emmons, Kent, Ricketts, et al. (2019). In the United States, the 80s and 90s have been characterized by the rise of the College skill premium. Increased inequalities have been attributed to skill-biased technical change. The work of Katz and Murphy (*i.e.*, Katz and Murphy (1992)) shows that a "standard" model is able to capture the evolution of the hierarchy of wages as the result of an increased demand of employers for graduates (and for the employment of

[6] *Average treatment effect on the treated* and *Average treatment effect*, resp.

[7] It follows that the devaluation of degrees may be underestimated for at least two reasons: firstly because it is a dynamic phenomenon, taking a few years of career to reveal itself fully, and secondly, because a subset of types do suffer from devaluation while others do not.

[8] But of course, simplicity comes with some benefits in terms of tractability and interpretation.

women). Fluctuations of the skill premium are directly related to the supply of graduates. Card and Lemieux (2001) have then showed differences in the evolution of skill premia across age groups and emphasized that the main force favoring the relative wages of younger graduates is the smaller growth of their number in the generations born after 1950 in the US, the UK and Canada. Goldin and Katz (2008) propose a historical view of wages over more than a century in the US.[9] This line of research has led to an analysis of labor-market *polarization*, and the so-called "Ricardian" model of the allocation of skills to tasks, allowing a study of occupational downgrading (see, *e.g.*, Acemoglu (1999), Autor, Katz, and Kearney (2008), Acemoglu and Autor (2011)).

In the UK, Blundell, Green, and Jin (2022) propose an explanation for the fact that the proportion of UK workers with university degrees tripled between 1993 and 2015 while simultaneously the time trend in the college wage premium remained flat: during the period, firms opted for more decentralized organization forms, UK firms took advantage of an increased supply of graduates and chose to pick up the technologies and organizational forms already developed in the US. Ichino, Rustichini, and Zanella (2022) study the higher education expansion in the UK from 1960 to 2004 with the help of a general equilibrium Roy model. They find that the expansion is associated with a decline of the average intelligence of graduates and that it mainly benefited relatively less intelligent students from advantaged socioeconomic backgrounds.

Until the turn of the millenium, facts were giving the impression that the evolution of wages and skill-premia had been different in France and in the US, with no development of inequalities due to higher education in the former country. Indeed, the work of Verdugo (2014) shows that France has experienced a *great compression* of the hierarchy of wages until 2008. But, finally, it may be that similar phenomena have been at work in the two countries and in the recent years. In the United States, Beaudry, Green, and Sand (2014, 2016) have shown the existence of a trend shift around the year 2000. The share of the working population commonly allocated to cognitive-task occupations has ceased to grow at the turn of the century, while the share of graduates was still increasing. The result was an increased probability of occupational downgrading, with various adverse consequences for the less qualified workers. After 2000, the wage curves of the 4-year College graduates have "flattened" and the starting wages went down, and these facts cannot simply be explained by the business cycle. The situation of France is similar.

Studying composition effects, Carneiro and Lee (2011) show that enrollment growth is likely to have caused a decrease in the quality of student selection, explaining a drop of 6% in the College skill-premium between 1960 and 2000, in the United States.[10] Belzil and Hansen (2020) reach similar

[9] According to Autor, Goldin, and Katz (2020): "The largest part of increased wage variance in the twenty-first century comes from rising inequality among college graduates...".

[10] Their result relies on an identification assumption: there are College-enrollment differences in the individuals' regions of birth that can be exploited to disentangle the effect of quantity from that of quality. See also Carneiro, Heckman, and Vytlacil (2011), who estimate the *marginal treatment effect* of College.

conclusions, comparing the 1979 and 1997 cohorts of the NLSY survey, using structural econometric methods. Ashworth, Hotz, Maurel, and Ransom (2021) use the same NLSY data, and study closely related questions with the help of a structural model with a latent factor structure.

Returns to experience have recently been the object of renewed interest: see Dustmann and Meghir (2005), Kambourov and Manovskii (2009) and Jeong, Kim, and Manovskii (2015). On the dynamics of wages over the life-cycle, see *e.g.*, Huggett, Ventura, and Yaron (2011), Magnac, Pistolesi, and Roux (2018), Guvenen, Karahan, Ozkan, and Song (2021). Our analysis shows that the age-earnings profiles of individuals with different latent types (and with different levels of human capital) also have different slopes as in — for instance — Guvenen (2007).

For a general treatment of finite mixture models, see McLachlan and Peel (2000), Bouveyron, Celeux, Murphy, and Raftery (2019). The estimation methods used here have been employed in various contributions. Discrete or discretized latent structures are not a novelty in economics, and go back (at least) to Heckman and Singer (1984). The sequential EM algorithm that we use to obtain preliminary estimates has been proposed by Arcidiacono and Jones (2003) and applied by several researchers[11]

2 Context and Data

In this section, we first briefly describe the French education system and present the data.

2.1 The French context

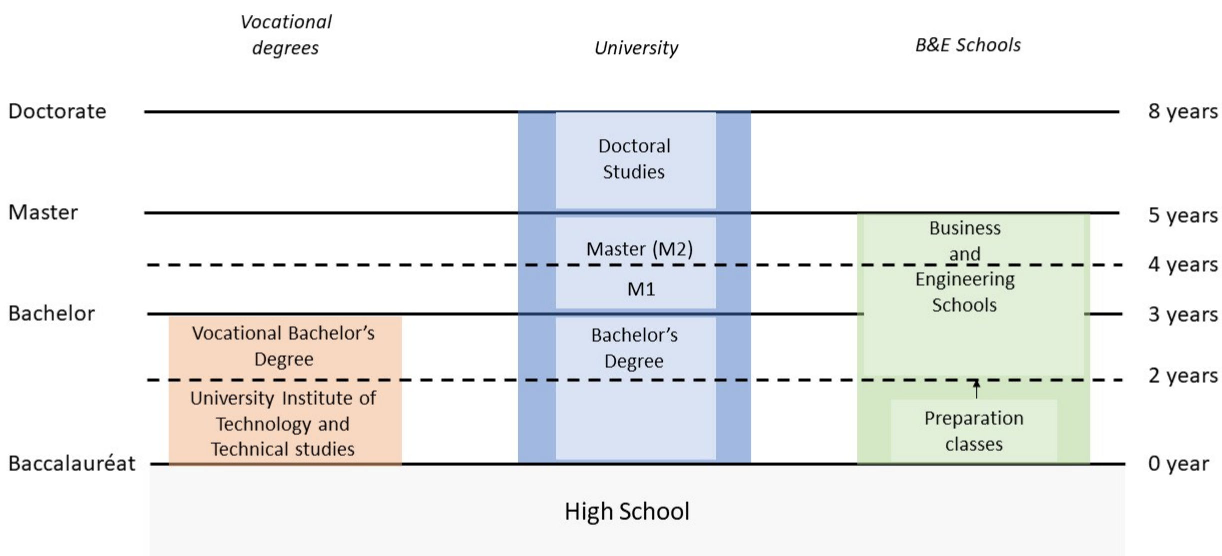
In France, (as in many other countries) the share of higher-education graduates has constantly grown in the past decades. In 2012, the share of higher-education graduates, including the French equivalent of the associate's and two-year vocational degrees, reaches 42% in the 25-29 age bracket, while this share is only 12.5% in the 60-64 age group. The number of higher-education students has reached 2.73 million in 2019-2020.[12] Between 1990 and 2015, the overall rate of growth of enrollment in higher-education institutions reaches 37%. The growth in enrollment has the potential to flood the labor market with graduates and there are concerns that an excess supply of Masters would cause a drop in their wages.

[11]See, in particular, Beffy, Fougère, and Maurel (2012), Arcidiacono, Aucejo, Maurel, and Ransom (2016). In addition, Gary-Bobo, Goussé, and Robin (2016), Cassagneau-Francis, Gary-Bobo, Pernaudet, and Robin (2021), and Cassagneau-Francis (2021) present other applications of finite mixtures. Finally, see the recent manuscript of Corblet (2022), exploring the same data sets, but with other methods.

[12]Figures published online by the French Ministry of Higher Education and Research, *i.e., Ministère de l'enseignement supérieur de la recherche et de l'innovation.*

In 2002, French universities have implemented the BMD reform (*i.e.*, the *Bachelor-master-doctorate* reform), *i.e.*, a set of measures adapting the French higher education system to European standards. The reform has set up an architecture based on three academic levels: Bachelor (*i.e., Licence*), Master, PhD (*i.e., Doctorat*).[13] The system is described by Fig. 1.

Figure 1: **The French Higher-Education System**



After high-school graduation (*i.e., baccalauréat*), typically at 18, students may go to work or continue studying. This depends to a large extent on the type of baccalauréat, that can be vocational or general. There is a group of vocational degrees requiring two or three years of education that can be compared to Associate's degrees in America. Undergraduate studies in universities lead to a Bachelor's degree after three years of College. The students who continue after three years in universities typically enter a two-year Master program. We distinguish the first year (called M1, standing for Master 1), 4 years after high-school graduation, from the second year (called M2, standing for Master 2) and requiring 5 years of study. The reason for this distinction is selective admissions. Until recently, the French public universities were not allowed to select students at the entry of M1 years. The tradition was that selective admission was permitted only at the entry of the second, M2 year (and this was the rule applied during our observation period). In the period covered by our data, the M1 is still not selective in principle. Yet, some universities used capacity constraints to limit admissions. But in this system, M2 Master graduates are obviously special. Finally, there also exists engineering schools and business schools that are typically independent institutions and have nothing to do with universities. Some are public, some are private (mainly nonprofit) institutions. The best such schools deliver a degree after five years of study, but the

[13] Higher education in France is now structured by European standards. Before this reform, the French system was not very different, and it is easy to find a correspondence between the pre-reform and post-reform degrees. The division of institutions in three categories: universities, vocational institutes, and schools has survived.

first two years are devoted to preparation classes. Admission is typically selective, sometimes very selective, in all French higher-education schools. They admit students after a competitive entry exam. The schools' degrees are equivalent to Master's M2 degrees but the selection of students is of course much more rigorous in schools than in universities, at least in principle.[14] This is the reason why we single out the business and engineering school categories, including only students with at least 5 years of study after high-school graduation in this category.

Aggregation of degrees In the following, we aggregate the highest degrees of individuals in 5 categories: 1°) *Below High-School Degree*, including dropouts without any certificate and secondary vocational certificates[15]; 2°) *High-School Degrees*, including all students for whom the *baccalauréat* is the highest degree. Many of these individuals in fact earned a vocational certificate, the *baccalauréat professionnel*, and in contrast, many of the classical baccalaureates have been enrolled in various higher-education institutions and therefore eventually earned a more advanced degree; 3°) *Some College and Bachelors* includes all the students whose highest achievement is the equivalent of an Associate degree[16], plus all the bachelors, *i.e.*, the French *Licence* and the M1, *i.e.*, the first year of master programs; 4°) Master degrees, typically the degree of a two-year graduate program requiring 5 years of study (M2); 5°) The degrees of all business and engineering schools, also requiring 5 years.

2.2 Data; CEREQ Generation Surveys

For the estimation work presented here, we stacked three large samples of young male workers. The samples, called Generation Surveys (*i.e.*, *Enquêtes Génération*) are produced by a French public institution called CEREQ.[17] Since 1992, every 5 years, the CEREQ draws a large representative sample of individuals who all left the educational system during the same year, with a large variety of educational achievement levels.[18] We will consider the CEREQ Generation surveys of 1998, 2004 and 2010 only.[19] The workers are followed during 7 years; the labor market experience of each worker is tracked, month by month, by means of interviews; questions are asked after 3 years, 5 years and 7 years. The data takes the following form : a listing of individuals with, for each, a long list of possible control variables (including family-background characteristics) and a fine description of

[14]Some engineering or business schools — not the best — admit students directly after high-school graduation. In addition, business schools recently developed 3-year bachelor programs that are less selective, enrolling students after high-school.

[15]In particular, the CAP, *i.e., certificat d'aptitude professionnelle*, the BEP, *i.e., brevet d'études professionnelles.*

[16]The degrees of the IUT, *i.e., Instituts Universitaires de Technologie*, called DUT, of the STS, *i.e., Sections de techniciens supérieurs*, called BTS, and other vocational degrees requiring less than 3 years of study.

[17]The CEREQ (*i.e., Centre d'études et de recherches sur les qualifications*, see https://www.cereq.fr.

[18]For instance, interviews, after 7 years, yield a sample of around 16,000 male and female individuals in Generation 1998.

[19]The first survey, launched in 1992, has a slightly different structure.

degrees and certificates. For each individual, we also have a list of employment and unemployment spells, giving the monthly wage at the beginning and the end of each employment spell, and giving the rate of employment (*i.e.*, full time, part time, etc., expressed as a percentage of full time work, between 0 and 1). In addition, we observe the wages at the moment of the interviews, after 3, 5 and 7 years. Wages are given as monthly nominal salaries, including bonuses, net of compulsory social security and medical-insurance contributions, but gross of the income tax.[20] Individuals typically leave the education system on a given month during the base year, but not all on the same month. Thus, there is some variation of the beginning month, so that young workers accumulate different amounts of potential experience during the period covered by the survey. Individuals also take a variable number of months to find a first job.[21] To sum up, we observe employment variables for a sample of individuals every month from 1998 until 2005, from 2003 to 2011 and from 2009 to 2017, but we observe wages only at some dates: at the endpoints of employment spells and at the moment of interviews.

Stacking the three *Generation* surveys of 1998, 2004 and 2010, we obtain a standard, but unbalanced panel. The panel is unbalanced for two reasons. Firstly, we do not observe the individuals of a given survey in all periods from January 1998 to December 2017. Secondly, we do not observe the wages in the middle of employment spells. For details on sample construction, see Online Appendix K. Descriptive statistics are presented in Appendix A.

2.3 Do we Observe a Devaluation of Degrees?

We defined the devaluation of a degree above as an absolute decrease in the expected real salary of workers conditional on holding the degree.[22] Relative devaluations refer to drops in the College skill premia or more generally to a decrease in the ratio of average wages conditional on two different degree categories. The main questions that we ask are simple: do we observe a devaluation of degrees over this period of 20 years, and if a devaluation did indeed occur, what are its likely causes? In particular, can it be attributed to changes in the selection of students?

For estimation, we limited ourselves to full-time wages. With the restriction to full-time wage observations, we clearly maximize the chances of selecting individuals in relatively good health, with relatively good jobs and good pay.[23] If, given this kind of selection, we observe a devaluation,

[20] This is not exactly the usual take-home pay, but note that before January 2019, the French income tax was not withheld from wages. The definition used here was therefore, for most individuals, the most easily observable and most salient expression of their income.

[21] In practice, the realization of interviews after 3, 5 or 7-years last 3 or 4 months (from October to December).

[22] Our definition of devaluation is simple and empirical. We can estimate the average real wage of a student holding a certificate of some given category after 7 years of career, in 2005, 2011 and 2017 and compare these averages at several points in time. If the average real wage decreased, we say that this particular category of degrees is devalued.

[23] The advantage of doing this is to avoid possible errors in the reporting of part-time work and part-time wages,

it is therefore all the more significant.

We started with a preliminary analysis of the data using standard econometric methods, including the usual panel-data *within* estimators. Our preliminary study shows a devaluation of higher-education degrees, and more specifically Master's degrees, of the order of 10%, between 1998 and 2017. Part of the devaluation is in fact due to a drop in returns to experience. We also studied the effect of the business cycle on wages, in a simple way. The Online Appendix D gives a presentation of these preliminary results.[24]

To summarize the preliminary analysis, we believe that the devaluation of higher-education degrees is most likely due to an excess supply of graduates. Yet, we know that there are competing explanations. The value of the degrees under scrutiny depends on (at least) two other factors: the selection of student skills and the quality of education. Both factors contribute to the graduates' human capital, and therefore to productivity and wages. It is a common contention that the quality of students enrolled in advanced programs has gone down in the recent years (this is heard in France and elsewhere). The quality of the teaching may also have decreased with time, and the two phenomena go hand in hand. At this point, with the help of standard econometric methods, it is almost impossible to decide if the selection of talents enrolled in higher education has changed in the past twenty years. To push the investigation further, we therefore propose a model of unobserved heterogeneity.

3 The Model

To model the beginning of careers of three cohorts of young men under unobservable heterogeneity, we assume that the distributions that we see are generated by a finite mixture of distributions, each point in the mixture being a latent, unobservable type of individual.

Let c denote the cohort of the individual with $c \in \{1998, 2004, 2010\}$. In each cohort, we follow individuals across time from the moment they leave school to the moment of the survey seven years later.

We denote by t the elapsed time period (in months) since the first individuals of the first cohort left school. Note that, at $t = 1$, individuals do not have the same age, as some individuals just graduated from high school and enter the labor market while others just graduated from university. Individuals are indexed by i, with $i = 1, \dots, N$. Let h index the highest level of education reached

given that we do not observe the exact number of hours, and that most of full-time jobs correspond to 140 hours per month.

[24] Our preliminary analysis, in essence, is also exposed in Argan and Gary-Bobo (2021) — but the latter article is written in French.

by the individual with $h = 1, \ldots, H$. Let $\chi_h(i)$ be the dummy that indicates whether individual i has reached education level h. Let X_{it} denote a vector of observed characteristics of i at time t. We decompose X_{it} in two subsets of variables, Z_i, the set of time-invariant variables and Ξ_{it}, the set of time-varying characteristics, so that $X_{it} = (Z_i, \Xi_{it})$.

Let W_{it} denote the observed real salary of i at time t. Let $w_{it} = \ln(W_{it})$. To obtain real wages, we deflated nominal wages, using the French consumer price index.[25] We also observe the employment rate of individual i at date t, denoted e_{it}. The latter variable takes on a finite number of values only, $e_{it} \in \{0, .3, .5, .6, .8, 1\}$; $e = 1$ represents full-time employment, and numbers between 0 and 1 measure the hours of part-time jobs as a fraction of a standard full-time job. Using the convention that $e_{it} = 0$ for all periods t such that i has not yet left the educational system, we therefore also measure effective experience, denoted x_{it}, as the cumulative hours of work, that is, for $t > 1$,

$$x_{it} = \sum_{\tau=1}^{t-1} e_{i\tau}, \tag{1}$$

where $x_{i1} = 0$.

We assume that individuals belong to one of a finite number of unobserved groups, called *types*. Let K be the number of latent types and let k index types. We denote $\theta_k(i)$ the dummy that indicates whether individual i is of type k.

3.1 Wage equation

Note that individual i's wage is not observed each month (for each t). The wage is observed at the onset and at the end of employment spells, and at the moment of the survey. Let T_i be the subset of dates t at which we observe a wage for individual i.

We can now specify the wage equation. For $t \in T_i$ and for an individual i of type k, we set

$$w_{itk} = \alpha_{0k} + \sum_{c=1}^{C} \chi_c(i) \left(\delta_{0ck} + \sum_{h=1}^{H} \chi_h(i) \left(\gamma_{0chk} + \beta_{0chk} x_{it} \right) \right) + X_{it}\eta_{0k} + \epsilon_{itk}, \tag{2}$$

where ϵ_{itk} is a normal error term with a zero mean and variance σ^2_{wk}, where $\chi_c(i)$ is a dummy indicating if the individual is in cohort c and $(\alpha_{0k}, \beta_{0chk}, \gamma_{0chk}, \delta_{0ck}, \eta_{0k})_{h=1,.,H,k=1,.,K,c=1,.,C}$ is a vector of parameters. In addition, note that x_{it} is effective experience as defined by equation 1. Given this, the expression for the observed wage of individual i at period t is,

$$w_{it} = \sum_{k=1}^{K} \theta_k(i) \left[\alpha_{0k} + \sum_{c=1}^{C} \chi_c(i) \left(\delta_{0ck} + \sum_{h=1}^{H} \chi_h(i) \left(\gamma_{0chk} + \beta_{0chk} x_{it} \right) \right) + X_{it}\eta_{0k} \right] + \epsilon_{it}, \tag{3}$$

[25] We used the CPI published by the National Statistical Institute, *i.e.*, INSEE. All wages ar expressed in 2013 euros.

where $\epsilon_{it} = \sum_k \epsilon_{itk}\theta_k(i)$.

Note that the model is very flexible insofar as all the parameters of the wage equation are free to vary with type k; returns to education and experience vary with the cohort c and returns to experience may also depend on educational attainment h.

3.2 Employment equation

We model the employment level e_{it} at each date by means of an Ordered Probit model. Recall that e_{it} takes on discrete values between 0 and 1 that measure individual i's rate of employment in period t. Let G be the number of levels of employment, denoted $\mathbf{e}_g$, with $g = 1, .., G$ and $1 \geq \mathbf{e}_{g+1} > \mathbf{e}_g \geq 0$. We define,

$$P_k(e_{it}|X_i, x_{it}, h_i) = \Pr(e_{it} = \mathbf{e}_g \mid X_{it}, x_{it}, h_i, k) = \Pr\left[\mathbf{c}_{gk} \leq \rho_{itk} + \zeta_{itk} \leq \mathbf{c}_{g+1,k}\right], \tag{4}$$

where

$$\rho_{itk} = \sum_{c=1}^{C} \chi_c(i) \left(\delta_{1ck} + \beta_{1ck} x_{it} + \sum_{h=1}^{H} \gamma_{1chk} \chi_h(i) \right) + X_{it}\eta_{1k}, \tag{5}$$

where the $\mathbf{c}_{gk}$ are the thresholds (*i.e.*, cuts) of the Ordered Probit, and $\mathbf{c}_{0k} = -\infty$. The ζ_{itk} are independent random variables with a standard normal distribution and $(\beta_{1ck}, \gamma_{1chk}, \delta_{1ck}, \eta_{1k})_{h=1,.,H,k=1,.,K,c=1,2,3}$ is a vector of parameters to estimate. Remark that all parameters are free to vary with k.

We denote E_i the subset of dates t at which e_{it} is observed. This model is estimated mainly thanks to observations e_{it} at the beginning and the end of each employment spell of individual i. In addition, there are some observations in the middle of a spell. Typically, this happens when, at the end of the survey period, an individual is currently employed, and these observations correspond to truncated spells. Typically, at a date t corresponding to the beginning of an employment spell, the employment rate e_{it} jumps to 1, or a positive value smaller than 1 in the case of a part-time job. At a date t corresponding to the last period of a full-employment spell, we observe $e_{i,t+1} = 0$ if i becomes unemployed, or $0 < e_{i,t+1} \leq 1$ if i changes for a part-time job.

This model is clearly a kind of reduced form, but it is rich and flexible enough to capture the possibility that probabilities of finding a job at any t depend on accumulated experience, degrees, the cohort, and the type.

3.3 Education equation

Finally, we model the level of education with the help of a multinomial logit model. This approach provides a simple way of modelling individual investment in education. We denote Λ the probability of choosing education h, that is,

$$\Lambda_k(h|Z_i) = \Pr(h_i = h \mid Z_i, k) = \Pr\left[\mathbf{u}_{ihk} = \max_{j\in\{1,\ldots,H\}} (\mathbf{u}_{ijk})\right], \tag{6}$$

where the utility u_{ihk} of an individual i of type k choosing education level h is defined as $\mathbf{u}_{ihk} = \mathbf{v}_{ihk} + \xi_{ihk}$, with

$$\mathbf{v}_{ihk} = \alpha_{2hk} + \sum_{c=1}^{C} \delta_{2chk}\chi_c(i) + Z_i\eta_{2hk}, \tag{7}$$

where ξ_{ihk} is a random variable that follows a Gumbel distribution (*i.e.*, Type 1 extreme-value distribution) and where we want to estimate the following vector of parameters: $(\alpha_{2hk}, \delta_{2chk}, \eta_{2hk})$ where $h = 1, \ldots, H$, $k = 1, \ldots, K$, and $c = 1, \ldots, C$.

Clearly, this model is again a reduced form. The description of education choices is static. In addition, the model has a "triangular" structure because degrees explain experience and degrees and experience explain wages. In other words, wages do not appear in the choice equations. But as discussed in the literature, the ex ante wage expectations of individuals should in principle appear in the choice equations, instead of the ex post, effectively observed wages of each individual.[26] This would require a model of wage expectations depending on the latent types — a possible extension of our approach. Since education will depend on the latent groups, we can say that types capture differences in expectations in a rudimentary way. The multinomial choice equation may be viewed as an auxiliary part of the model, yet, it permits us to estimate choice probabilities that depend on the latent types.

3.4 Identification

We estimate the model by maximization of the log likelihood. We typically use the sequential EM algorithm to obtain preliminary estimates, and then use a standard ML algorithm. The model's likelihood is derived in Appendix B.

The maximum likelihood method provides us with estimated values and standard deviations for all parameters, $(\alpha, \beta, \gamma, \delta, \eta, \sigma, \mathbf{c})$ and the *prior* probabilities of types p_k. We present here the results obtained when we fix $K = 3$. We discuss the choice of the number of types, using information and entropy criteria, in Online Appendix J. An important output of the estimation algorithm is the *posterior* probability that individual i is of type k, that is,

$$p_{ik} = \Pr(k|X_i, y_i). \tag{8}$$

[26]On this theme, see Beffy, Fougère, and Maurel (2012) and Arcidiacono, Hotz, Maurel, and Romano (2020). On *ex ante* returns to schooling, on the separation of what a student can forecast at the time of educational decisions, based on private information, from the risk in future wages, *i.e.*, the separation of risk from heterogeneity in the observed distribution of wages, there is an important literature; see Cunha and Heckman (2007), Carneiro, Hansen, and Heckman (2003), Cunha, Heckman, and Navarro (2005).

The probability p_{ik}, can be expressed with the help of Bayes' rule and the likelihood, as indicated in Appendix B.

Identification and Nonparametric Identification. Our main identifying assumption is that wage observations (and employment rates) are independent conditional on accumulated experience, observable characteristics (degrees) and latent types. Parametric identification of the wage equation is obtained under standard conditions (see McLachlan and Peel (2000)). The ordered probit and the multinomial logit would be parametrically identified in the case of a single type. In addition, a static discrete choice model, if estimated separately, does not permit the identification of latent choices. We will come back to this point below.

We first discuss the identification of a wage equation with a latent structure. The discussion on the possibility of *nonparametric identification* can be based on the results of Allman, Matias, and Rhodes (2009). In a nutshell, our wage equation alone would allow us to identify a latent type structure and its parameters nonparametrically, up to a relabeling of types, *i.e.*, we would obtain, for given K, the probabilities of types p_k and the conditional c.d.fs $G_t(w|k)$ of wages w at time t. So, in principle, we could get rid of the normality assumption and still estimate the wage model with a set of latent types and their associated probabilities. More precisely, to achieve full nonparametric identification, according to the theorems of Allman *et al.* (2009), we need three groups of variables that are independent conditional on the latent types, plus a condition that the conditional distributions $G(.|k)$, $k = 1, \dots, K$, are linearly independent.[27] The latter condition is reasonable if types are really different. So, the main problem is to find three conditionally independent random measures of types: we now show that the three measures are at hand.

We can apply the general theorems if we also condition with respect to observable characteristics. The employment rate profile of any individual, and therefore this individual's profile of accumulated experience, can be described by a finite number of states or cells, since employment rates e_{it} are discretized. Other observed characteristics such as the educational achievement h and the family-background variables are typically dummy variables (if a control is continuous, it can be discretized). It follows that we can bin the entire population in a finite number of cells. Given our assumption on wages (and the wage equation above), *in each of these cells*, and *conditional on the latent type*, wage observations made at different dates t are independent. In our panel, at least three different values of w are available for each i. Now, let K be the number of types. For each k, we identify in each cell X a probability $p(k|X)$ and an array of distributions $G_t(w|k, X)$. Given that we know the distribution of observable variables $\phi(X)$, we easily derive $p_k = \sum_X \phi(X)p(k|X)$, etc. It follows that a latent type structure can be nonparametrically identified from the distribution of wages.

[27]See particularly Theorem 8 in Allman *et al.* (2009). On this topic, further results are proved and estimation methods are provided in Bonhomme, Jochmans, and Robin (2016).

A more difficult problem is to nonparametrically identify a finite latent structure for the joint distribution of wages, employment rates and educational choices. The theorems of Allman *et al.* (2009) cannot be applied because education determines employment and wages, and because employment (in fact, experience) determines wages: the three variables cannot be assumed independent conditional on the latent types.

The literature on the identification of dynamic discrete choice models[28] provides us with some tools that can be applied to the study of our model. Our Ordered Probit model, used to predict the employment rate at each t, which is a specific discrete choice model, is nonparametrically identified using the results of Kasahara and Shimotsu (2009). In the latter paper, the key features permitting nonparametric identification of a finite mixture are: *(i)* the observation of individual choices during a sufficiently large number of periods (*i.e.*, the length of the panel), *(ii)*, the number of different values that time-varying control variables can take; and *(iii)*, the fact that latent types react differently to changes in the control variables. Our panel is sufficiently long; the accumulated experience varies with time in many possible ways; it is reasonable to assume that each type reacts differently to changes in effective experience: nonparametric identification is at hand.

Finally, the education choice model is static and it follows that a finite mixture of multinomial choice models cannot be identified in isolation. Yet, if we fix the number of types and know their probabilities, we can obtain the choice model for each type simply by means of a weighted likelihood-maximization algorithm, as in the M-step of an EM algorithm, in spite of the fact that the model is static. The finite mixture of multinomial choice models can therefore be identified jointly with the wage equation, since the latter provides the type probabilities that are needed to estimate its parameters. In other words, the wage model provides an auxiliary equation for the finite mixture of Multinomial Logits. To conlude this discussion, it is possible to obtain a nonparametric identification result for the complete model, but it is a nontrivial problem to prove such a result rigorously, and this problem is beyond the ambitions of the present article.

4 Policy-relevant parameters; ATEs and ATTs

We will use the model, the estimated values of parameters and the posterior probabilities of types for each individual, denoted p_{ik}, to compute policy-relevant parameters. In particular, we can study ATTs, ATEs and the effect of unobserved selection on various outcomes. It is also possible to study the heterogeneity of treatment effects and we can compute ATEs and ATTs conditional on type k.

[28] See also Magnac and Thesmar (2002), Hall and Zhou (2003).

4.1 Policy-relevant parameters; ATEs and ATTs: Method

Let $y_t(z)$ denote the potential value of any outcome, at time t, for individuals with observable characteristics z.[29] We first define an *average treatment effect conditional on type* k and education level h at time t, denoted $ATE(h,k,t)$. Let $h=0$ denote the level of individuals without any degree (high-school dropouts): these individuals are our reference point. This conditional treatment effect is defined as follows,

$$ATE(h,k,t) = \mathbb{E}[y_t(h)|k] - \mathbb{E}[y_t(0)|k]. \tag{9}$$

The (unconditional) average treatment effect at time t, for individuals with level h is then defined as follows,

$$ATE(h,t) = \sum_k p_k ATE(h,k,t), \tag{10}$$

where the p_k are the prior probabilities of types defined above.

For any vector of observable characteristics z, let $\chi_z(i)=1$ if and only if $z_i = z$ and $\chi_z(i) = 0$ otherwise. We use the observations y_{it} of the outcome for individuals i. To estimate $\mathbb{E}[y_t(h)|k]$ we use the statistic,

$$\hat{\mathbb{E}}[y_t(h)|k] = \frac{\sum_i y_{it}\hat{p}_{ik}\chi_h(i)}{\sum_i \hat{p}_{ik}\chi_h(i)} = \frac{\sum\limits_{\{i|h_i=h\}} y_{it}\hat{p}_{ik}}{\sum\limits_{\{i|h_i=h\}} \hat{p}_{ik}}, \tag{11}$$

where $\hat{p}_{ik}$ is the estimated posterior probability that i belongs to group k, computed by Bayes's law as indicated above by expression (27). Basically (11) is an estimation of $\mathbb{E}(y|h,k)$, using the sample. In a similar fashion, we define,

$$\widehat{ATE}(h,t) = \sum_k \hat{p}_k \widehat{ATE}(h,k,t), \tag{12}$$

where $\hat{p}_k$ is the estimated prior probability, and where,

$$\widehat{ATE}(h,k,t) = \hat{\mathbb{E}}[y_t(h)|k] - \hat{\mathbb{E}}[y_t(0)|k]. \tag{13}$$

Now, to compute the ATT (average effect of treatment on the treated), we have,

$$\mathbb{E}[y_t(h)|h',k] = \mathbb{E}[y_t(h)|k] \quad \text{for all } (h',h). \tag{14}$$

This is equivalent to the familiar *conditional independence assumption* of the treatment-effects literature, except that conditioning is with respect to the unobservable type k, and education h is the treatment here. In other words, the expected counterfactual (or potential) outcome of a type k

[29] For instance, the potential average wage of an individual with degree h and t months of potential experience, *i.e.*, $w_t(h)$, is an outcome of interest, as well as the average wage of an individual with characteristic z in cohort c, that we can also denote $w_c(z)$ (with a slight abuse of notation). In a similar fashion, we define the employment rate $e_t(z)$; the accumulated level of effective experience $x_t(z)$, etc.

with degree h', if instead of h' they had chosen a degree h, is just the mean outcome of individuals with degree h, knowing type k. Under this assumption, we have

$$ATT(h,k,t) = ATE(h,k,t), \tag{15}$$

and it is easy to show that,

$$ATT(h,t) = \sum_k p(k|h) ATE(h,k,t), \quad \text{where} \quad p(k|h) = \frac{p(h,k)}{p(h)}. \tag{16}$$

Now, obviously, to estimate ATT, we use $\widehat{ATE}(h,k,t)$ and the estimated conditional probability $\hat{p}(k|h)$ which is itself the ratio of[30]

$$\hat{p}(h,k) = \frac{1}{N} \sum_i \hat{p}_{ik} \chi_h(i) \tag{17}$$

and

$$\hat{p}(h) = \sum_k \hat{p}(h,k) = \frac{1}{N} \sum_i \chi_h(i). \tag{18}$$

Finally, $\widehat{ATT}$ is just obtained by putting hats on p and ATE in equation 16.

With the help of posterior probabilities, we can estimate the probability of choosing h, knowing unobservable type k and observable characteristic z as follows,

$$\hat{p}(h|k,z) = \frac{\hat{p}(h,k,z)}{\hat{p}(k,z)} = \frac{\sum_i \hat{p}_{ik} \chi_{hz}(i)}{\sum_i \hat{p}_{ik} \chi_z(i)}, \tag{19}$$

where $\chi_{hz}(i) = 1$ iff $z_i = z$ and $h_i = h$ and $\chi_{hz}(i) = 0$ otherwise.

In the case of wages, we do not observe w_{it} at every t and for every i but we can take averages over several periods, if needed, and the definitions would be changed accordingly, in an obvious manner. For instance, we will consider averages over all the periods t of a given cohort c.

4.2 Estimation of the discounted sum of earnings. Simulations

We can estimate the "human capital", *i.e.*, the discounted sum of earnings of an individual i with observable characteristics Z_i and type k, using the estimated model. This outcome is interesting to compare types, because it summarizes all the differences in wages (returns to education and experience), employment rates and educational achievement. To this end, we simulated a fictitious sequence of employment, experience and wages for each individual i in the sample. Then, using weights p_{ik}, we averaged the expected-discounted fictitious sequence of earnings of each i. We provide details on the method of simulation in Appendix C.

[30] It is easy to check that $\sum_h \sum_k \hat{p}(h,k) = 1$.

5 Results

We can now present our estimation results. We start with the distribution of types. Next we compute the ATTs and ATEs of education and the simulated discounted earnings, type by type, and overall. Then, we present the ML estimates of the model's other parameters and we discuss returns to education, returns to experience and educational choices, conditional on unobserved types. This leads us to propose an interpretation of the three types that we find: the types are clearly different, with a clear hierarchy.

To estimate the model, in addition to the variables discussed in Section 3, we use the following list of variables X_t: the student's location in geographical space, indicated by dummies (Urban, Peri-Urban and Rural), the father's occupation (the father-is-a-professional dummy); and the macroeconomic unemployment rate. Further explanations about controls are given below.

5.1 Probability of types $k = 1, 2, 3$

Table 1 presents the estimated probabilities of types when $K = 3$: in our sample 42% of young men are of type 1, 36% of type 2 and 22% of type 3. The type frequencies are very precisely estimated. Before we provide an interpretation of these types —in other words, who do these types represent? — it is important to check if these types generate a good classification (*i.e.*, a near partition) of the population. The quality of classification is said to be good if each individual i belongs to a given group k with a sufficiently high probability, say, ideally, with $p_{ik} \simeq 1$ for some k. It may happen that a minority of individuals remains hard to categorize, and for these, we would find $p_{ik} \simeq 1/K$, or alternatively, they sometimes belong to a subset of $K' < K$ types with a high probability. Visual inspection of the histograms of the estimated values p_{ik} for each k will immediately show if the classification is fuzzy. Figure 2 shows that the classification is in fact very good. Most individuals have values of p_{ik} close to 0 or 1.

Table 1: Estimated probability of types

Type	1	2	3
Probability	0.42	0.36	0.22
Standard error	(.006)	(.006)	-

Table 2: Distribution of types by cohort

Type	1	2	3
1998	0.40	0.36	0.23
2004	0.42	0.37	0.21
2010	0.45	0.33	0.22

Given our results, it seems that the types are not simply fictitious disembodied categories used to fit the distribution of employment and wages: they are likely to correspond to real people. It remains to understand which observable characteristics help recognizing a given type.

As explained above, our ML estimation results permit us to examine the distribution of types conditional on any characteristic or set of characteristics. It is sufficient to compute the arithmetic average of posterior probabilities p_{ik} in the subset of individuals i sharing the given characteristic(s). The distribution of types by cohort is presented in Table 2. It appears that this distribution is very stable across cohorts. None of the types is negligible — all the prior probabilities (or frequencies) are above 20%. In the coming subsections, we will see how individual wages and employment rates depend on types. The multinomial, discrete-choice part of our model gives a more detailed account of individual education decisions.

5.2 ATEs, ATTs and Discounted Earnings: Results

We now present the values of ATEs and ATTs, as well as the simulated values of discounted earnings by type.

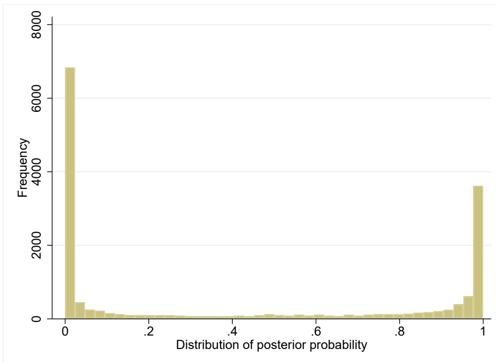
5.2.1 ATEs and ATTs. Changes in the Selection of Students

The ATE and ATT parameters are defined above, in subsection 5.1., by Equations 12, 13 and 16, 17, 18. Figure 3 gives: a) the ATE; b) the ATT; and c) the percentage variation (ATT − ATE)/ATE for full-time wages. The left-hand part of each panel of Fig. 3 displays the results for wages observed during the first year while the right-hand part of each panel displays the results obtained with wages observed during the seventh year of career. The reference education $h = 0$ is the category of 'less than high-school degree' (including the dropouts).

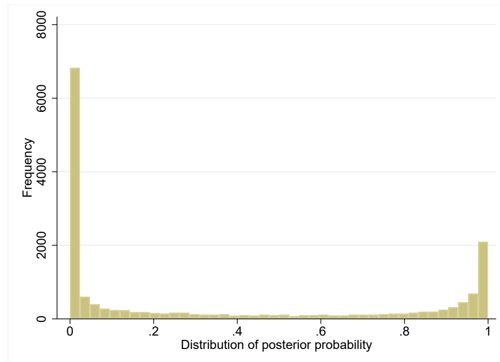
The first striking result is that ATE is consistently larger than ATT for the high-school graduates, *i.e.*, the students for which the high-school degree is the highest degree. This confirms our intuition that this education level does not select the most productive students.

Figure 2: Empirical distribution of posterior type probabilities

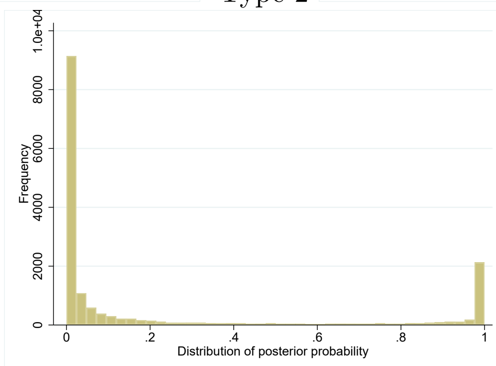
Type 1



Type 2



Type 3

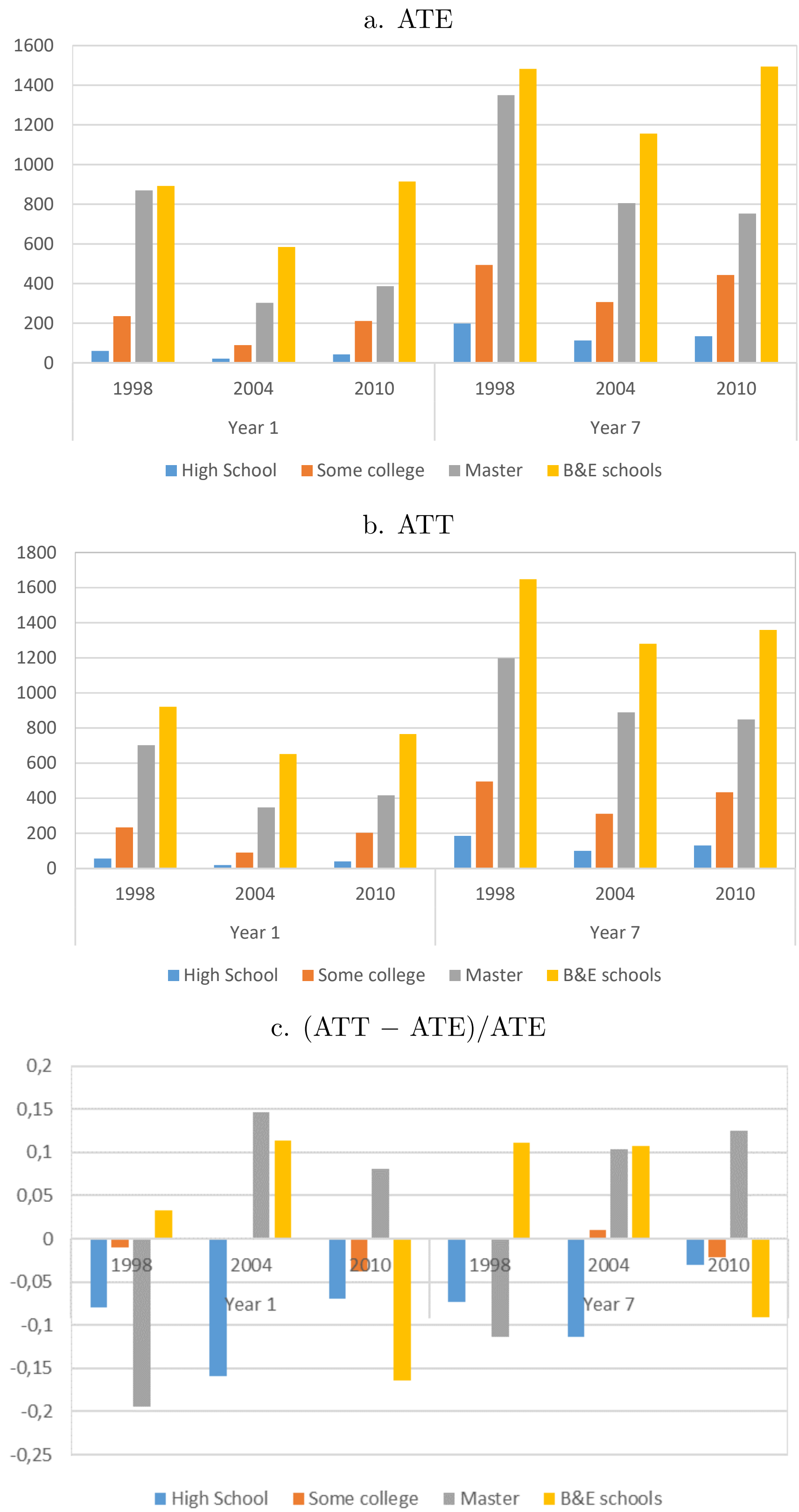


The second striking result that is visible is the drop in the ATE of Master degrees: between 2017 and 1998, the absolute variation ΔATE after 7 years of career is around -600 euros per month. The absolute variation ΔATT, during the same period and for the same degrees is around -400 euros (per month), a smaller drop. The bottom panel of Fig. 3 shows that in 1998, we had ATE>ATT for the Master program graduates. These graduates were therefore less well selected than the average population, but this difference reversed in the later years. Indeed, in the 2004 and 2010 cohorts, panel *c* of Fig. 3 clearly shows that the selection of Master graduates has improved. The reversal is particularly visible after 7 years. This result is surprising, because we expected a selection of lesser quality students in these programs, due to the sharp increase in enrollment. A consequence of this observation is that an 'excess supply' of graduates could be the main explanation for degree devaluation, because it is not the result of adverse selection.

Next, the situation is more complicated than it may seem at first glance, because the evolution of selection is exactly the opposite for business-school and engineering-school graduates. The difference ATT-ATE clearly decreased between the 1998 and 2010 cohorts. There is a lot of evidence about the constant growth of business schools in France. These schools have been growing since the 1970s and new schools opened. The growth of aggregate enrollment accelerated in the recent years, in spite of increasing tuition fees. The growth was possible only at the cost of less selectivity.[31] The

[31] Indeed, it is well-known that the French business schools developed teaching programs, like the "bachelors" that

Figure 3: ATE and ATT when education is the treatment and full-time wages are the outcome



interpretation of our results is therefore straightforward. In sharp contrast, University degrees, in spite of the growth of enrollment, have markedly improved selection at the master's level. In fact, if we put business and engineering schools and the doctorates aside, the master's degree has become the only really selective instrument of French universities.

5.2.2 Simulations of the Model. Discounted Earnings Conditional on Type k

We simulate sequences of employment rates and wage rates $(\tilde{e}_{itk}, \tilde{w}_{itk})$. This allows us to compute the discounted expected earnings during the periods $t \in \{1, \dots, T\}$. We choose a discount factor $\delta = .99$ (per month) and for every (i, k), we compute,

$$\tilde{W}_{ik} = \frac{(1-\delta)}{(1-\delta^T)} \sum_{t=1}^{T} \delta^{t-1} \tilde{e}_{itk} \exp(\tilde{w}_{itk}).$$

$\tilde{W}_{ik}$ is a weighted average and has the dimension of monthly earnings. Then, we compute the weighted arithmetic mean, using the estimated probabilities p_{ik}. For each type k, we compute,

$$H_k = \frac{\sum_{i=1}^{N} \tilde{W}_{ik} \hat{p}_{ik}}{\sum_{i=1}^{N} \hat{p}_{ik}}.$$

See Appendix C for a detailed description of simulations. The simulations are based on the full estimated model. The value of H_k can be computed in subsamples, conditional on c or h or both. The results are given by Table 3. The figures are rather low for Type 1. This is due, not only to smaller monthly wages, but also to a lot of unemployment. In addition, we compute these values conditional on the cohort c, denoted $H_k(c)$.

Table 3: Discounted earnings by type and cohort, *i.e.*, $H_k(c)$

Type	1	2	3
All cohorts	746	1215	1329
1998	740	1173	1091
2004	745	1252	1393
2010	758	1246	1334

We see a clear hierarchy of types. Type 1 did not experience a devaluation (a drop of $H_k(c)$ with c) but this devaluation hit the other two types, between the 2004 and 2010 cohorts. The discounted values H_k are a synthetic indicator summarizing the effects of returns to degrees, returns to experience and unemployment. We see that Type 3 has lost 4.2% between 2004 and 2010.

cannot be compared with the traditional programs of the French "grande école". Note that the individuals in the business-school category here, as already mentioned, have completed 5 years of study after high-school graduation, and can be compared with Master's graduates.

Table 4: Average discounted earnings by type, degree and cohort, *i.e.*, $H_k(c, h)$

Cohort	1998			2004			2010		
Type	1	2	3	1	2	3	1	2	3
Less than High School	652	1017	819	635	994	934	510	865	930
High-School Degree	711	1050	1094	723	1113	1145	664	1072	952
Some College and Bachelors	802	1239	1382	801	1273	1507	772	1288	1496
Masters (M2)	1412	2362	1121	1017	1628	1879	837	1357	1974
Bus. & Eng. Schools	1327	1924	1861	1151	1786	2215	1874	2411	1443

Table 5: Discounted earnings by degree and cohort, *i.e.*, $H(h, c)$

Cohort	Degree	Actual	Counterfactual	Difference	Percent Variation
1998	Less than High School	813	824	-11	-0.1%
	High-School Degree	916	924	-8	-0.8%
	Some College and Bachelors	1105	1097	$+9$	$+0.8\%$
	M2	1589	1689	-100	-6.2%
	Bus. & Eng. Schools	1782	1669	$+113$	$+6.3\%$
2004	Less than High School	806	831	-25	-3.1%
	High-School Degree	932	956	-24	-2.6%
	Some College and Bachelors	1146	1125	$+21$	$+1.8\%$
	M2	1530	1426	$+105$	$+6.8\%$
	Bus. & Eng. Schools	1769	1611	$+158$	$+8.9\%$
2010	Less than High School	691	720	-30	-4.3%
	High-School Degree	828	862	-34	-4.1%
	Some College and Bachelors	1121	1103	$+17$	$+1.5\%$
	M2	1414	1261	$+153$	$+10.8\%$
	Bus. & Eng. Schools	1862	1954	-92	-4.9%

Table 4 shows the discounted earnings by type, degree and cohort, *i.e.*, $H_k(c, h)$. It is striking to see that, in terms of discounted earnings, devaluation took place for the less-than-high-school and the high-school degrees of Type 1 and 2. This is due to worse employment conditions because the wage rates increased, mainly as a consequence of an increased minimum wage. In the 'Some College and Bachelors' category, the devaluation hits Type 1 only. The devaluation of Masters is confirmed for Types 1 and 2, but not for Type 3: we give more details on this result below, in subsection 6.3. The interpretation of results obtained for business and engineering schools is more delicate: macroeconomic conditions probably play a role in explaining the unstable performances of Type 3 (but Type 3 is characterized by relatively less stable jobs, as compared to Type 2, as we will see in subsection 6.3 below).

Do we see selectivity changes in the sub-populations of graduates? Table 5, in column 3 (*i.e.*, 'Actual') gives the average value of $H_k(h, c)$, weighted by the conditional probabilities $p(k|h, c)$, while column 4 (*i.e.*, 'Counterfactual') gives the average of $H_k(h, c)$ weighted by probabilities $p(k|c)$. The fifth column of Table 5 gives the difference *Actual* $-$ *Counterfactual*. This difference measures the extent to which individuals are positively or negatively selected at various educational levels. The less-than-high-school-degree and high-school-degree holders earn on average less than if this population had the distribution of types of the whole population. The figures in the Selection column are negative in the three cohorts, but the difference between Actual and Counterfactual is small. In contrast, the situation of M2 degree holders has changed with time. Selection was clearly negative in 1998 (theses graduates seem less able than the general population), but the selection becomes positive in the 2004 and 2010 cohorts. The number of students enrolled in master programs has increased, but in fact, these university degrees have selected students that seem better than the average in a certain sense: they tend to have a higher type.

Next, the discounted earnings of engineering and business-school graduates has followed a completely different path: it seems that the quality of the selection of schools has deteriorated with time. These results confirm the findings obtained above with ATE and ATT when observed wages are the outcome and education is the treatment.

5.3 Parameters estimates

We now consider in turn the ML estimation results of the three building blocks of our model: the wage equation, the employment equation; the education choice equation. The results again show a clear hierarchy of types.

5.3.1 Wage equation

Complete ML estimates of the wage equation are presented in Appendix D, in Tables 10, 11 and 12. Table 10 presents the wage returns to experience by cohort, type and education level (β_{0chk}). Table 11 presents the returns to education by cohort and type (γ_{0chk}). Table 12 presents the other parameters of the wage equation ($\alpha_{0k}, \delta_{0ck}, \eta_{0k}$). A glance at Figures 4 and 6 will show the main insights that can be drawn from the wage equation.

Figure 4: Monthly Returns to Experience by Type, Educational Attainment and Cohort

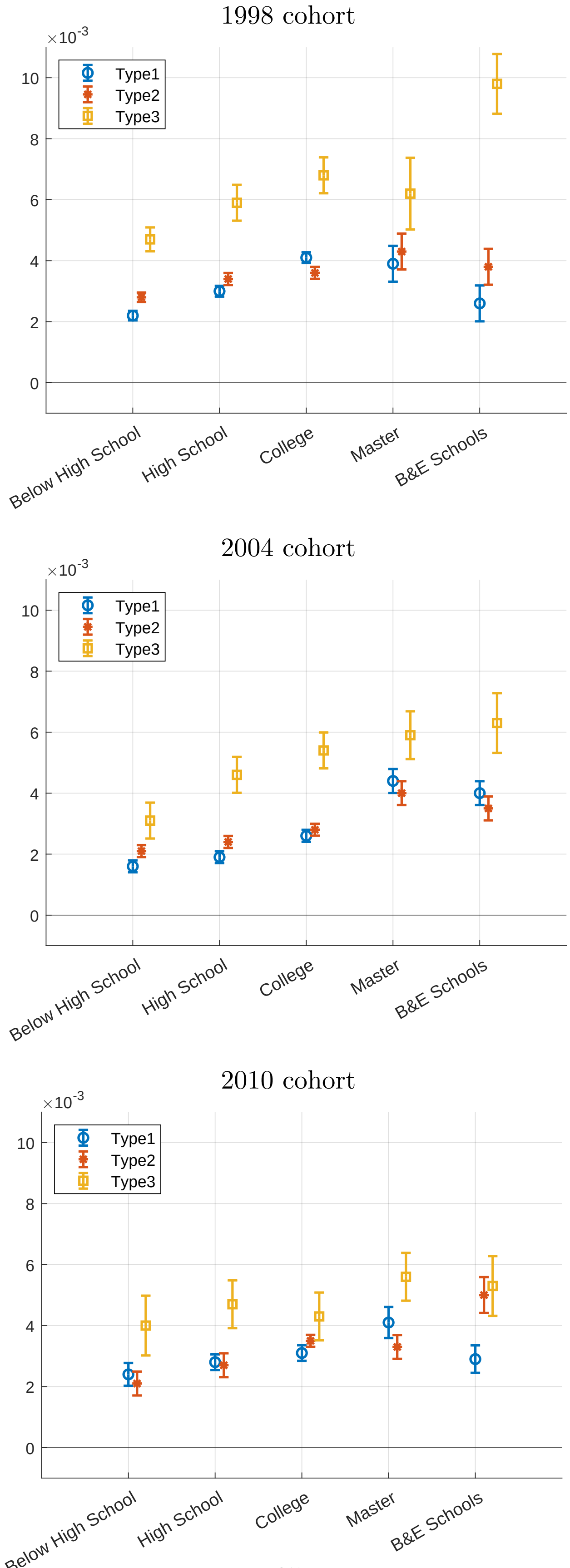
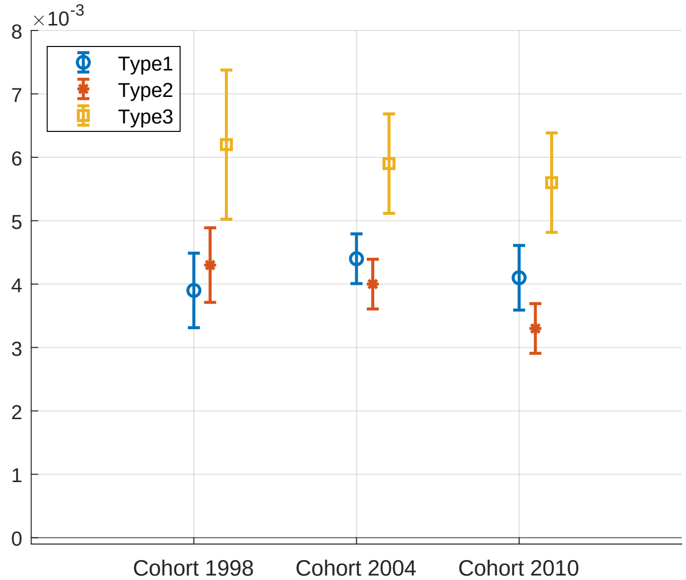


Figure 5: Evolution of Returns to Experience by Type: Masters



Returns to experience. Figure 4 shows the returns to experience in the three cohorts. Remember that these returns are average percentages of wage growth by month. The colors (and intervals) are distinguishing the 3 types, while each group of three intervals corresponds to a level of educational achievement. The most striking phenomena are, firstly, that type 3 (yellow intervals) has markedly higher returns to experience than the other types; secondly, that returns to experience typically increase with the level of educational achievement[32]; thirdly, returns to experience have tended to decrease with time, between 1998 and 2017, in particular for type 3. This is shown on Fig. 5 for the Master's degree holders. We see an "erosion" of the returns to experience.[33] The fall in returns to experience has been particularly important for the type 3 individuals who graduated from business and engineering schools. An exception is the return to type-2, business-school graduates, that has strongly increased in the most recent cohort.

Returns to education. We now turn to estimated returns to education or returns to degrees. These returns can also be viewed as returns at zero experience determining average starting salaries. The three panels of Figure 6 give the returns to education for the three cohorts, the three types (represented by three intervals with different colors) and the 5 education levels.

The education levels (5 groups of three bars) are clearly ranked (following the common, and expected hierarchy). In the beginning, in the 1998 cohort, type 2 gives the impression of dominating the highest educational levels, but in the 2004 cohort, we see a clear and consistent hierarchy of types:

[32] Business schools are an exception to these rules.

[33] In particular, the drop appears in the 2004 cohort; it then stabilized in the 2010 cohort for types 2 and 3, or it slightly increased again, without catching up the 1998 level for type 1.

Figure 6: Returns to Education by Type, Educational Attainment and Cohort

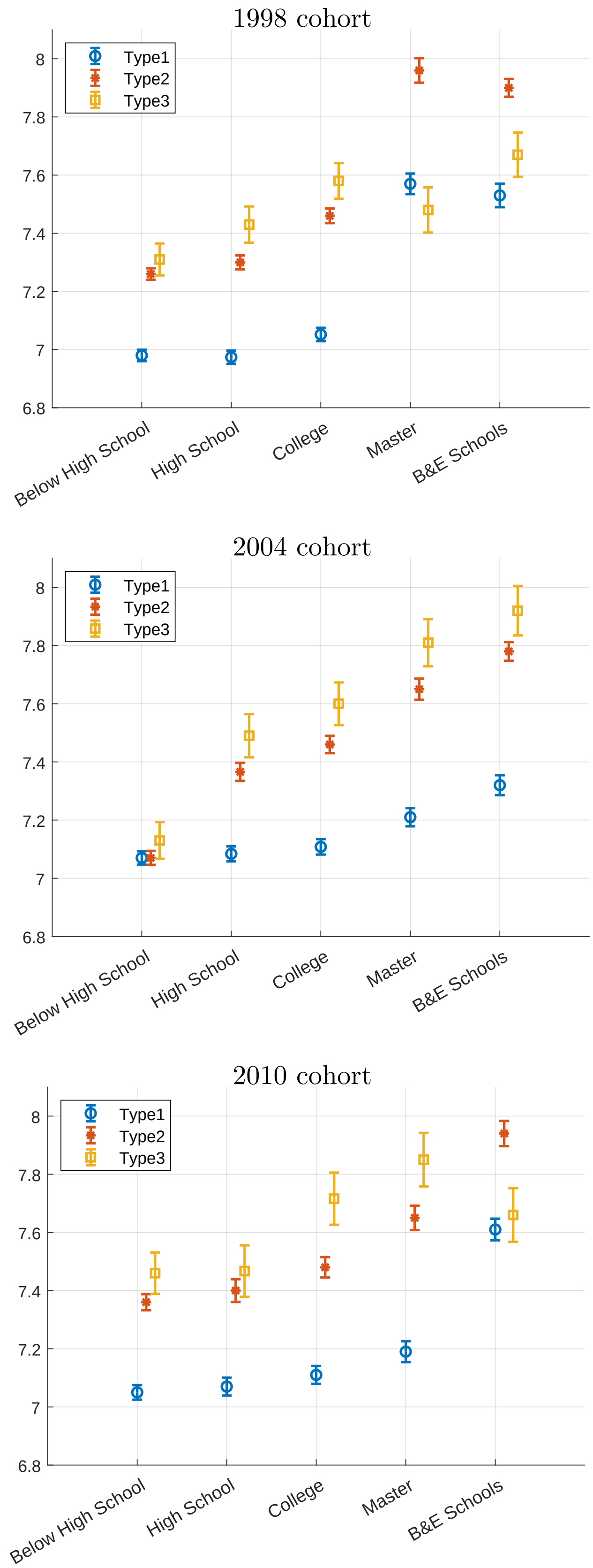
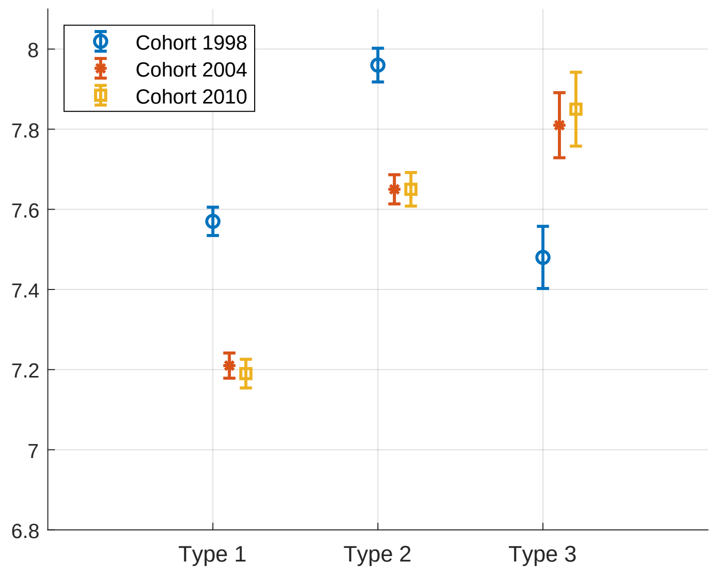


Figure 7: Devaluation of Master's Degrees 1998-2017

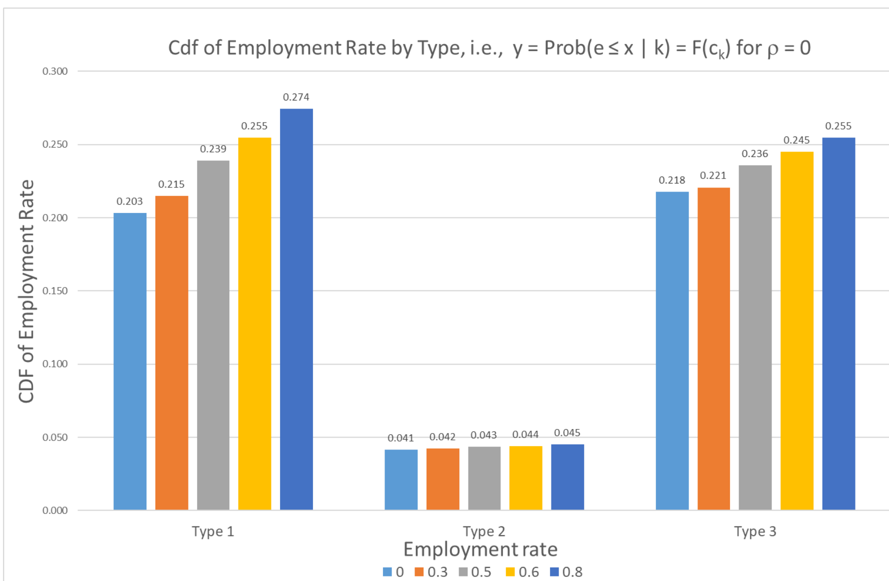


type 3 is simply the best everywhere; type 1 has the smallest returns and type 2 has median returns everywhere. The 2010 cohort confirms the hierarchy of types (with the exception of business schools).

Do we observe a devaluation of some degrees? Figure 7 now groups types by cohort on the same picture, to appreciate the possible devaluation of returns to degrees, in the case of Masters. Note that, on this picture, different colors now represent different cohorts. It is very striking that the average wages, conditional on type and a Master's degree have been devalued for types 1 and 2, but not for type 3. Devaluation of Masters' degrees is confirmed, but it is heterogeneous. If we compute the weighted variation of log-wages from the 1998 cohort to the 2010 cohort, using the type frequencies of Table 6 as weights, we find a drop of $.0747 = \Delta w \simeq 7.676 - 7.602$, and $e^{-.0747} - 1 \simeq -.072$, that is, a 7.2% devaluation of Masters' degrees. This corresponds to the result that can be obtained with a simple regression of log-wages on cohort and degree dummies (see Table 16 in Online Appendix D). A similar computation would show that there is no devaluation for the *Some College and Bachelors'* level. Yet, we observe a *relative* devaluation of this level, which is directly visible from the estimated coefficients displayed by Table 11. There is a relative devaluation of Bachelors with respect to the Less-than-High-School-Degree level. As already noted, this relative devaluation is mainly due to the fact that minimum-wage regulations protect the real value of wages from depreciation at the lowest educational level. In contrast, the devaluation of Master graduates' average real wages is *absolute*.

Impact of some control variables. Table 12, in Appendix D, provides the estimated coefficients of some important control variables. We control the wage equation for the macroeconomic unemployment rate. This is a way of controlling for the impact of the business cycle on wages. The impact on type 1 is not very significant. This is probably due to the fact that type 1 tends to reach education levels at which wages are protected by the minimum wage legislation. But the impact of overall unemployment is clearly negative for types 2 and 3, as expected: we find mildly procyclical real wages. Secondly, we find a significant and positive effect of the *father is a professional* dummy. This latter effect is much stronger for type 3 (five times more than the effect on type 1). This dummy indicates individuals whose father's occupation requires higher-education degrees: executives, doctors, lawyers, engineers, teachers, etc. The reference individual belongs to the 1998 cohort, lives in urban areas, has a father which is not a "professional" in the above sense. Indications of geographic origin are significant too: the rural and peri-urban individuals earn (slightly) smaller wages.[34]

Figure 8: Employment Rates by Type: Analysis of Ordered Probit Cuts



5.3.2 Employment equation

Estimates of the Ordered-Probit parameters are presented in Table 14, Online Appendix B. The Ordered Probit shows a striking feature of type 2 individuals. This is visible if we look at the Ordered-Probit cuts. Figure 8 gives a representation of these cuts. To be more precise, the table

[34]The Peri-urban is a heterogeneous category including neither purely urban nor purely rural individuals: it typically includes suburban and smalltown France. Note that, unlike in America, the French suburban individual generally does not have a well-to-do background. The urban individual is more likely to come from a privileged background.

gives $\Pr(e \leq x|k) = F(\mathbf{c}_k)$, where $x \in \{0, .3, .5, .6, .8\}$, F is the standard normal c.d.f and $\mathbf{c}_k$ is the corresponding Ordered-Probit cut. In other words, we consider an individual with all controls set equal to 0 — hence we have $\rho = 0$ —, and conditional on type k, we compute the cumulative probabilities that this individual has an employment rate e smaller than x. Figure 8 very clearly shows that type 2 has a very small probability of unemployment (around 4%) and a high probability of full employment of 95.5%. In contrast type 1 and type 3 have, respectively, a 72.6% and a 74.5% probability of being fully employed when $X = 0$. These results give the impression that type 2 finds a job quickly and stays in this job: the matching of type 2s with employers seems very stable as compared to that of the other types. As a counterpart, these individuals obtain smaller wages at the start and, as time passes, obtain smaller pay raises than type-3 individuals. Online Appendix H gives further details on the Ordered Probit and the reason for the observed differences between Type 2 and the other types in terms of employment. In particular, we study the possibility that Type 2 has a preference for the public sector.

5.3.3 Education choices

Table 6 gives the empirical values of conditional probabilities $\hat{p}(k|h, c)$, using the estimated posterior probabilities; this table shows that type 1 is more prevalent among individuals with the shortest education, in particular in the most recent cohorts. In the years 2010-17 (*i.e.*, the third cohort), more than 50% of those who did not go to college are type-1 individuals. In contrast, in the same category, we find around 30% of type-2 individuals and less than 20% belong to type 3. The distribution was closer to uniform in 1998. In 2010, there are less type-3 students in the *some college* category and a greater proportion of them in the business and engineering schools, as compared to 1998. We also see that 71% of the individuals with a Master degree are members of type 2 or type 3 in 2010, in contrast with 62% of the same categories in 1998. Also in contrast, the distribution of types among students who graduated from schools (*i.e.*, business and engineering schools) is closer to uniform in 2010 as compared to 1998 and 2004, where a large majority were members of type 2. To sum up, we see that the mix of types has changed, conditional on degrees.

Table 6 shows that the sorting of students has increased with time in universities and for those with an attainment below or equal to high-school graduation. Business and Engineering schools are an exception since sorting has decreased, schools admitting more members of types 1 and 3.

Table 6: Mix of types by education level and cohort

	$p(k\|h,c)$								
	1998 cohort			2004 cohort			2010 cohort		
Probability of type ...	1	2	3	1	2	3	1	2	3
Less than High-school Degree	0.43	0.33	0.25	0.50	0.33	0.18	0.52	0.32	0.16
High-school Degree	0.42	0.37	0.21	0.48	0.34	0.18	0.54	0.25	0.21
Some College and Bachelors	0.38	0.40	0.22	0.38	0.38	0.23	0.40	0.41	0.19
Masters	0.38	0.29	0.33	0.29	0.39	0.32	0.29	0.38	0.33
Bus. Engin. School Degrees	0.21	0.51	0.28	0.19	0.57	0.24	0.37	0.27	0.36

Table 7: Probability of reaching an education level given the type and cohort

	$p(h\|k,c)$								
Conditional on cohort ...	1998			2004			2010		
and conditional on type ...	1	2	3	1	2	3	1	2	3
Less than High-school Degree	0.43	0.37	0.43	0.34	0.26	0.24	0.29	0.24	0.18
High-school Degree	0.26	0.25	0.22	0.29	0.23	0.22	0.31	0.20	0.25
Some College and Bachelors	0.26	0.30	0.26	0.27	0.31	0.32	0.24	0.34	0.23
Masters	0.03	0.02	0.04	0.06	0.09	0.13	0.08	0.14	0.18
Bus. Engin. School Degrees	0.02	0.06	0.05	0.03	0.11	0.08	0.08	0.08	0.16

Table 8: Probability of reaching an education level given the type and cohort: aggregation of education levels

	$p(h\|k,c)$								
Conditional on cohort ...	1998			2004			2010		
and conditional on type ...	1	2	3	1	2	3	1	2	3
High-school Degree and Less	0.69	0.62	0.65	0.63	0.49	0.46	0.60	0.44	0.43
Some College and Bachelors	0.26	0.30	0.26	0.27	0.31	0.32	0.24	0.34	0.23
Masters and School Degrees	0.05	0.08	0.09	0.09	0.20	0.22	0.16	0.22	0.34

Table 7 provides a different point of view on the same reality and displays conditional probabilities of choosing level h, given the type k and cohort c, that is, $\hat{p}(h|k,c)$. This table confirms the observations already made, and also that the types are far from being completely characterized by their education level. As time passes, types seem to specialize more but all of them are characterized by shifts towards longer studies. To clarify the differences in educational 'choices' or (attainments) of the different types, Table 8 aggregates the degrees in three groups, with a clear hierarchy, and the essential phenomenon appears: 60% of type-1 students end up with a high-school degree or less in the last cohort, Bachelors (and the equivalent of Associates's degrees) are increasingly the common choice of type 2, while the type 3 (and to a lesser extent the type 2) are more concentrated at the top of the degree scale. It seems that the differentiation of type 2 and type 3 has increased with time (because their educational 'choice' patterns were very close in the 1998 cohort). Table 7 shows that there has been a rush on Master programmes and schools, and a certain flight from the lowest levels and the 'some college' category. The types differ only in the intensity of these changes.

The estimated parameters of the Multinomial Logit, describing education choices, are presented in Online Appendix A, Table 13. We discuss the impact of family background by type in Online Appendix I.

5.4 Conclusion of the analysis of unobserved heterogeneity: who are the types 1, 2 and 3?

We can now summarize a reasonable interpretation of types.

1°) Type 1 has smaller returns to experience and smaller returns to education than other types. These individuals also tend to study less than other types. It seems that it is the group of individuals with a smaller ability.

2°) Type 2 occupies a median position in terms of returns to education, between Type 1 and Type 3, but closer to Type 3. Type 2 also occupies the median position in terms of returns to experience, but this time, closer to Type 1 than to Type 3. Type 2 is strongly characterized by a high employment rate, around 95%. To sum up, the Type 2 have a good level of ability and find stable jobs but their earnings grow slowly as compared to Type 3.

3°) Type 3 is clearly the 'top type' in the sense that these individuals are strongly characterized by markedly higher returns to experience. They also obtain the highest returns to degrees but have a much smaller employment rate than type 2, around 75%.

None of the types is easily predicted, or characterized by certain values of observable control variables, but we find some unsurprising correlations with family and geographical background variables.

5.5 Are unobservable types determined by, or correlated with, neglected observable characteristics?

By construction, types are supposed to be orthogonal to observable characteristics present in our estimation model. However, are they correlated with omitted pre-market variables? Can we observe the determinants of types?

To study this point, we used regularized regressions of type probabilities on pre-market variables, and more precisely, an *elastic net* method to select the variables correlated with type among all available controls. The elastic net is a *regularized regression* method that linearly combines the L_1 and L_2 penalties of the lasso and ridge regression methods. We first assign to each individual his most likely type (*i.e.*, the type with the highest ex-post probability). Then, using a ridge regression with a multinomial model where the hyper-parameters are estimated by cross-validation, we estimate which variables are correlated with the types. A first reassuring result is that observable variables do not help predicting correctly the types. The confusion matrix shows very poor prediction results. However, among selected variables, results show that Type 2 and Type 3 are more prevalent among individuals who did not repeat a grade before junior high school, while Type 1 is more prevalent among individuals who have repeated a grade before high school. Type 3 is less prevalent among individuals who lived in rural areas at age 11 whereas types 1 and 2 are more prevalent among them. In addition, Types 2 and 3 are less prevalent among individuals whose parents work in agriculture, whereas Type 1 is more prevalent among members of this group. Type 1 is more common among individuals living in the south and west of France whereas Type 3 is more common among individuals living in Paris or the Paris region. Finally, Type 3 is also associated with individuals whose parents have a university degree. In Online Appendix C, Table 15 gives the results of the elastic net procedure. We conclude that types are somewhat correlated with some observable characteristics, but that types are not just dummies for omitted observable characteristics. For instance, family background is not a good predictor of types; the three types are present in all families.

6 Conclusion

In this article, we studied the evolution of wages during the early years of career of a large panel of individuals, in France. We stacked three surveys covering the first 7 years of career of young workers in France, from 1998 until 2017. The dataset takes the form of an unbalanced panel. We estimated a model describing the education choices, the accumulation of effective experience and individual wages simultaneously. Unobserved heterogeneity is handled by means of a finite set of latent individual types (a finite mixture model). Each type has its own Mincerian log-wage equation, its own employment-rate equation and education-choice model. The full model is estimated by means of standard Maximum Likelihood methods, using a sequential EM algorithm to find preliminary estimates. On top of a full set of type-dependent parameters, the estimation procedure yields the prior probabilities of types and, using Bayes' rule, the posterior probability that each individual belongs to any given type (*i.e.*, a probabilistic *classification* of individuals). From these results we compute policy-relevant parameters, such as the ATT and ATE of various education levels. The overall ATTs and ATEs can be expressed as averages of type-dependent treatment effects. So we obtain a representation of unobserved heterogeneity. This allowed us to show that the variation in time of the average real wages of workers, given a type of degree, is in some interesting cases the average of devaluations for some types, and wage increases for other types. In a similar fashion, the returns to experience and experience accumulation are themselves heterogeneous. The devaluation (*i.e.*, absolute drops) of the real wages of Master's degrees holders is an average of divergent evolutions conditional on type. Overall, between 1998 and 2017, and after 7 years of career, the absolute variation of a Master degree holder's ATE is a drop of around 600 euros per month, if we treat the high-school dropouts as the untreated. The parallel variation of the ATT is a smaller drop of around 400 euros. We observe that the selection of students (or the quality of students) has improved with time in French Master programs, in spite of the growth in enrollment. We conclude that the observed devaluation is likely to be due to an excess supply of graduates because it cannot be attributed to a lesser average quality or productivity of the graduates.

7 References

References

ACEMOGLU, D. (1999): "Changes in unemployment and wage inequality: An alternative theory and some evidence," *American economic review*, 89(5), 1259–1278.

ACEMOGLU, D., AND D. AUTOR (2011): "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of labor economics*, vol. 4, pp. 1043–1171. Elsevier.

ALLMAN, E. S., C. MATIAS, AND J. A. RHODES (2009): "Identifiability of Parameters in Latent Structure Models with Many Observed Variables," *Annals of Statistics*, 37(6A), 3099–30132.

ARCIDIACONO, P., E. AUCEJO, A. MAUREL, AND T. RANSOM (2016): "College Attrition and the Dynamics of Information Revelation," Working Paper 22325, National Bureau of Economic Research.

ARCIDIACONO, P., V. J. HOTZ, A. MAUREL, AND T. ROMANO (2020): "Ex Ante Returns and Occupational Choice," *Journal of Political Economy*, 128(12), 4475–4522.

Arcidiacono, P., and J. B. Jones (2003): "Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm," *Econometrica*, 71(3), 933–946.

Argan, D., and R. J. Gary-Bobo (2021): "Les diplômes français se sont-ils dévalorisés?," *Revue Economique, forthcoming.*

Ashworth, J., V. J. Hotz, A. Maurel, and T. Ransom (2021): "Changes across cohorts in wage returns to schooling and early work experiences," *Journal of labor economics*, 39(4), 931–964.

Autor, D., C. Goldin, and L. F. Katz (2020): "Extending the Race between Education and Technology," *AEA Papers and Proceedings*, 110, 347–51.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008): "Trends in US wage inequality: Revising the revisionists," *The Review of economics and statistics*, 90(2), 300–323.

Beaudry, P., D. A. Green, and B. M. Sand (2014): "The declining fortunes of the young since 2000," *American Economic Review*, 104(5), 381–86.

——— (2016): "The great reversal in the demand for skill and cognitive tasks," *Journal of Labor Economics*, 34(S1), S199–S247.

Beffy, M., D. Fougère, and A. Maurel (2012): "Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter?," *The Review of Economics and Statistics*, 94(1), 334–347.

Belzil, C., and J. Hansen (2020): "Reconciling changes in wage inequality with changes in college selectivity, using a behavioral model," Institut Polytechnique de Paris.

Blundell, R., D. A. Green, and W. Jin (2022): "The U.K. as a Technological Follower: Higher Education Expansion and the College Wage Premium," *Review of Economic Studies*, 89, 142–180.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2016): "Estimating multivariate latent-structure models," *Annals of Statistics*, 44(2), 540–563.

Bouveyron, C., G. Celeux, T. B. Murphy, and A. E. Raftery (2019): *Model-based clustering and classification for data science: with applications in R*, vol. 50. Cambridge University Press.

Card, D., and T. Lemieux (2001): "Can falling supply explain the rising return to college for younger men? A cohort-based analysis," *The quarterly journal of economics*, 116(2), 705–746.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003): "2001 Lawrence R. Klein Lecture Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice*," *International Economic Review*, 44(2), 361–422.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101(6), 2754–81.

Carneiro, P., and S. Lee (2011): "Trends in quality-adjusted skill premia in the United States, 1960-2000," *American Economic Review*, 101(6), 2309–49.

Cassagneau-Francis, O. (2021): "The Returns to Higher Education by Cognitive and Non-cognitive Abilities," Discussion paper, UCL, London.

Cassagneau-Francis, O., R. Gary-Bobo, J. Pernaudet, and J.-M. Robin (2021): "A Non-parametric Finite Mixture Approach to Difference-in-Difference Estimation, with an Application to On-the-job Trainig and Wages," Discussion paper, Sciences Po, Paris.

Corblet, P. (2022): "The Decreasing Returns to Experience for Higher Education Graduates in France: An Occupational Analysis," Discussion paper, Sciences Po, Paris.

Cunha, F., J. Heckman, and S. Navarro (2005): "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, 57(2), 191–261.

Cunha, F., and J. J. Heckman (2007): "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling," *Labour Economics*, 14(6), 870–893, Education and Risk.

Dustmann, C., and C. Meghir (2005): "Wages, experience and seniority," *The Review of Economic Studies*, 72(1), 77–108.

Emmons, W. R., A. H. Kent, L. Ricketts, et al. (2019): "Is college still worth it? The new calculus of falling returns," *Federal Reserve Bank of St Louis Review*, 101(4), 297–329.

Gary-Bobo, R. J., M. Goussé, and J.-M. Robin (2016): "Grade retention and unobserved heterogeneity," *Quantitative Economics*, 7(3), 781–820.

Goldin, C., and L. F. Katz (2008): *The race between education and technology.* harvard university press.

Guvenen, F. (2007): "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?," *American Economic Review*, 97(3), 687–712.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021): "What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?," *Econometrica*, 89(5), 2303–2339.

Hall, P., and X.-H. Zhou (2003): "Nonparametric estimation of component distributions in a multivariate mixture," *The Annals of Statistics*, 31(1), 201 – 224.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018): "Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking," *Journal of Political Economy*, 126(S1), S197–S246.

Heckman, J. J., and B. Singer (1984): "A Method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica*, 52, 271–320.

Heckman, J. J., and E. Vytlacil (2005): "Structural equations, treatment effects, and econometric policy evaluation 1," *Econometrica*, 73(3), 669–738.

Huggett, M., G. Ventura, and A. Yaron (2011): "Sources of Lifetime Inequality," *American Economic Review*, 101(7), 2923–2954.

Ichino, A., A. Rustichini, and G. Zanella (2022): "College Education, Intelligence and Disadvantage: Policy Lessons from the UK, 1960-2004," .

Jeong, H., Y. Kim, and I. Manovskii (2015): "The price of experience," *American Economic Review*, 105(2), 784–815.

Kambourov, G., and I. Manovskii (2009): "Occupational specificity of human capital," *International Economic Review*, 50(1), 63–115.

Kasahara, H., and K. Shimotsu (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77, 135–175.

Katz, L. F., and K. M. Murphy (1992): "Changes in relative wages, 1963–1987: supply and demand factors," *The quarterly journal of economics*, 107(1), 35–78.

Magnac, T., N. Pistolesi, and S. Roux (2018): "Post-Schooling Human Capital Investments and the Life Cycle of Earnings," *Journal of Political Economy*, 126(3), 1219–1249.

Magnac, T., and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70(2), 801–816.

McLachlan, G., and D. Peel (2000): *Finite Mixture Models*, Wiley series in probability and statistics: Applied probability and statistics. Wiley.

Valletta, R. G. (2018): "Recent flattening in the higher education wage premium: Polarization, skill downgrading, or both?," in *Education, skills, and technical change: Implications for future US GDP growth*, pp. 313–342. University of Chicago Press.

Verdugo, G. (2014): "The great compression of the French wage structure, 1969–2008," *Labour Economics*, 28, 131–144.

A Descriptive Statistics

Descriptive statistics. Table 9 presents the descriptive statistics of our sample. The share of individuals who graduated from university (with a master degree), from a business or an engineering school has increased substantially. We also observe that their average real log wage has decreased. In contrast, the share of individuals with less than a high-school degree has decreased, but their average log-wage has increased. In general, we observe that the average real monthly wage of full-time employees increases with education. More interestingly, we remark that there is an increase in the range of the share of individuals working full-time across education degrees. Whereas in 1998, the share of individuals working full time varied from 62% (low education level) to 68% (high education level), in 2010, the same share varied from 46% among the individuals without a high-school degree to 76% for individuals who graduated from a business (or engineering) school.

Table 9: Descriptive Statistics:
Average Employment Rate and Full-Time Real Log-Wages

Cohort	Education	Individuals	Obs.	Average Employment	Average FT Real Log-Wage
1998	Less than High School	3026	35285	0.65	7.16
	High-School Degree	1815	19216	0.64	7.23
	Some College, Bachelors	2038	20902	0.64	7.39
	Masters	203	1826	0.64	7.72
	Bus. and Engin. Sch. Deg.	301	2777	0.69	7.85
	All	7383	80006	0.65	7.28
2004	Less than High School	1601	20475	0.61	7.23
	High-School Degree	1406	16038	0.65	7.27
	Some College, Bachelors	1628	16580	0.68	7.39
	Masters	470	4363	0.70	7.65
	Bus. and Engin. Sch. Deg.	395	3461	0.73	7.80
	All	5500	60917	0.65	7.36
2010	Less than High School	870	10855	0.50	7.22
	High-School Degree	917	10538	0.58	7.27
	Some College, Bachelors	964	9554	0.65	7.40
	Masters	419	3869	0.67	7.62
	Bus. and Engin. Sch. Deg.	351	2673	0.76	7.78
	All	3521	37489	0.60	7.39

B Likelihood

We derive the model's likelihood function. Individuals are indexed by $i = 1, \ldots, N$. Recall that the log-wage w_{it} is observed in a subset T_i of periods. The probability density of w_{it}, conditional on observed characteristics and latent type k, is denoted as follows,

$$p(w_{it}|x_{it}, X_i, h_i, k) = f_k(\epsilon_{itk}), \tag{20}$$

where f_k is the pdf of a normal distribution with mean 0 and standard deviation σ_{wk} and ϵ_{itk} is defined by equation (2). Now, denote $w_i = (w_{it})_{t \in T_i}$ and $x_i = (x_{it})_{t_i \in T_i}$. We have

$$\Pr(w_i|x_i, X_i, h_i, k) = \prod_{t \in T_i} p(w_{it}|x_{it}, X_i, h_i, k).$$

The probability of observing an employment rate conditional on past observed employment rates, exogenous characteristics, education level h and latent type k can be written as follows:

$$P_k(e_{it}|X_i, x_{it}, h_i) = \Pr(e_{it}|X_i, x_{it}, h_i, k) = \prod_{g=1}^{G} \left[F(\mathbf{c}_{g+1,k} - \rho_{itk}) - F(\mathbf{c}_{g,k} - \rho_{itk})\right]^{Q_{itg}}, \tag{21}$$

where

$$Q_{itg} = \begin{cases} 1 & \text{if } e_{it} = \mathbf{e}_g \\ 0 & \text{otherwise} \end{cases},$$

and F is the cumulative distribution function of the standard normal distribution. Finally, we denote the probability of choosing education level h_i conditional on observable characteristics Z_i and latent type k as follows,

$$\Lambda_k(h|Z_i) = \frac{\exp(\mathbf{v}_{ihk})}{\sum_{j=1}^{H} \exp(\mathbf{v}_{ijk})}. \tag{22}$$

Let now y_i denote the vector of outcomes of individual i, namely, observed wages w_{it}, observed employment rates e_{it} and the observed education (*i.e.*, highest degree) h_i. Let E_i be the set of periods during which i's employment rate e_{it} is observed. Let $e_i = (e_{it})_{t \in E_i}$. Recalling that $x_{it} = \sum_{\tau=1}^{t-1} e_{i\tau}$, we can write the conditional probability of e_i as follows,

$$\Pr(e_i \mid X_i, h_i, k) = \prod_{t \in E_i} \Pr(e_{it}|x_{it}, X_i, h_i, k), \tag{23}$$

where $x_{i\tau} = 0$ if i enters the labor market at time τ for the first time.

Then, we can write the contribution to likelihood of an individual i with type k as,

$$\begin{aligned} L_{ik} = L_{ik}(y_i|X_i) &= \prod_{t \in T_i} p(w_{it}|x_{it}, X_i, h_i, k) \prod_{t \in E_i} \Pr(e_{it}|x_{it}, X_i, h_i, k) \Pr(h_i|X_i, k) \\ &= \left(\prod_{t \in T_i} f_k(\epsilon_{itk})\right) \left(\prod_{t \in E_i} P_k(e_{it}|x_{it}, X_i, h_i)\right) \Lambda_k(h_i|Z_i), \end{aligned} \tag{24}$$

where ϵ_{itk} is defined by equation (2).

Now, integrating over latent types k, the contribution to likelihood of individual i can be written,

$$L_i(y_i|X_i) = \sum_{k=1}^{K} p_k L_{ik}(y_i|X_i), \tag{25}$$

The model Likelihood is $L = \prod_{i=1}^{N} L_i$, so that the Log-Likelihood is

$$\ln L = \sum_{i=1}^{N} \ln \left[\sum_{k=1}^{K} p_k L_{ik} \right], \tag{26}$$

The *posterior* probability that individual i is of type k is denoted p_{ik}; it can be expressed with the help of Bayes' rule and the likelihood, as follows,

$$p_{ik} = \Pr(k|X_i, y_i) = \frac{p_k L_{ik}}{\sum_{j=1}^{K} p_j L_{ij}}. \tag{27}$$

The posterior probabilities are a crucial ingredient in many useful computations. For all $k = 1, \ldots, K$, we have,

$$p_k = \frac{1}{N} \sum_{i=1}^{N} p_{ik}. \tag{28}$$

It is easy to see that the latter equation is a necessary condition for likelihood maximization[35], and it follows that this relation between priors p_k and posteriors p_{ik} holds when we use their numerical, estimated values.

C Simulations

The simulations can be decomposed in a few steps.

Step 1.

1(a). We first recursively simulate the employment level $\tilde{e}_{itk}$ for each (i, t), and each k, $t = 1, \ldots, T$, $k = 1, \ldots, K$ and $T = 84$.

We start with $t = 1$ and then increment. We initialize experience by setting $\tilde{x}_{i1k} = 0$. We draw a random number $\tilde{\zeta}_{itk}$ for each (itk), with $\tilde{\zeta}_{itk} \sim \mathcal{N}(0, 1)$. Then, we use the ordered probit as estimated by ML. More precisely, if it happens that

$$c_{gk} - \rho_{itk} \leq \tilde{\zeta}_{itk} \leq c_{g+1,k} - \rho_{itk},$$

where ρ_{itk} is given by 5 above, then we set $\tilde{e}_{itk} = e_g$. To compute ρ_{itk} we use $x_{it} = \tilde{x}_{itk}$ for $t > 1$. Recall that $e_g \in \{0, .3, .5, .8, 1\}$.

[35] Indeed, it is equivalent to $\partial \ln L / \partial p_k = 0$, for $k = 2, \ldots, K$ where we set $p_1 = 1 - \sum_{k=2}^{K} p_k$.

1(b) Compute the accumulated experience $\tilde{x}_{itk} = \sum_{\tau<t} \tilde{e}_{itk}$, with $\tilde{x}_{i1k} = 0$.

Step 2. Given the sequences $(\tilde{e}_{itk}, \tilde{x}_{itk})$, we compute a sequence of expected log-wages for each (i, t, k) (no need to draw a random shock here). Using the estimated values of the parameters, we set, for each (i, t, k),

$$\tilde{w}_{itk} = \mathbb{E}[w_{itk}|\tilde{x}_{itk}, X_i] = \alpha_{0k} + \beta_{0k}\tilde{x}_{it} + \sum_{h=1}^{H} \gamma_{0hk}\chi_h(i) + X_i\eta_{0k}.$$

Step 3. Given the simulated sequences $(\tilde{e}_{itk}, \tilde{w}_{itk}, \tilde{x}_{itk})$ we can now compute the discounted expected earnings during the periods $t \in \{1, \ldots, T\}$. We choose a discount factor δ and for every (i, k), we compute

$$\tilde{W}_{ik} = \frac{(1-\delta)}{(1-\delta^T)} \sum_{t=1}^{T} \delta^{t-1} \tilde{e}_{itk} \exp(\tilde{w}_{itk}).$$

$\tilde{W}_{ik}$ has the dimension of monthly earnings[36]

Then, we compute the weighted arithmetic mean, using the estimated probabilities p_{ik}. For each type k, we compute,

$$H_k = \frac{\sum_{i=1}^{N} \tilde{W}_{ik}\hat{p}_{ik}}{\sum_{i=1}^{N} \hat{p}_{ik}}.$$

We can also compute expected-discounted values conditional on a degree h. So, we define $I(h) = \{i|h_i = h\}$ and we compute,

$$H_k(h) = \frac{\sum_{i\in I(h)} \tilde{W}_{ik}\hat{p}_{ik}}{\sum_{i\in I(h)} \hat{p}_{ik}},$$

which measures the average expected-discounted earnings of a type k, knowing the degree h. This type of conditioning can be performed with any other subsample (for instance, the sons of executives, or the sons of executives with degree h, etc.). [37]

[36] We choose a yearly discount rate of 0.9845. This corresponds to a monthly discount rate $\delta = 0.9987$. $\tilde{W}_{ik}$ is a weighted average of expected monthly earnings with weights $(\delta^{t-1}(1-\delta))/(1-\delta^T)$.

[37] These computations can be improved, if needed, by simulating employment and wage trajectories several times in the same fashion and then taking the simple arithmetic averages of all simulated values of H.

D Full tables: Wage Equation

Table 10: Wage equation. Returns to experience

Type		1	2	3
Experience ×				
1998 cohort	Below High-school degree	0.0022	0.0028	0.0047
		(.00008)	(.00008)	(.0002)
	High school degree	0.0030	0.0034	0.0059
		(.00009)	(.0001)	(.0003)
	Some College and Bachelors	0.0041	0.0036	0.0068
		(.00009)	(.0001)	(.0003)
	Masters	0.0039	0.0043	0.0062
		(.0003)	(.0003)	(0006)
	Bus. Engin. School degree	0.0026	0.0038	0.0098
		(.0003)	(.0003)	(.0005)
2004 cohort	Below High-school degree	0.0016	0.0021	0.0031
		(.0001)	(.0001)	(.0003)
	High-school degree	0.0019	0.0024	0.0046
		(.0001)	(.0001)	(.0003)
	Some College and Bachelors	0.0026	0.0028	.0054
		(.0001)	(.0001)	(.0003)
	Masters	0.0044	0.0040	0.0059
		(.0002)	(.0002)	(.0004)
	Bus. Engin. School degree	0.0040	0.0035	0.0063
		(.0002)	(.0002)	(.0005)
2010 cohort	Below High-school degree	0.0024	0.0021	0.0040
		(.00019)	(.0002)	(.0005)
	High-school degree	0.0028	0.0027	0.0047
		(.00013)	(.0002)	(.0004)
	Some College and Bachelors	0.0031	0.0035	0.0043
		(.00013)	(.0001)	(.0004)
	Masters	0.0041	0.0033	0.0056
		(.00026)	(.0002)	(.0004)
	Bus. Engin. School degree	0.0029	0.0050	0.0053
		(.00023)	(.0003)	(.0005)

Table 11: Wage equation. Returns to education

Type		1	2	3
1998 cohort	High-school degree	-0.006	0.04	.12
		(.006)	(.007)	(.015)
	Some College and Bachelors	0.072	0.20	0.27
		(.006)	(.008)	(.014)
	Masters	0.59	0.70	0.17
		(.015)	(.019)	(.028)
	Bus. Engin. School degree	0.55	0.64	0.36
		(.018)	(.012)	(.027)
2004 cohort	High-school degree	0.014	0.016	0.03
		(.006)	(.01)	(.020)
	Some College and Bachelors	0.038	0.11	0.14
		(.007)	(.009)	(.019)
	Masters	0.14	0.30	0.35
		(.011)	(.014)	(.026)
	Bus. Engin. School degree	0.25	0.43	0.46
		(.013)	(.011)	(.029)
2010 cohort	High-school degree	0.02	0.04	0.007
		(.009)	(.014)	(.027)
	Some College and Bachelors	0.06	0.12	0.256
		(.009)	(.011)	(.028)
	Masters	0.14	0.29	0.39
		(.013)	(.016)	(.030)
	Bus. Engin. School degree	0.56	0.58	0.20
		(.014)	(.017)	(.030)

Table 12: Wage equation. Controls

Type	1	2	3
2004 cohort	0.09	0.09	0.15
	(.006)	(.007)	(.016)
2010 cohort	0.07	0.10	0.15
	(.008)	(.01)	(.023)
Father is a professional	0.012	0.019	0.06
	(.003)	(0.004)	(.006)
Peri-urban	-0.006	-0.010	-0.039
	(.003)	(.003)	(.007)
Rural	-0.018	-0.022	-0.025
	(.002)	(.003)	(.006)
Unemployment rate	0.002	-0.015	-0.020
	(.001)	(.001)	(.003)
Constant	6.98	7.26	7.31
	(.01)	(.01)	(.028)

A Online Appendix

Multinomial Logit; Full Estimation Results

Table 13: Multinomial logit: Education choice

Type		1	2	3
Below High-School Degree			*Ref.*	
High-School Degree	× 2004 cohort	0.30	0.29	0.59
		(.09)	(0.10)	(.13)
	× 2010 cohort	0.58	0.16	0.98
		(.1)	(.14)	(.15)
	×Father is a professional	0.87	0.67	1.04
		(.10)	(.13)	(16)
	×Peri-urban	0.003	-0.11	0.006
		(.09)	(.10)	(.14)
	× Rural	0.23	-0.31	-0.17
		(.08)	(0.10)	(.12)
	Constant	-0.68	-0.33	-0.75
		(.07)	(.08)	(.09)
Some College and Bachelors	× 2004 cohort	0.18	0.37	0.77
		(.09)	(.10)	(.12)
	× 2010 cohort	0.28	0.46	0.68
		(.1)	(0.12)	(.16)
	×Father is a professional	1.19	1.31	1.77
		(.10)	(.12)	(.15)
	× Peri-urban	-0.12	-0.25	-0.07
		(.09)	(.10)	(.13)
	× Rural	0.05	-0.62	-0.28
		(.08)	(.09)	(.12)
	Constant	-0.65	-0.15	-0.69
		(.07)	(.07)	(.09)
Masters	× 2004 cohort	0.85	1.74	1.79
		(.17)	(.19)	(.19)
	× 2010 cohort	1.29	2.13	2.33
		(.18)	(.20)	(.21)
	×Father is a professional	2.09	1.97	2.16
		(.14)	(.15)	(.18)
	× Peri-urban	-0.25	-0.57	-0.66
		(.17)	(.17)	(.20)
	× Rural	-0.09	-0.73	-0.66
		(.16)	(.16)	(.18)
	Constant	-3.14	-2.90	-2.59
		(.16)	(.18)	(.16)
Bus. and Engin. School Degr.	× 2004 cohort	0.45	0.99	1.07
		(.21)	(.14)	(.19)
	× 2010 cohort	1.56	0.61	1.98
		(.20)	(.18)	(.20)
	× Father is a professional	2.16	2.31	2.60
		(.16)	(.14)	(.18)
	× Peri-urban	-0.02	-0.67	-0.30
		(.18)	(.16)	(.20)
	× Rural	-0.22	-0.85	-0.42
		(.20)	(.15)	(.20)
	Constant	-3.39	-2.04	-2.64
		(.18)	(.12)	(.16)

B Online Appendix:
Ordered Probit; Full Estimation Results

Table 14: Ordered Probit: Individual Employment Rate

Type		1	2	3
1998 cohort			*Ref.*	
2004 cohort		-0.11	-0.22	-0.04
		(.02)	(.04)	(.04)
2010 cohort		-0.23	-0.28	0.07
		(.03)	(.04)	(.05)
1998 cohort	High-school Degree	0.08	-0.07	0.15
		(.02)	(.04)	(.03)
	Some College and Bachelors	0.10	-0.05	0.21
		(.02)	(.03)	(.03)
	Masters	0.19	0.19	0.09
		(.06)	(.09)	(.06)
	Bus. Engin. School Degrees	0.22	-0.07	0.33
		(.06)	(.05)	(.06)
	Experience	0.0179	0.0214	0.0176
		(.0004)	(.0008)	(.0006)
2004 cohort	High-school Degree	0.11	0.11	0.15
		(.02)	(.04)	(.04)
	Some College and Bachelors	0.19	0.20	0.35
		(.02)	(.04)	(.04)
	Masters	0.32	0.22	0.32
		(.04)	(.05)	(.05)
	Bus. Engin. School Degree	0.32	0.19	0.38
		(.05)	(.05)	(.06)
	Experience	0.0167	0.0207	0.0190
		(.0004)	(.0006)	(.0007)
2010 cohort	High-school Degree	0.20	0.19	-0.01
		(.03)	(.06)	(.05)
	Some College and Bachelors	0.30	0.32	0.24
		(.03)	(.04)	(.06)
	Masters	0.31	0.18	0.41
		(.04)	(.06)	(.07)
	Bus. Engin. School Degrees	0.93	0.56	0.20
		(.07)	(.08)	(.06)
	Experience	0.0209	0.0266	0.0206
		(.0005)	(.001)	(.0009)
Father is a professional		-0.07	-0.04	0.03
		(.013)	(.02)	(.02)
Peri-urban		0.07	0.04	0.07
		(.013)	(.02)	(.02)
Rural		0.10	0.06	0.13
		(.012)	(.02)	(.02)
Unemployment		-0.11	-0.18	-0.11
		(.006)	(.007)	(0.01)
Cuts	0-0.3	-0.83	-1.734	-0.78
		(.061)	(.066)	(0.09)
	0.3-0.5	-0.79	-1.724	-0.77
		(.061)	(.067)	(0.09)
	0.5-0.6	-0.71	-1.712	-0.72
		(.062)	(.068)	(0.09)
	0.6-0.8	-0.66	-1.707	-0.69
		(.062)	(.069)	(0.09)
	0.8-1	-0.60	-1.693	-0.66
		(.062)	(.070)	(0.09)

C Online Appendix: Results obtained with the Elastic Net Method

Table 15 reports the results of the elastic net method applied to the most-likely-type indicators. The coefficients of the explanatory variables selected by the algorithm appear in the table (otherwise, the entry is blank); these variables are significant. There are four groups of three columns, one for each type in each group of three. The first three column groups correspond to the three cohorts, 1998, 2004, 2010. The last group of three columns reports results obtained when the three cohorts are stacked. There are indicators of the father's and the mother's occupation and indicators of the region of origin listed in the bottom half of the table. Grade repetition; rural origin; parents are farmers; mother is a graduate; Corsica; South-West of France (Occitanie and Aquitaine); West Indies and Islands; Paris are the most salient indicator variables: this is not particularly surprising.

Table 15: Elastic-net regressions of posterior probabilities

	1998			2004			2010			All cohorts		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
1 year late				0,254	-0,114	-0,141						
2+ years late				0,173	-0,094	-0,079						
Repeated a grade							0,035	0,000	-0,237	0,20	-0,04	-0,16
Has not moved	0,011	-0,005	-0,006									
Urban area	-0,002	-0,011	0,013							0,004	-0,009	0,004
Peri-urban										-0,13	0,11	0,02
Rural area	0,127	0,008	-0,134	-0,002	0,023	-0,021				0,02	0,04	-0,06
Father is French										0,01	0,01	-0,01
Foreigner										-0,06	0,00	0,06
Mother is French										0,04	0,02	-0,06
French acquired										-0,04	-0,05	0,09
Foreigner							-0,046	0,000	0,000	-0,01	0,00	0,00
Father : worker										-0,06	0,04	0,03
unemployed							0,090	-0,078	0,000	0,05	0,03	-0,09
retired										-0,05	0,07	-0,03
at home (has worked)							0,090	-0,078	0,000			
at home (never worked)												
training										-0,11	-0,07	0,18
deceased										-0,04	-0,07	0,11
no answer										0,15	-0,01	-0,14
Mother : unemployed	0,018	-0,015	-0,003									
at home (has worked)							0,000	-0,049	0,000			
at home (never worked)							-0,019	0,000	0,000			
no answer				0,127	-0,095	-0,031						
Father : farmer	0,044	-0,016	-0,028							0,27	-0,11	-0,16
Craftsman, business										-0,03	-0,03	0,07
White collar	-0,081	0,020	0,060							-0,03	0,01	0,02
Technician				-0,155	0,036	0,119				-0,09	0,02	0,07
White collar							0,028	0,000	0,000	0,03	-0,01	-0,02
Blue collar							0,000	0,019	0,000	0,00	0,02	-0,02
Does not know										0,05	-0,03	-0,03
Mother : farmer	0,139	-0,053	-0,086							0,16	0,03	-0,19
Craftsman, business	-0,028	-0,008	0,035							-0,09	-0,06	0,15
White collar										-0,08	-0,02	0,09
Technician										0,00	-0,01	0,01
Blue collar										0,07	0,02	-0,09
Does not know				0,137	-0,081	-0,056	0,000	-0,116	0,000	0,11	-0,10	0,00
Auvergne-Rhone-Alpes										-0,03	0,06	-0,03
North (Hauts de France)										0,01	-0,08	0,07
Provence-Alpes-Cote d'Azur				0,000	-0,001	0,001	0,232	-0,148	0,000	0,12	-0,16	0,04
East (Grand Est)	-0,182	0,058	0,124							-0,14	-0,02	0,16
Occitanie	0,095	-0,079	-0,016	0,083	-0,023	-0,060	0,093	0,000	0,000	0,22	-0,09	-0,13
Normandie				-0,025	0,038	-0,014	-0,100	0,000	0,000	0,00	0,00	0,00
Nouvelle-Aquitaine	0,175	-0,066	-0,109	0,028	-0,011	-0,017				0,20	-0,05	-0,14
Centre-Val de Loire										0,00	0,04	-0,03
Bretagne										0,03	0,02	-0,05
Corse										0,24	0,03	-0,27
Pays de la Loire	0,009	0,000	-0,010	0,000	0,004	-0,004	0,000	0,016	0,000	0,08	0,08	-0,16
Paris	-0,125	0,056	0,069				0,000	0,000	0,000	-0,31	-0,02	0,33
Ile-de-France	-0,024	-0,008	0,032	-0,032	-0,022	0,055	-0,031	0,000	0,025	-0,16	-0,02	0,19
Ile-de-France, St Denis (93)										-0,06	0,00	0,06
West Indies, Islands (DOM)							0,152	0,000	0,000	0,23	-0,17	-0,06
Father, graduate							0,000	0,000	0,056			
does not know							0,130	0,000	0,000			
Mother, graduate							0,000	0,000	0,204			
does not know							0,151	0,000	0,000			

D Online Appendix: A Preliminary Analysis Using Standard Econometric Methods

D.1 The devaluation of degrees.

D.1.1 Estimation in sub-samples of students holding the same degree

We now present some preliminary results obtained with the unbalanced panel stacking the surveys of 1998, 2004 and 2010. The details are given in several appendices. We first consider the most classical log-wage regression. Log-wages are regressed on potential experience and dummies interacting the education level with the cohort. Potential experience is denoted z_{it}. Potential experience is defined as the number of months elapsed since the individual left the educational system. In essence, we studied variants of the following regression:

$$w_{it} = a + \sum_k \mathbb{1}_k(i)(b_k + c_k z_{it} + d_k z_{it}^2) + \epsilon_{it}, \tag{29}$$

where (a, b_k, c_k, d_k) are parameters, ϵ_{it} is a random error with a zero mean and $b_1 = 0$. Index k can denote the education level, the cohort (the survey), gender, or any interaction of the three. Variable $\mathbb{1}_k(i)$ is a dummy equal to 1 when i has characteristic k.

Table 16: Devaluation of Degrees

	Less than High-school	High-School	Some College Bachelors	Masters (M2)	Bus. and Eng. Schools
2004	0.0663***	0.0477***	0.00272	**-0.0671*****	**-0.0497*****
	(0.00287)	(0.00359)	(0.00390)	(0.0114)	(0.00966)
2010	0.0574***	0.0472***	0.0161***	**-0.0918*****	**-0.0644*****
	(0.00391)	(0.00421)	(0.00460)	(0.0116)	(0.01000)
Constant	7.164***	7.225***	7.388***	7.717***	7.846***
	(0.00167)	(0.00238)	(0.00257)	(0.00961)	(0.00719)
Observations	37868	26659	28835	6389	6261
Individuals	5497	4138	4630	1092	1047

Note. Results obtained by means of OLS on the panel obtained by stacking three 7-year Generation surveys 1998, 2004 and 2010. The dependent variable is the logarithm of the monthly real wages of male individuals with a full-time job. The 1998 cohort is the reference. Stars indicate degrees of statistical significance of the estimated coefficients; * for a p-value<0.05, ** for a p-value<0.01 and *** for a p-value<0.001.

We start with a test showing the devaluation, on average, of higher education degrees (we set

$c_k = d_k = 0$). Table 16 gives the results of 5 simple sub-sample regressions, one for each education level, of log-wages on dummies indicating the cohort, estimated by OLS on pooled data, without any control for experience. Taking the 1998 cohort as a reference, we find a significant devaluation for some degrees, the drop in average real wages being particularly clear, of the order of -9% for the Master's (M2) and -6% for the Engineering and Business school degrees, between the 1998 and 2010 cohorts. In contrast, the corresponding results for attainment levels below or equal to high-school graduation (*i.e.*, below the French *baccalauréat*) did not suffer any devaluation. On the contrary, in these categories, we see only real-wage increases. This striking difference can be attributed to minimum-wage regulations. Indeed, the real-value of the minimum wage rose by 26% between 1992 and 2012. This substantial growth protected the less skilled working full-time from the devaluation observed at the other end of the hierarchy of degrees. Of course, these results do not take care of the fact that years of education are an endogenous variable; our coefficients are not average treatment effects, only ATTs.

D.1.2 The Decline of Returns to Experience

In fact, we can show that a substantial part of the observed devaluation takes the form of a decrease in the returns to potential (or effective) experience during the first 7 years of career. If we add a control for potential experience in linear form (*i.e.*, now allowing for $c_k > 0$ and keeping $d_k = 0$), we find that the returns to potential experience decreased with time (see Table 17 in Appendix E). Returns to potential experience are of the order of 3% per year. They increase with the number of years of education. Appendix E gives the within-group estimates of returns to potential and effective experience obtained with our data.[38]

D.2 Unemployment, the Business Cycle and the Supply of Graduates

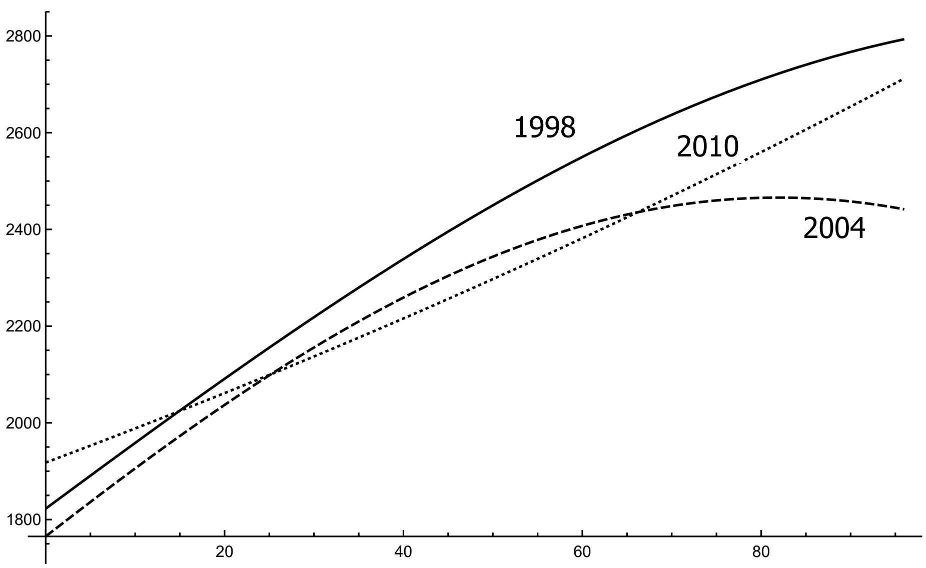
A look at the overall unemployment rate is needed, for the observed devaluation of degrees could entirely be due to a business-cycle effect. We consider the national unemployment rate, which is correlated with GDP. Variations of the unemployment rate do have an impact on real wages in our dataset. The real wages are sticky and mildly procyclical. In any case, we will control for the variation of national unemployment to capture a business-cycle effect. A discussion of this point is developed in Appendix F.

[38] A comparison of OLS and *within* estimators shows that OLS estimates of returns to effective experience by means of regression (29) are upward biased. If the bias is small for basic secondary degrees, it is in contrast substantial in the case of higher-education degrees.

E Online Appendix: The Returns to Experience. Fixed-effects, *Within* Estimators

To summarize the information conveyed by many regression coefficients, we can plot the wage $W = e^w$ as a function of potential experience and potential experience squared for top and bottom degrees. Coefficients of the Mincer regression (29) are all precisely estimated. We used these coefficients to draw the curves of Figures 9 and 10. We see the devaluation of top degrees in the case of male students on Fig. 9: the dotted line representing the Mincer curve of the 2010 cohort starts slightly above the other two, but after a year and a half of career, the young workers of the 1998 cohort were better off. This result is no longer true for the lowest degrees on Fig. 10.[39] The latter result can be almost entirely attributed to the growth of the French minimum wage. These pictures are of course not statistical tests; they should be viewed as a convenient representation of estimated coefficients, but with our data, the returns to experience are very precisely estimated (see below).

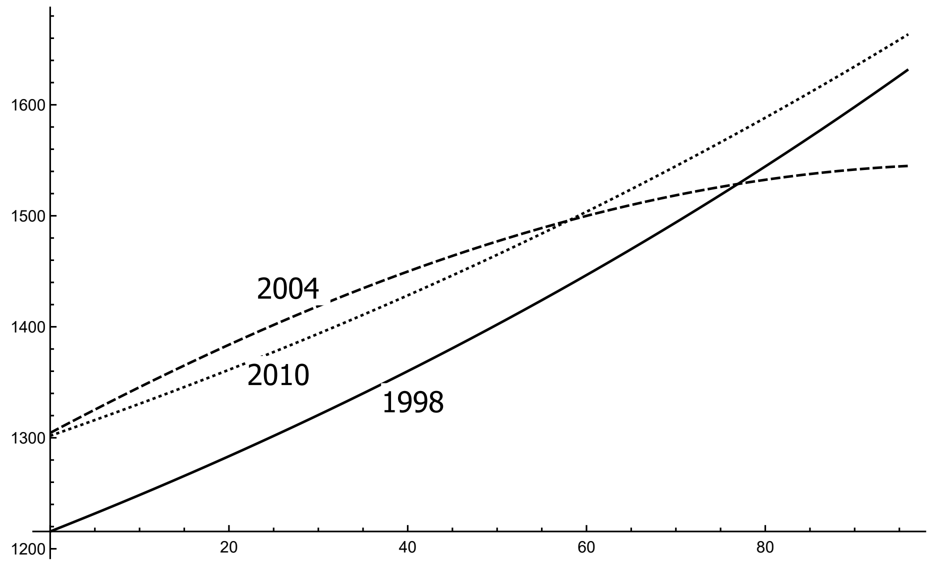
Figure 9: **Men. Masters and 'Schools' Degrees**



Comparison of returns to education and experience of male workers holding *Master's, business and engineering school degrees* in three 7-year Generation surveys, 1998, 2004, 2010. Months of potential experience are on the x-axis; monthly real wages (2013 euros) are on the y-axis.

[39] Appendix **??** shows that the picture is somewhat different for young women, whose salaries resisted devaluation much better: after 5 years (*i.e.*, 60 months), the 1998 curve catches up the 2010 curve. In Appendix **??**, Fig. **??** shows that the less educated women did not experience any devaluation during our 20 year period.

Figure 10: **Men. High-School Degree and Less**



Comparison of returns to education and experience of male workers with an educational achievement lower than or equal to high-school graduation *i.e.*, the French *baccalauréat* in three 7-year Generation surveys, 1998, 2004, 2010. Months of potential experience are on the x-axis; monthly real wages (2013 euros) are on the y-axis.

Using the panel structure of our data, we can easily obtain fixed-effects, within-group estimates of returns to experience, either potential (*i.e.*, z_{it}) or effective (*i.e.*, x_{it}). We assume that the endogeneity of experience stems from additive individual effects (*i.e.*, the so-called fixed effects).[40]

We again study the wages of full-time jobs, but we use all employment spells to compute an individual's effective experience (as explained above). Effective experience is potentially highly endogenous, because individuals with the best characteristics on the labor market also accumulate more experience. OLS estimates of the returns to effective experience should therefore overestimate these returns. This is what we find.

Table 17 permit a comparison of estimated returns of both potential and effective experience in the three cohorts and for three aggregate levels of educational attainment. The estimates of c_k are monthly returns to experience.[41] We obtain yearly returns, denoted γ_k, with the formula, $\gamma_k = (1 + c_k)^{12} - 1$. All c_k coefficients are significant at the .1 percent (*i.e.*, 10^{-3}) level.[42] Table 17 gives coefficients γ_k, where c_k is estimated with two different methods: by OLS on pooled data and by the fixed-effects, *within* estimator (FE).

The first striking fact is that returns to experience are substantial, with values ranging from 2%

[40]Error terms can be written $\epsilon_{it} = u_i + \nu_{it}$ where ν_{it} have a zero mean and are independent of explanatory variables and u_i. Terms u_i are individual effects depending on i that do not vary with time t. The within estimator will then produce unbiased estimates of c_k and d_k.

[41]In Appendix **??**, Table **??** gives the equivalent results for the subsample of women.

[42]We assumed $d_k = 0$ here.

to 6% per year. The OLS returns to potential experience seem to be only slightly biased (if we compare the estimated coefficients with the corresponding fixed-effects coefficients). In contrast, as expected, the OLS returns to effective experience are biased upwards, and all the more since the attainment level is high. Returns to experience typically increase with the education level, in all cohorts. But the most important feature of Table 17 is that returns to experience fell between 1998 to 2004, and they fell more for the highest degrees of attainment.

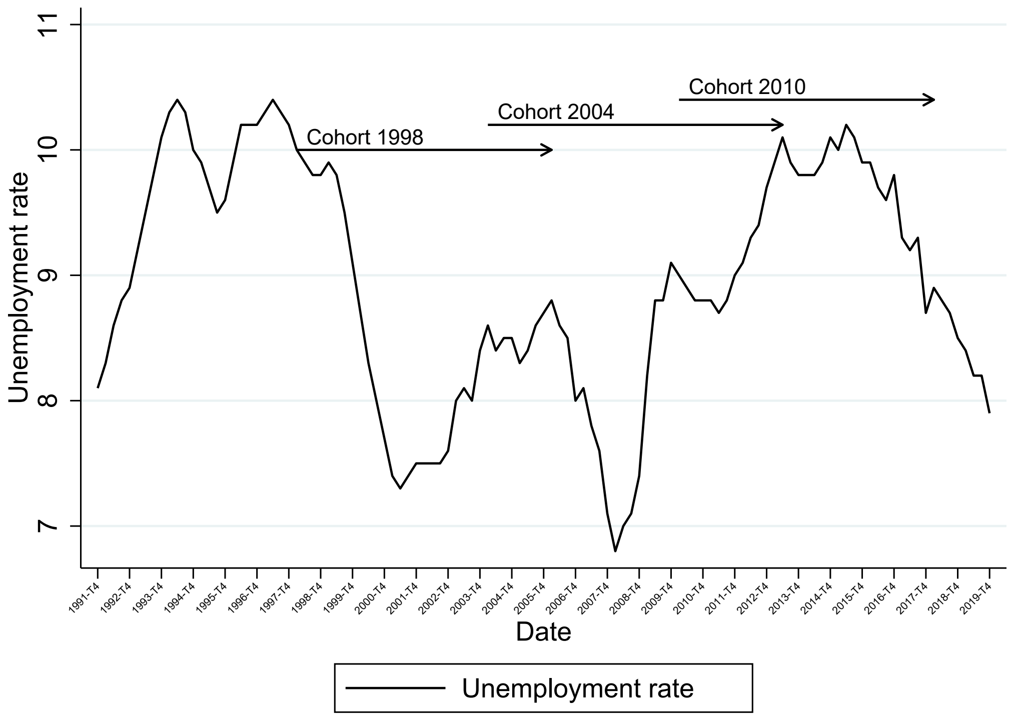
Table 17: Yearly returns to potential and effective experience of men

MEN		Potential Experience		Effective Experience	
$\gamma_k = (1 + c_k)^{12} - 1$		OLS	FE	OLS	FE
High School and less	1998	0.0339	0.0372	0.0444	0.0437
Some College and Bachelors		0.0511	0.0533	0.0638	0.0572
Masters and schools		0.0572	0.0564	0.0733	0.0585
High School and less	2004	0.0200	0.0224	0.0320	0.0289
Some College and Bachelors		0.0325	0.0320	0.0421	0.0377
Masters and schools		0.0468	0.0449	0.0665	0.0499
High School and less	2010	0.0237	0.0309	0.0411	0.0403
Some College and Bachelors		0.0393	0.0387	0.0498	0.0431
Masters and schools		0.0449	0.0442	0.0603	0.0477

Note. Results obtained with pooled data stacking the 7-year Generation surveys of 1998, 2004 and 2010, considering males only. The dependent variable is the logarithm of real-wages of individuals with a full-time job. For potential experience as well as for effective experience, the first column on the left gives the OLS estimates, the second column on the right gives the *within*, fixed-effects estimates. Regressions are weighted, using the CEREQ survey weights. All the displayed c_k coefficients are significant at the 1% level.

F Online Appendix: Impact of the Business Cycle. Variations of the National Unemployment Rate

Figure 11: **Unemployment rate; France, 1992-2019**



In France, from 1998 to 2002, the national unemployment rate dropped from 10.3% to 7.9%. Then, unemployment grew again and reached 9% in 2010 and 10.4% in 2015, as shown by Figure 11. In spite of the relatively better macroeconomic conditions of 1998-2001, the devaluation of higher-education degrees is already visible when we compare the surveys of 2004 and 1998, as shown by Table 16. The first years of *Génération* 2004 are characterized by a relatively small national rate of unemployment. In the middle of the 7-year period, *i.e.*, after 2007, the effects of the great recession started to be felt, slowing down wage increases. The 2010 survey is characterized by a relatively higher unemployment rate reaching 10%. In the quarters following 2010, France came back to the high unemployment situation of the beginning of 1998. In spite of the swings of macroeconomic unemployment, during the entire period, we observe the downward trend of the real wages of university and engineering-school graduates, while at the same time, the real wages of young workers with less than a high-school degree grew. There is no simple explanation of the evolution of real wages in terms of business-cycle fluctuations.

The jobs of educated workers are stable as compared to that of unskilled labor. It could still be true that the variation in real wages is caused by the business cycle because real wages are more flexible at these levels of education.[43] Yet, as we will see, other factors are likely to be responsible for (most of) the observed devaluation. We checked that variations of the unemployment rate do have a significant impact on real wages. Real wages are sticky and hence mildly procyclical. But the estimated returns to education by degree and cohort (or zero-experience wages), taking the 1998 high-school dropouts as a reference, are robust to the introduction of a control for variations of national unemployment.[44] These estimates are particularly stable for higher levels of educational attainment. Indeed, if we control for the variation of unemployment, we still find a drop of around 5% in the return of a 2-year Master's degree, relative to 1998.[45] Returns to experience also seem to be procyclical. We conclude that we should control for variations of the national unemployment rate, but that the devaluation phenomenon is not closely related to the business cycle.

G Online Appendix: Adding a Control for the Business Cycle

Table 18 displays a subset of estimated coefficients from a Mincer-type equation where log-wages are regressed on degree indicators interacting with gender and the cohort, plus terms where potential experience interacts with gender, degree and the cohort, and finally, on the rate of growth of the overall, macroeconomic unemployment rate (more precisely, the variation of the logarithm of the national rate of unemployment). We compare the results obtained with or without a control for the variation of the rate of unemployment.[46] A glance at Table 18 shows that the variation of unemployment, measuring the effect of the business cycle, has a weak impact on men's skill premia at zero experience. Yet, the unemployment variable has a significant impact on wages. These results how the robustness of the devaluation of university degrees emphasized above.

[43]In France, the minimum wage legislation is the main cause of rigidity at lower levels.

[44]In Online Appendix G, we show the result of a log-wage regression in which we control for the variation of the macroeconomic unemployment rate; see Table 18.

[45]See Online Appendix G, Table 18.

[46]This measure of unemployment variation is used here because it yields better results than the French GDP or the rate of variation of the French GDP.

Table 18: Returns to degrees at zero experience, with or without control for the variation of unemployment

Men	1998		2004		2010	
Control for Unemployment	No	Yes	No	Yes	No	Yes
Dropouts	.	.	0.103***	0.106***	0.0552***	0.0620***
	(.)	(.)	(0.0118)	(0.0118)	(0.0140)	(0.0141)
Vocational Degree	0.0212**	0.0216**	0.128***	0.131***	0.127***	0.133***
	(0.00725)	(0.00726)	(0.00818)	(0.00821)	(0.0109)	(0.0110)
High-School Degree	0.0546***	0.0549***	0.150***	0.153***	0.132***	0.138***
	(0.00830)	(0.00830)	(0.00778)	(0.00779)	(0.00891)	(0.00900)
Associate's	0.107***	0.107***	0.215***	0.218***	0.210***	0.216***
	(0.00745)	(0.00746)	(0.00799)	(0.00801)	(0.00862)	(0.00869)
3 years of College (L3)	0.259***	0.259***	0.219***	0.222***	0.262***	0.268***
	(0.0249)	(0.0250)	(0.0139)	(0.0139)	(0.0293)	(0.0293)
4 years of Colleges (M1)	0.271***	0.271***	0.244***	0.247***	0.304***	0.310***
	(0.0159)	(0.0159)	(0.0131)	(0.0131)	(0.0332)	(0.0332)
Master's (M2)	0.488***	0.489***	0.420***	0.422***	0.442***	0.448***
	(0.0201)	(0.0201)	(0.0151)	(0.0151)	(0.0125)	(0.0126)
Business Schools	0.467***	0.467***	0.596***	0.599***	0.495***	0.501***
	(0.0375)	(0.0376)	(0.0397)	(0.0398)	(0.0276)	(0.0275)
Engineering Schools	0.615***	0.615***	0.571***	0.573***	0.587***	0.593***
	(0.0168)	(0.0168)	(0.0153)	(0.0153)	(0.0147)	(0.0147)
Growth of Unemployment		-0.161***				
		(0.0330)				
Constant	7.029***	7.026***				
	(0.00562)	(0.00568)				
Observations	16,2452					

Note: Results obtained by OLS on pooled data stacking three Generation surveys 1998, 2004 and 2010. The dependent variable is the logarithm of the real wages of individuals with a full-time job. The table gives the coefficients and standard deviations of two regressions giving the returns to degrees at zero experience, with degree indicators interacted with cohort and gender, with or without control for the variation of overall unemployment. The 1998 high-school dropouts are the reference group. Potential experience interacted with cohort and degree dummies are introduced in these regressions, but their coefficients are not reported, to lighten the table. Stars indicate the significance of estimated coefficients; * for p-value < 0.1, ** for p-value < 0.05 et *** for p-value < 0.01. Regressions are weighted using Céreq's survey weights.

H Online Appendix. Differences in Employment Rates by Type

Our estimates of the employment equation exhibit a few other interesting properties. The probability of full-employment generally decreases with the cohort, and particularly for types 1 and 2. In a certain sense, this contributes to the devaluation of degrees. The probability of a higher rate of employment typically increases with educational achievement. The impact of effective experience on the probability of employment is always positive and significant, showing the existence of a virtuous circle of employment (employment today begets more employment in the future) and this effect is higher for the 2010 cohort than for older cohorts; it seems also stronger for type 2 than for types 1 and 3.

Can we provide an intuition for the reason why Type 2's jobs are so stable? A first reason has to do with a possible contrast between the public and private sectors (because public jobs are typically much more stable than private sector equivalents in France). Indeed, Table 19 shows that Type 2 is slightly more frequent in the public sector (85% of the Type 3 are in the private sector, as compared to only 78% for Type 2). So, this seems to be an important difference between Type 2 and Type 3.

Table 19: Proportion of individuals employed in the public sector, by type

	Overall	Type 1	Type 2	Type 3
1998				
Private	.81	.81	.78	.85
Public	.19	.18	.22	.15
2004				
Private	.83	.81	.83	.88
Public	.17	.19	.17	.11
2010				
Private	.80	.78	.78	.85
Public	.20	.22	.22	.15

Figure 12: Employment Rates by Type: Simulations

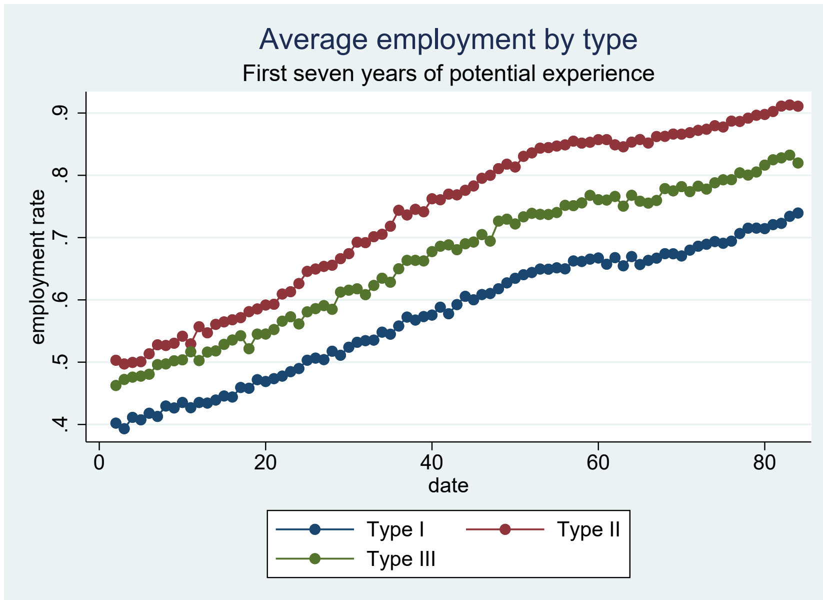


Table 20: Distribution of firm size (number of employees) conditional on type k

	Overall	Type 1	Type 2	Type 3
1998				
Small Firms $[1,9]$	.26	.34	.22	.19
Medium Firms $[10,49]$	.27	.31	.24	.24
Large Firms ≥ 50	.47	.36	.54	.57
2004				
Small Firms $[1,9]$	.26	.33	.23	.21
Medium Firms $[10,49]$	.28	.30	.28	.25
Large Firms ≥ 50	.46	.37	.49	.54
2010				
Small Firms $[1,9]$	.25	.32	.21	.20
Medium Firms $[10,49]$	.24	.25	.24	.21
Large Firms ≥ 50	.51	.43	.55	.59

Firm size could be an even more important source of job stability. The Generation Surveys give indications on the size of the employing firm during each employment spell. Table 20 shows that Type 1 is evenly distributed in the three size categories, while between 50 and 60% of Types 2 and 3 are in large firms. Table 20 does not show a real difference between Type 2 and Type 3 when it comes to firm size. We conclude that observable characteristics of jobs and employing firms help understanding the specific visible features of Type 2, but only to a certain extent. We'll come back below to the fact that types are not well explained by omitted controls.

Table 21: Average number of employment spells, by type

	Overall	Type 1	Type 2	Type 3
1998	2.95	3.12	2.68	3.07
2004	3.15	3.45	2.90	2.98
2010	2.98	3.17	2.69	3.02

We can check that the average number of employment spells of Type 2 is smaller than two while Types 1 and 3 have an average number of spells greater than three. This is shown on Table 21 and confirms the greater job stability of Type 2.

Finally, a look at simulated employment and experience paths clearly shows the differences between types. Simulations confirm that Type 2 has the highest employment rate, as shown by Figure 12.

I Impact of Family Background by Type

When the father is a professional, the probability of reaching a higher education level is significantly increased for all types — the probability of reaching the top levels is markedly increased.

On the role of an educated father (*i.e.*, "father is a professional") we can provide more indications. Table 22 gives the estimated percentage of individuals whose father is a professional (knowing that this category includes mainly educated fathers: teachers, engineers, doctors, executives etc.). Table 22 shows that the proportion of professional fathers is increasing with time (this is due to the fact that the years of education of the population are increasing). In addition, we see that educated fathers are more prevalent among Type-3 individuals. But the professional father is far from perfectly correlated or predicted by the type.

Table 22: Frequency of a professional father conditional on cohort and type

	Type 1	Type 2	Type 3
All cohorts	19.7%	21.1%	23.8%
1998	13.6%	16.8%	19.2%
2004	24.0%	23.6%	26.1%
2010	25.2%	27.1%	30.9%

Table 23: Probability of reaching an education level given the type, cohort and a professional father ($i.e.$, $z = 1$)

	Conditional on professional father, $p(h\|k, c, z = 1)$								
Conditional on cohort ...	1998			2004			2010		
and conditional on type ...	1	2	3	1	2	3	1	2	3
Less than High-school Degree	0.23	0.14	0.13	0.16	0.10	0.08	0.12	0.09	0.06
High-school Degree	0.28	0.19	0.22	0.27	0.15	0.14	0.29	0.16	0.16
Some College and Bachelors	0.35	0.41	0.44	0.36	0.37	0.38	0.28	0.38	0.26
Masters 2	0.07	0.06	0.07	0.14	0.15	0.22	0.14	0.22	0.27
Bus. Engin. School Degrees	0.07	0.20	0.14	0.07	0.22	0.18	0.17	0.14	0.26

Table 23 is particularly striking. It gives the estimated probabilities of reaching (choosing), the various educational levels, conditional on cohort, type and the fact that the 'father is a professional'. This Table is all more striking if we compare it to the unconditional equivalent, that is, Table 7. As time passes, the sons of professionals have been deserting the lowest educational levels and invading the highest levels, in particular, the Type 3s and, to a lesser extent, the Type 2s with a professional father, the most impressive "invasion" being that of Master programs by individuals with this family background.

J Online Appendix: Choice of the Number of Types K; Robustness

Until now, we estimated the model with three types, but the number of types is a choice that must be justified. The choice of a number of types K is in principle not easy. We devote the next subsection to this question. It happens that 3 types seems to be the right choice. The model has been estimated with panel data stacking three surveys. But would the results be different had we estimated the same model three times separately with the help of each survey? It happens that the results do not change much if we try this form of sub-sample estimation.

J.1 Number of Types

The question of the number of types is crucial because the set of types provides a model of the unobservable factors generating the well-known endogeneity problems: mainly the endogeneity of education and experience in the wage and employment equations.

The difficulty comes from the well-known fact that the log-likelihood of the model with K types, denoted $\mathcal{L}(K)$, is typically increasing and concave: an additional type will always lead to some improvement of $\mathcal{L}(K)$, but with decreasing marginal values. If K is too small, the types are themselves heterogenous melting pots of individuals. If K is too large, there is a risk that the types do not represent real individuals but are just improving the approximation of the distribution of wages, education and employment by a finite mixture of normal distributions. We know that, in essence, any distribution can be approximated by a mixture of normals, to any desired degree of precision, and in our case, a large K may simply be a form of over-fitting.

To choose the number of types K, we in fact combine several criteria. The usual criteria penalizing the likelihood for a high number of parameters, the Akaike and the Bayesian Information Criteria (resp. AIC and BIC, see Akaike (1974), Schwarz (1978)) will in principle reach a minimum for some value of K, but are not well adapted to the choice between K and $K + 1$.[47] AIC tends to overestimate the correct number of components (AIC pushes towards over-fitting). BIC corrects for these difficulties but tends to underestimate K.[48] These criteria are useful, but they do not

[47] If q is the number of parameters, N the number of observations and $\mathcal{L}$ is the log-likelihood, then $AIC = 2q - 2\mathcal{L}$ and $BIC = q\ln(N) - 2\mathcal{L}$.

[48] For references on these problems and the discussion of other information criteria, see Celeux and Soromenho (1996).

measure the quality of classification. So we use other criteria, based on *entropy* and penalizing the fact that types are difficult to distinguish.

An individual i is well-classified or well categorized as type k if $p_{ik} \simeq 1$. The quality of classification provided by the model is high if all (or most) individuals are well classified. When K increases, we often quickly reach a point at which the p_{ik} values are mostly far away from 1 and 0. Visual inspection, on Figure 2 shows that with our model, the quality of classification is good for $K = 3$.

To push the analysis further, we estimated the model for different values of K and looked at different criteria, including entropy, to choose the best model. The difficulty here is that the number of parameters (and time needed for estimation) quickly increases with K (it is already difficult to estimate our model with 4 types). Table 24 presents the values of different criteria when K varies from 1 to 4.

There exists a tension between Information and Entropy criteria. Celeux and Soromenho (1996) have proposed a choice criterion based on the notion of entropy, called the *Normalized Entropy Criterion*, or NEC. In our context, entropy $\mathcal{E}$ must be defined as follows,

$$\mathcal{E}(K) = -\sum_{i=1}^{N}\sum_{k=1}^{K} \hat{p}_{ik} \ln(\hat{p}_{ik}), \tag{30}$$

where $\hat{p}_{ik}$ is the estimated value of the posterior probability p_{ik}. It is easy to check that $\mathcal{E}(1) = 0$ and $0 \leq \mathcal{E}(K) \leq N \ln(K)$, where N is the number of observations i.[49] Entropy is minimal (and equal to zero) when partitioning is perfect.[50] We can divide entropy by its maximum value to obtain an index taking values in $[0, 1]$. Define $\mathfrak{E}(K) = (N \ln(K))^{-1}\mathcal{E}(K)$. This index should be minimized.

To define the NEC, we consider the gains, in terms of the Log-Likelihood, with respect to $K = 1$, that is $\mathcal{L}(K) - \mathcal{L}(1)$. Entropy is now divided by this gain. NEC is defined as follows,

$$\text{NEC}(K) = \frac{\mathcal{E}(K)}{\mathcal{L}(K) - \mathcal{L}(1)}. \tag{31}$$

Another simple criterion that measures the quality of classification is the Average Hirschman-Herfindahl Index. This index is defined as follows,

$$H(K) = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} \hat{p}_{ik}^2 \tag{32}$$

Note that H is equal to 1 if all observations i are perfectly classified. In addition we have, $1/K \leq H(K) \leq 1$. It follows that the lower bound of H is decreasing with K.[51] A normalized index can be

[49] The entropy is maximal when $p_{itk} = 1/K$ for all k and all i. Entropy is maximal when types cannot be distinguished because any observation can belong to every group with the same probability $1/K$.

[50] Indeed, if for all i, there exists a type $k = k(i)$ such that $p_{ik} = 1$, then, $\mathcal{E}(K) = 0$.

[51] On the use of H in classification problems, see Windham and Cutler (1992).

constructed as follows. For $K > 1$, define $\mathfrak{H}(K) = (K.H(K) - 1)/(K - 1)$. We have $0 \leq \mathfrak{H}(K) \leq 1$. H and $\mathfrak{H}$ may increase with K ; if these indices drop, this is because the quality of classification deteriorates as K increases.

Table 24: Selection Criteria for the Number of Types

Criterion	1 type	2 types	3 types	4 types
Number of parameters	85	158	231	304
Log-Likelihood $\mathcal{L}(K)$	-167,263	-150,893	-143,745	-141,210
$\mathcal{L}(K) - \mathcal{L}(1)$	0	16,370	23,517	26,053
Adj. R^2 of wage regression	.402	.563	.585	.635
AIC	334,696	302,102	287,952	283.028
BIC	335,351	303,319	289,732	285,370
Average Herfindahl (H)	-	0.89	0.84	0.79
Normalized Herfindahl ($\mathfrak{H}$)	-	0.78	0.76	0.72
Entropy $\mathcal{E}$	-	2825	4391	6160
$\mathfrak{E}$	-	0.2484	0.2436	0.2708
NEC	-	0.172	0.186	0.236
Individuals N	16,404	16,404	16,404	16,404

The figures of Table 24 are derived from EM estimations of the full model with $K = 1, 2, 3$ and 4 types.

The most important information shown by Table 24 is that the Log-Likelihood increases markedly until $K = 3$. The marginal gain of adding a fourth type is clearly smaller. So, three types seems a reasonable choice at first glance. A difficulty is that AIC and BIC are always decreasing — they probably reach a minimum for $K > 4$ — but lead to the same conclusion that $K = 3$ is reasonable. The Average Herfindahl and Normalized Herfindahl indices suggest $K = 2$ as the best choice. The normalized entropy $\mathfrak{E}$ clearly indicates $K = 3$, while Celeux and Soromenho's NEC indicates $K = 2$, but NEC doesn't increase much between $K = 2$ and $K = 3$ while it increases a lot more between $K = 3$ and $K = 4$. We therefore choose $K = 3$ as our compromise: not too many parameters, a good classification of individuals and the gains if $K \geq 4$ are apparently small.

J.2 Estimation of the model by cohorts separately

The model presented above has been estimated with a sample stacking three cohorts of males. The model is very flexible in the sense that most parameters vary by cohort and by type. For this reason, in a nutshell, we find very similar results when the model is estimated with three types on each of the three cohorts separately. Table 25 very clearly shows that the classification of individuals in three types is very stable to the extent that we find a closely related classification if we estimate the model in a single cohort. Table 25 gives the correlation matrices of the p_{ik} estimated in the three-cohort model with the estimated p_{ik} obtained in a three-type version of the same model, estimated on a single cohort. The structure of these correlation matrices, with a positive diagonal and high coefficients around .9 and negative off-diagonal values show that our three-type structure does not strongly depend on the fact that we stacked three cohorts. In addition, the full estimation results on the three cohorts taken separately do not show big differences.[52] This is reassuring, because one could have suspected that the structure of the economy has changed with time in a manner that our variables do not explain well. Yet, the three-cohort model is very flexible with most coefficients depending on the cohort: this very flexibility probably explains that subsample estimation does not lead to markedly different results.

Table 25: Correlation coefficients of the posterior probabilities of types p_{ik} estimated in a model with three cohorts, with the corresponding probabilities estimated in a model estimated with a single cohort

	Columns: three-cohort model types								
Rows: single-cohort	1998			2004			2010		
model types	1	2	3	1	2	3	1	2	3
1	0.93	−0.54	−0.48	0.73	−0.35	−0.50	0.94	−0.61	−0.47
2	−0.49	0.84	−0.36	−0.40	0.65	−0.25	−0.47	0.90	−0.42
3	−0.53	−0.27	0.91	−0.44	−0.35	0.94	−0.56	−0.28	0.97

K Online Appendix: Construction of the Sample

In this work we exploit the CEREQ surveys called *Enquêtes Generations à 7 ans*, from 1998, 2004 and 2010. The surveys provide observations during the first 7 years of career of a large representative sample (*i.e.*, cohort) of individuals. The sample includes only individuals who left the educational system during the first survey year (*i.e.*, 1998, 2004 or 2010) and did not return to education during the 7 years of the observation period, except maybe for short on-the-job training sessions. Each of the three stacked surveys contains 3 files: *employment spells, non employment spells and individual characteristics*, the three files form a dataset containing the sequence of employment and unemployment (or non employment) spells for each individual during 7 years

Changes in working hours during employment spells are described. In 1998, the employment-spells dataset contains 47,936 observations, the unemployment dataset contains 30,329 observations and

[52] The complete cohort-by-cohort results are available upon request.

the individuals' file contains 16,040 observations. The corresponding figures are 39,101, 22,724, and 12,365 in the 2004 survey; these figures are respectively 26,056, 16,467, and 8,882 in the 2010 survey.

In each survey, we start by removing the employment spells that are labelled as *family help (i.e., aide familial or afa), self-employed (i.e., à son compte or asc), undescribed summer jobs (i.e., vac).* This amounts to removing 3,148 employment spells in 1998, 3,572 employment spells in 2004, 2,076 employment spells in 2010. It follows that an individual who is always self-employed (or categorized as *afa*, or *vac*) in the first 7 years after having left the educational system disappears from the data. Then, we merge the employment and non-employment data sets: each individual's history appears with a sequence of employment and non-employment spells. In 1998 we have 75,117 spells, in 2004 58,253 spells, in 2010 40,467 spells.

Individuals are interviewed at the end of their 3rd, 5th and 7th year. They are asked to describe their recent history and their situation at the very moment of the call. So, for each individual, we have 3 additional observations that are the description of their situation at the month of the interview. We recover this information from the 3rd and 5th year of each cohort (*i.e.*, survey) for the individuals observed at the end of the 7th year and we add these data to the 7th year survey. This increases the number of point observations in each cohort, that, at this point are: 29,986 in 1998, 23,011 in 2004, 16,153 in 2010.

We deleted the employment spells that lack the working time information; as a consequence, we lose 413 observations in 1998, 66 observations in 2004 and 1,536 observations in 2010.

At this point the beginning and the end of each spell plus the observations at the time of the survey are kept as observations of the individual. Each row of the database becomes an observation (i, t) in the labor market of an individual i (either employed or not), at a date t. At this point the number of observations are : 171,258 in 1998, 133,211 in 2004, 91,174 in 2010.

Each individual enters the dataset the month after the end of his(her) education. There is a date system for each cohort. *Beginning* is the date when an individual in the cohort can be first observed, while *End* is the date of the last observation of the dataset:

- Cohort 1998. Beginning: 1 = January 1998; End: 96=December 2005.
- Cohort 2004. Beginning: 1 = November 2003; End: 98 = December 2011.
- Cohort 2010. Beginning: 1 = November 2009; End: 98 = December 2017.

At this point, the dataset can be described as follows:

- Cohort 1998: 15,950 individuals that are observed on average 10.74 times;
  (minimum 1, 1st quartile 6; median 10; 3rd quartile 14; maximum 54)

- Cohort 2004: 12,233 individuals that are observed on average 10.89 times; (minimum 1, 1st quartile 6; median 10; 3rd quartile 14; maximum 63)
- Cohort 2010: 8,774 individuals that are observed on average 10.39 times; (minimum 1, 1st quartile 6; median 9; 3rd quartile 13; maximum 45)

Then, we build the experience variable as the sum of working time up to time $t-1$. For each spell we add the information regarding the accumulated experience at time $t-1$ at the beginning, and the end of the spell.

Now, using the individual dataset we create the variables: father is a professional, place of residence at grade 6 entry and the education level (*i.e.*, degree category). The detail for these variables, for each cohort, can be found in the tables below.

The real salary is computed in July 2013 euros.

Then, we remove individuals lacking an observation of the father's occupation and of the residence at grade 6 entry. This leads us to delete 1,071 individuals in the 1998 cohort, 647 individuals in the 2004 cohort and 856 individuals in the 2010 cohort. *Finally, we take the subset of males.* The final dataset for each cohort includes 16,404 individuals, among which:

- Cohort 1998: 80,006 observations for 7,383 individuals;
- Cohort 2004: 60,907 observations for 5,500 individuals;
- Cohort 2010: 37,489 observations for 3,521 individuals.

We stack the three cohorts and generate a unique dataset. We generate a cohort variable c taking values 1998, 2004 or 2010, and a common calendar for the three cohorts where $1 =$ January 1998 and $240 =$ December 2017.

Table 26 lists the degree types that have been aggregated in each of the categories used for estimation.

Table 26: Aggregation of Degrees

Education level	Education level detail
	1998 Cohort
Less than High School	SEGPA, reached grades 7 to 11, first year of CAP or BEP, CAP without degree, BEP without degree, CAP, BEP,

Continues on the next page...

	MC post CAP-BEP, Bac Pro without degree, Brevet or Bac techno without degree, finished grade 12 without degree
High-School Degree	Bac Pro, Bac techno, Bac général, 2 years of College without degree BTS or DUT without degree
Some College, Bachelors	DEUG, BTS DUT, Bac + 3, Bac + 4 IUFM : admitted, IUFM : not admitted
Masters	Bac + 5 and more Excluded: Doctorate and advanced medical degrees
Bus. and Engin. Sch. Deg.	Business Schools, Engineering schools
	2004 Cohort
Less than High School	without degree, CAP, BEP, MC
High-School Degree	Bac pro, Bac techno, Bac général
Some College, Bachelors	Bac+2, DEUG Licence pro, L3, M1
Masters	M2 Humanities, Business adm., Law, M2 Maths, Sciences, Technology, Health, Physical education
Bus. and Engin. Sch. Deg.	Business Schools, Engineering schools
	2010 Cohort
Less than High School	Without degree, CAP, BEP, MC
High-School Degree	Bac Pro, Brevet de Technicien, Brevet Professionnel Bac Techno, Bac général
Some College, Bachelors	BTS or DUT other Bac+2 Bac+2/3, Licence pro L3, other Bac+3 M1, Bac+4
Masters	M2 Humanities Business adm. Law

Continues on the next page...

... table 26 (continued)

	M2 Maths Sciences Technology
	other Bac+5
Bus. and Engin. Sch. Deg.	
	Bac+5 Business Schools, Engineering schools

Note. Without degree *i.e., Non-diplômé* means without the diploma or certificate: students who studied but were never granted the degree. Bac is shorthand for baccalauréat (high-school graduation). *Bac pro* means baccalauréat professionnel. *Bac techno* means baccalauréat technologique. Both categories are vocational versions of terminal high-school degrees. CAP and BEP and MC (*i.e.*, mention complémentaire) are pre-bac vocational certificates. *Brevet* is a certificate typically obtained at the end of grade 9. DEUG means two successful years of College. DUT and BTS are vocational degrees, equivalent to the American associate degree. L3 is a Bachelor (three years of College). *Licence pro* is a three-year higher-education vocational degree. The IUFM are preparation schools for primary school-teachers. SEGPA means special education for students with difficulties (grades 6-9).

Table 27 gives, for each cohorts the definition of the Urban, Peri-Urban and Rural areas used to construct the corresponding indicators. A difficulty comes from the fact that exact definitions changed with the years, but the classification of cities and towns has not changed much.

Table 28 lists the occupation categories included (and not included) in the definition of the dummy variable called “Father is a professional”.

Table 27: Area of residence at grade 6 entry

Residence area	Residence area, detail
	1988 Cohort
Urban area	municipality belonging to an urban cluster
Peri-urban	municipality belonging to a peri-urban, outer suburban zone
Rural area	Municipalities belonging to a rural-zone labor market
	Other localities of rural zones
	Municipality belonging to the periphery of a rural labor market
	Ultramarine Municipalities (West Indies, etc.)
	Foreigner, Unknown
	2004 Cohort
Urban area	Urban cluster
Peri-urban	Mono-polarised Municipality
Rural area	Multi-polarised Municipality, Rural space
	2010 Cohort
Urban area	Large urban areas (more than 10 000 jobs),
	Intermediate urban areas (5 000 to 10 000 jobs)
Peri-urban	Periphery of large and intermediate urban areas
Rural area	Multi-polarized Municipalities in large urban areas,
	Small clusters (less than 5 000 jobs), Periphery of small clusters,
	Other Multi-polarized Municipalities,
	Isolated communes out of the influence of clusters
	Foreign, Ultramarine communes

Table 28: Occupation of the Father

Occupation of the Father	Occupation of the Father, detail
Not a "professional"	Farmer, Craftsman, Storekeeper, Entrepreneur,
	Technician, Foreman, Salesman, Associate professional,
	White collar worker, Blue collar worker, unknown
Father is a "professional"	Executive, Engineer, Learned profession, Professor

Note: "Professional" here is a category including the French *professions intellectuelles supérieures*, typically requiring an advanced higher-education degree.