I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18053

Is There a Devaluation of Degrees? Unobserved Heterogeneity in Returns to Education and Early Experience

Damiano Argan
Robert Gary-Bobo
Marion Goussé

JULY 2025

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18053

Is There a Devaluation of Degrees? Unobserved Heterogeneity in Returns to Education and Early Experience

Damiano Argan
LUISS

Robert Gary-Bobo
Université Paris 1 Panthéon-Sorbonne

Marion Goussé
CREST-ENSAI and IZA

JULY 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Is There a Devaluation of Degrees? Unobserved Heterogeneity in Returns to Education and Early Experience*

We study how the returns to higher education evolved in France during a period of educational expansion. We study possible changes in the mix of unobservable characteristics of the graduate population. Using a finite mixture model with latent types, we estimate type-specific log-wage, experience accumulation, and education-choice equations. We find that expected real wages declined for higher-education degrees, and that this result is not driven by adverse selection. Returns to education and experience decreased for certain unobserved types but increased for others. The composition of types among Master's graduates suggests improved student selection over time, despite rising graduate numbers.

JEL Classification: C33, I21, I24, I26, J22, J24, J31

Keywords: wages, returns to education, returns to experience, human capital, selection, unobserved heterogeneity, finite-mixture models, latent types

Corresponding author:
Marion Goussé
CREST-ENSAI
51 Rue Blaise Pascal
35172 Bruz
France

E-mail: marion.gousse@gmail.com

* Research supported by CREST and Labex Ecodec (LabEx ECODEC; ANR-11-LABX-0047) and CEPREMAP, Paris. This article is also part of project ISEProD (Industrial Structure and the European Productivity Growth Divergence) that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (awarded to Fabiano Schivardi, grant agreement No 835201). D. Argan thanks the participants of the EUI Microeconometrics Working Group for their useful comments, Andrea Mattozzi, Andrea Ichino, and Fabiano Schivardi for their support and advice, and Anatole Cheysson for being a constant source of insightful discussion. The authors thank Stéphane Bonhomme, Arnaud Maurel and Christian Belzil for discussions and Jean-Marc Robin, for inspiration.

1 Introduction

In the recent decades, university and college enrollment has surged in many countries. The global number of students pursuing tertiary education more than doubled between 2000 and 2020 (UNESCO, 2022; OECD, 2025). It is a common point of contention that this growth might lead to an excess supply of graduates, potentially reducing the real wages of college graduates relative to those of high-school graduates, *i.e.*, causing a drop in the "college premium". Indeed, a decline in returns to education is observed and studied in many countries.[1] When a decline occurs, it is critical to disentangle the possible contributing factors, including an excess supply of graduates, potential decline in teaching quality and changes in student selection criteria. In this paper, we examine the evolution over time of returns to education among young French workers during a period of educational expansion.[2]

When individual returns to education are heterogeneous, the population of graduates can be represented as a mix of unobserved student types. Changes in the average wage difference between education levels, and across cohorts, are uninformative of changes in the average treatment effect of education over time, because the unobserved mix of student types may change over time. These changes may be particularly important during periods of higher education expansion. Similarly, differences in IV estimates of the return to education across cohorts — assuming the same instrument is available for different cohorts — also reflect changes in the set of compliers and their returns to education over time.[3] In this paper, we study changes in the average treatment effect (hereafter ATE) of education over time using a model that accounts for unobserved heterogeneity

[1] See Valletta (2018), Emmons et al. (2019), Banh et al. (2024), Hanushek et al. (2023). Decline in returns to unobserved skills is studied in Lochner et al. (2025).

[2] Nimier-David (2023) documents the large expansion of higher education in France from 1990 to 2003.

[3] In the presence of heterogeneous responses to treatment, Heckman and Vytlacil (2005) write: "Two economists analyzing the same dataset but using different instruments will estimate different parameters that have different economic interpretations" and "the cure may be worse than the disease," when it comes to correcting for least squares bias with the help of instruments. The interpretation of IV estimates is even more delicate if we want to explore changes in treatment effects across several cohorts of individuals.

by means of a system of latent types.

Our analysis reveals that certain higher-education degrees, particularly Master's degrees, yielded lower real wages on average during the 2010–2017 period compared to 1998–2005. We find a 25% decline in the discounted expected real earnings of Master's degree holders over the first seven years of their careers, for a representative sample of the population (*i.e.*, a 25% drop in the ATE). Assessing variation in the return to a Master's degree based on average wage comparisons over time would have led to a significant underestimation of the drop in the return to education, and comparing the ratio of average wages between the educated and non-educated would have further increased the bias in the estimation. The reason is that the underlying composition of the graduate population changed significantly over time. We show that student selection at the Master's level has actually improved, while it has worsened at lower education levels. We find that the observed decrease in returns to Master's degrees is not attributable to a decline in student quality (*i.e.*, adverse selection), despite the growth in enrollment. The most likely explanation for the devaluation is therefore an excess supply of graduates.

We use a finite mixture model to treat the endogenous character of education choices, employment and effective experience. We assume that the endogeneity problems, typically arising in standard econometric regressions such as the Mincer equation, are entirely driven by the unobserved types. In other words, error terms are assumed conditionally independent of explanatory variables, *knowing the type*.[4] In essence, we estimate a model explaining individual wages, individual employment rates and education choices simultaneously with the help of panel data. Each latent 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.[5] The model is the product of three

[4]With a sufficiently large number of types, it is possible to approximate any distribution of wages and other dependent variables.

[5]In a variant used to check for robustness, we add firm effects and an equation describing the matching of

finite-mixture models for respectively, wages, employment and education, with a common set of latent types. The 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. The model is estimated by straightforward likelihood maximization.

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, in line with Heckman et al. (2018) who 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.

We show that our estimation method leads to the identification of a system of types that is stable and parsimonious in the following sense: (i) the types are estimated on a dataset that stacks several cohorts of students; (ii) if we reestimate the model on the cohorts taken separately, we find that the same types appear — in fact, we find types that are highly correlated with the original ones; (iii) the type system resists other robustness checks, like the introduction of firm classes; (iv) we find that three types do a good job at approximating unobserved student diversity.

An important output of finite-mixture models is the probability of belonging to a given group (*i.e.*, type), conditional on the individual's observed characteristics (hereafter the *individual posterior probabilities of types*). Our model's estimated coefficients and the posterior probabilities of types allow us to compute a number of *policy-relevant 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 *average treatment effect on the*

workers to a finite set of firm classes.

treated (ATT) and the ATE of a certain level of education. The ATE of a given education level is the average return to education obtained when the distribution of types in the sub-population achieving this level is the same as in the general population. Because of selection, ATE is a counterfactual value of average returns. In contrast, ATT is a computation of the average return, taking the mix of types in the specific subpopulation of graduates into account. It follows that the difference between ATE and ATT for a certain degree is a measure of selection for this degree.[6] In addition, our estimation results allow us to compute ATEs conditional on any latent type and hence to uncover the heterogeneity of returns to education and experience across types.

We estimate the model on a rich panel of young workers, the *Generation surveys*, and we focus on the subsample of men.[7] In these data, we follow the first seven years of career of three cohorts of French workers. Each cohort is defined by the year during which the worker left the educational system, namely, 1998, 2004 and 2010.

We can summarize some of our results as follows. The three types that we find are easily interpretable: there is an obvious ranking of types in terms of returns to education and experience. We can compute the ATE of degrees in terms of discounted expected real earnings over the first seven years of a career, and, for instance, we find a decrease in the ATE of Master's degrees of the order of 25% between the cohorts 1998–2005 and 2010–2017. Discounted expected real earnings is a synthetic measure of human capital that encompasses wages, employment rates, and accumulated experience. Its decrease for Master's graduates can be decomposed into a 12% drop in starting wages, a 9% drop in the annual return to experience, and a 7-percentage-point lower employment rate over the first seven years of a career. The corresponding decline in discounted expected real

[6]If ATT is greater than ATE for a given degree, we say that students are positively selected into the degree (and negatively selected if ATE is larger than ATT).

[7]A correct treatment of the earnings and choices of women would be more complicated. Yet, we know that the French women with full-time jobs observed a similar decrease in the return to Master than men (see, Argan and Gary-Bobo (2023)). We plan to include women in future research.

earnings based on observed outcomes (ATT) is about 11% over the same period. When we compute the difference in ATEs between a Master degree and a high-school diploma, the difference shows a drop of 37 percentage points overtime — or a 7.2 percentage point decrease in the return to one additional year of education (*i.e.*, from 16.4% in 1998 to 9.2% in 2010). The corresponding figure using the ATT is a decrease of less than 3 percentage points (a 0.6 percentage point drop in the gain from an additional year of education). There is a greater divergence between ATE and ATT when assessed through differences in earnings across education levels because the mix of student types changes over time at both the lower and higher ends of the educational ladder. Our approach allows us to assess changes in the composition of student types and their statistical significance. Individuals who dropped out of education at the lower end of the educational ladder were, in 2010, more negatively selected with respect to the general population than in 1998, while Master's graduates, who were not positively selected in a significant way relative to the rest of the population in 1998 became positively selected in 2010. We can study the determinants of this improved selection in Master's programs and show that it is partly due to an increase in the probability of graduating of "top types" coming from low socioeconomic backgrounds. Finally, our study shows that returns to education and experience are strongly heterogeneous and depend on unobservable characteristics — the latent types in our model. Depending on the type, the dynamics over time can be very different: in the case of the Master's degree, the "top type" experienced an increase in its average discounted earnings, while the other two types experienced a large drop.

To improve our understanding of the degree devaluation phenomenon, we estimated a variant of the model with a finite set of firm classes, a type-dependent equation describing the matching of workers to firms, and indicators of firm classes added as explanatory variables in the wage equation. The results show that the wage premia commanded by employment in top firm classes decreased with time, explaining a substantial part of the observed devaluation. The increased number of graduates led to a congestion effect, more students competing for jobs in the best firms led to a downward adjustment of wages, due to some excess supply of graduates. Hence, we conclude that the observed absolute devaluation of University Master's degrees, in France, is most likely due to an excess supply of graduates, because the selection of students has improved with time. In contrast, we find that in the French business and engineering schools, the enrollment of which has also grown substantially, the quality of student selection has decreased with time.

Literature. 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 (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 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. 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 et al. (2008), Acemoglu and Autor (2011)). In the UK, Blundell et al. (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.

Another literature argues that education expansion also changes the selection threshold for students. Ichino et al. (2024) study the higher-education expansion in the UK from 1960 to 2004 with the help of a general equilibrium Roy model. They find that 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. Also 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.[8] Belzil and Hansen (2020) reach similar conclusions, comparing the 1979 and 1997 cohorts of the NLSY survey, using structural econometric methods. Ashworth et al. (2021) use the same NLSY data, and study closely related questions with the help of a structural model with a latent factor structure.

Until the turn of the millennium, 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 et al. (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

[8]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 et al. (2011), who estimate the *marginal treatment effect* of College.

College graduates have flattened, the starting wages went down, and these facts cannot simply be explained by the business cycle. The situation of France is similar. In a recent paper, Corblet (2024) also argues that French higher-education expansion has caused occupational congestion whereby the share of higher education graduates employed in routine occupations rose, flattening their wage profiles. Her analysis, using the same data as ours, complements our findings.

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

For a general treatment of finite mixture models, see McLachlan and Peel (2000), Bouveyron et al. (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 a few researchers. See, in particular, Beffy et al. (2012), Arcidiacono et al. (2025); Gary-Bobo et al. (2016); Cassagneau-Francis et al. (2021), and Cassagneau-Francis (2021).

Our approach is sometimes called *soft classification*, in fact, this is probabilistic classification, because we only estimate the posterior probability of types for each observed individual. Models of probabilistic classification relying on a finite number of types can typically be estimated by Maximum Likelihood or EM algorithms.[9] There exists another, closely related approach, sometimes dubbed *hard classification*, where the set of observed individuals is partitioned in a number of

[9]See Appendix D for a discussion of identification properties.

groups (Bonhomme and Manresa (2015)).[10] An important advantage of both the 'soft' and 'hard' classification approaches is to reduce the number of parameters by grouping individuals, thus avoiding the incidental parameter bias. Both approaches are very flexible; both are in principle asymptotically equivalent under some technical assumptions.

In the following, Section 2 describes the data and some stylized facts. Section 3 presents the model. Section 4 discusses the computation of policy-relevant parameters. Section 5 presents the estimation results for the base model. Section 6 focuses on the evolution of returns to degrees and experience and presents variants of the ATE and ATT of education, allowing for a discussion of changes in selection. Section 7 presents the estimation of a variant of the model with firm effects and a treatment of firm-worker matching. We conclude in Section 8.

2 Data and Stylized Facts

2.1 The CEREQ Generation Surveys

We work with samples of young French workers called the Generation Surveys (*i.e.*, *Enquêtes Génération*), produced by CEREQ[11], a French public institution. Since 1992, every 5 years, the CEREQ draws a large sample of individuals who all left the educational system during the same year, with a representative variety of educational achievement levels. We will consider the CEREQ Generation surveys of 1998, 2004 and 2010.[12] In each of the three surveys, the workers are followed during 7 years; the labor market experience of each worker is tracked, month by month, by means

[10]This alternative approach relies on two-step estimators. In the first step, individuals are classified (*i.e.*, partitioned) with the help of a k-means algorithm. For instance, individuals can be separated by means of a vector of moments, based on observable characteristics. In a second step, the model equations are estimated, with a possibly different regression function for each class. Bonhomme and Manresa (2015); Bonhomme et al. (2022) study the statistical properties of associated estimators. Bonhomme et al. (2019) present an application to the study of wages with both firm and worker fixed effects, taking mobility into account. For a related, but different approach, see also Lentz et al. (2023).

[11]CEREQ (*i.e., Centre d'Etudes et de Recherches sur les Qualifications*), see https://www.cereq.fr.

[12]The first survey, launched in 1992, has a slightly different structure. For results obtained with these data, see Brodaty et al. (2014).

of interviews.[13] 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 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). Wages are given as monthly nominal salaries, including bonuses, net of compulsory social security and medical-insurance contributions, but gross of the income tax.[14] To sum up, we observe employment variables for a sample of individuals every month from 1998 until 2005, from 2004 to 2011 and from 2010 to 2017, but we observe wages only at some dates: at the endpoints of employment spells and at the moment of interviews. We first stack the three *Generation* surveys of 1998, 2004 and 2010 and we only keep men. Descriptive statistics of our sample are presented in Appendix B; additional details on sample construction are presented in Appendix A.

2.2 French Education Levels

We aggregate the highest degrees obtained by individuals in 5 categories: (1) *Below High-School Degree*, including dropouts without any certificate and secondary vocational certificates; (2) *High-School Degrees*, including all students who obtained their high-school diploma (*i.e.*, in French, *baccalauréat*); (3) *Some College and Bachelors* includes all the students whose highest achievement is the equivalent of an associate's degree or all vocational degrees requiring less than 3 years of study, plus all the bachelors; (4) *Master Degrees*, includes the degree of a two-year graduate program following the bachelor and requiring 5 years of study in total; (5) *Business and Engineering Schools*, includes degrees requiring 5 years of study and granted by French engineering and business

[13] Sampled individuals are called repeatedly and questions are asked after 3 years, 5 years and 7 years.

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

schools. The latter are independent institutions, separated from universities. Some are public, some are private (mainly non-profit) institutions. The best such schools deliver a degree after five years of study, but the 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 generally higher in schools than in universities.[15]

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 (see SIES (2022)). 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. Table 9 in Appendix B shows the evolution of the number of French students, enrolled in various institutions, across Generation surveys, from 1992 until 2017. The 'big push' of Master programs is very visible.

2.3 Observed Devaluation of Degrees.

We define the devaluation of a degree as an absolute decrease in the expected real salary of workers conditional on holding the degree.[16] 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.

[15]Some engineering or business schools admit students directly after high-school graduation. Business schools recently developed 3-year bachelor programs that are less selective, enrolling students directly after high-school.

[16]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.

Table 1: 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 stack the surveys of 1998, 2004 and 2010 and we regress observed log-wages on cohort dummies. Table 1 gives the results of 5 simple sub-sample regressions, one for each education level, of log-wages on dummies indicating the cohort.[17] 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.

We then examine differences in work experience between cohorts. These differences might explain

[17]We limit ourselves to full-time wages to avoid possible errors in the reporting of part-time work and part-time wages, as we do not observe the exact number of hours.

the diminishing returns to education. We measure effective experience in months at time t as the sum of all months worked from the time the individual left school until period t. For example, each month, we add one month to effective experience if the individual works full-time, and we add one fifth of a month if he or she works part-time at 20%. Table 2 shows the average experience accumulation in months, seven years after leaving education by cohort and level of education. It shows that workers who left school in 2010 accumulate experience at a lower pace than in the previous cohorts. In particular, workers with no diploma who left school in 2010 have barely accumulated 48 months of experience 7 years after graduation whereas similar workers who left school in 1998 had accumulated 66 months of experience.[18] Yet Masters and School students have barely lost experience across cohort which means it cannot explain the observe devaluation of degrees for these diplomas.

Table 2: Experience Accumulation across Cohorts and Education

	< High School	High School	Some College	Master	B & E schools
Cohort 1998	66.3	70.7	71.7	71.4	75.2
Cohort 2004	62.9	70.2	73.3	75.8	78.5
Cohort 2010	47.7	59.6	69.6	69.6	77.7

Note. Results obtained with pooled data stacking the 7-year Generation surveys of 1998, 2004 and 2010, considering males only. Average number of months of experience accumulated during the 7 years after leaving education.

We will now 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. Potential experience is defined as the number of months elapsed since the individual left the educational system. 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.

[18]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 in the beginning month. Besides, individuals also take a variable number of months to find a first job, so that young workers accumulate different amounts of potential experience during the period covered by the survey.

We regress log-wages on experience in linear form. We denote c_h the coefficient of experience for education level h. This coefficient gives the monthly returns to experience. To obtain yearly returns, denoted γ_h, we use the formula, $\gamma_h = (1 + c_h)^{12} - 1$. All c_h coefficients are significant at the 0.1 percent level. Table 3 gives coefficients γ_h, where c_h is estimated with two different methods: by OLS on pooled data and by the fixed-effects, *within*-student estimator (FE). We first observe that returns to experience are substantial, with values ranging from 2% to 7% 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 3 is that returns to experience fell between 1998 to 2010, and they fell more for the highest degrees of attainment.

These preliminary results show that there is a devaluation of higher-education degrees between 1998 and 2010.[19] This could be due to an excess supply of graduates, but there are competing explanations. The value of the degrees under scrutiny depends on two other factors at least: the selection of student skills and the quality of education. Both factors contribute to the graduates' human capital, and therefore to productivity and wages. The average talent or productivity of students enrolled in advanced programs might have gone down in the recent years. The quality of the teaching might also have decreased with time, and the two phenomena probably go hand in hand. Standard econometric methods are insufficient to analyze changes in student selection, in higher education, during the past twenty years. To push the investigation further, we therefore propose a model of unobserved heterogeneity.

[19]These results are also exposed in Argan and Gary-Bobo (2023)— this article is written in French.

Table 3: Yearly Returns to Potential and Effective Experience

MEN		Potential Experience		Effective Experience	
$\gamma_h = (1 + c_h)^{12} - 1$		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_h coefficients are significant at the 1% level.

3 The Model

To model the beginning of careers of three cohorts of young men under unobservable heterogeneity, we assume that unobserved heterogeneity is 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 $c \in C = \{1998, 2004, 2010\}$. In each cohort, we follow individuals across time from the moment they leave school to the moment of the survey seven years later. Individuals are indexed by i, with $i = 1, \ldots, N$. 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. Let h index the highest level of education reached by the individual with $h = 1, \ldots, H$. Let Z_{it} denote a vector of observed characteristics of i at time t.

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.[20] 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 $t > 1$,

$$x_{it} = \sum_{\tau=1}^{t-1} e_{i\tau}, \quad \text{with} \quad x_{i1} = 0. \tag{1}$$

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

We can now specify the 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. For $t \in T_i$ and for an individual i of type k, of education level h, from cohort c, we set

$$w_{itk} = \alpha_{0k} + \delta_{0kc} + \gamma_{0kch} + \beta_{0kch} x_{it} + \eta_{0k} Z_{it} + \epsilon_{itk}, \tag{2}$$

where ϵ_{itk} is a normal error term with a zero mean and variance σ^2_{wk}, where $(\alpha_{0k}, \beta_{0kch}, \gamma_{0kch}, \delta_{0kc}, \eta_{0k})_{h=1,.,H,k=1,.,K,c\in C}$ is a vector of parameters. Note that the model is very flexible insofar as all the parameters of the wage equation are free to vary with type k. Parameter α_{0k} is the constant type effect, parameter δ_{0kc} is the cohort effect varying by type, parameter

[20] We used the CPI from the National Statistical Institute (INSEE). Wages are expressed in 2013 euros.

γ_{0kch} is the return to education varying across education levels, types and cohorts and parameter β_{0kch} is the return to experience varying with type, education and cohort. Finally, η_{0k} are the effects of control variables assumed constant across cohorts and education levels but varying across types. In particular, the vector of variables Z_{it} includes the student's location when he was 11 years old, indicated by dummies (Urban, Peri-Urban and Rural), the father's occupation (the father-is-a-professional dummy) and the macroeconomic unemployment rate.

Given this, the expression for the observed wage of individual i at period t is,

$$w_{it} = \sum_{k=1}^{K} \theta_k(i) w_{itk}, \tag{3}$$

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 employment levels, denoted $\mathbf{e}_g$, with $g = 1, .., G$ and $1 \geq \mathbf{e}_{g+1} > \mathbf{e}_g \geq 0$. We define,

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

where

$$\rho_{itk} = \delta_{1kc} + \beta_{1kc} x_{it} + \gamma_{1kch} + \eta_{1k} Z_{it}, \tag{5}$$

where the $\mathbf{c}_{gk}$ are the thresholds of the Ordered Probit, $\mathbf{c}_{0k} = -\infty$ and $\mathbf{c}_{6k} = +\infty$. The ζ_{itk} are independent random variables with a standard normal distribution and $(\beta_{1kch}, \gamma_{1kch}, \delta_{1kc}, \eta_{1k})_{h=1,.,H,k=1,.,K,c\in C}$ is a vector of parameters to estimate. Remark that all parameters are free to vary with k. Parameter δ_{1kc} is the cohort effect varying by type, parameter γ_{1kch} is the return to education varying across education levels, types and cohorts and parameter β_{1kc} is the return to experience varying with type and cohort. Finally, η_{1k} are the effects of control variables assumed constant across cohorts and education levels but varying across types.

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, and $\bar{Z}_i$ the subset of observable time-invariant variables[21]. We have

$$\Lambda_k(h|\bar{Z}_i) = \Pr(h_i = h \,|\, \bar{Z}_i, k) = \Pr\left[\mathbf{u}_{ikh} = \max_{j\in\{1,\ldots,H\}} (\mathbf{u}_{ikj})\right], \tag{6}$$

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

$$\mathbf{v}_{ikh} = \alpha_{2kh} + \delta_{2kch} + \eta_{2kh}\bar{Z}_i, \tag{7}$$

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: $(\alpha_{2kh}, \delta_{2kch}, \eta_{2kh})$ where $h = 1,\ldots,H$, $k = 1,\ldots,K$, and $c \in C$. α_{2kh} is the education-level type effect, parameter δ_{2kch} is the cohort effect varying by type and education level, and parameter η_{2kh} are the effects of time-invariant control variables assumed constant across cohorts but varying across types and education levels.

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. In the literature, the ex ante wage expectations of individuals usually appears in the choice equations, instead of the ex post, effectively observed wages of each individual.[22] This would require a model of wage expectations depending on the

[21] In practice, we only exclude the unemployment rate from the variables in Z_{it}

[22] On this theme, see Beffy et al. (2012) and Arcidiacono et al. (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 et al. (2003), Cunha et al. (2005).

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 Matching of Workers with Firms

It may be that firm effects, more precisely, the matching of firms and workers contributes to the explanation of degree devaluation. As explained in the seminal paper of Rosen (1986), firms are viewed as differentiated products by workers, and workers are imperfectly substitutable inputs for firms. In equilibrium, graduates are sorted by firms on the basis of unobservable characteristics. It may be that the demand of firms for the labor of some worker types has varied with time.[23] A natural extension of our model is then to use a type-dependent, discrete-choice model explaining how workers are matched with some firm categories. Even with a limited number of firm classes, such a model may be difficult to estimate. This is why we estimated a relatively simple and parsimonious specification, described below, as a robustness check.

The dataset provides relatively detailed information about the firms employing individuals in each employment spell. To include firms effects in the wage equation, we create firm classes based on their predicted impact on wages and model the matching of types with firm classes in each period.[24] The procedure for classifying firms involves the following steps.

1. We first regress the log of real wages on individual characteristics (including the cohort, diploma, experience, father occupation, residency area, unemployment rate).

2. We extract the residuals from the first regression to capture the unexplained variation in wages

[23] In addition, imperfectly competitive conditions on the labor markets may explain part of the firm effects that we observe. A discussion of the recent literature on this topic can be found in, *e.g.*, Card (2022), Azar and Marinescu (2024).

[24] Note that we do not have a firm identifier in the dataset. Firms are only distinguished by a vector of observed characteristics.

after accounting for individual characteristics and we regress these residuals on firm characteristics (industrial sector, firm size, location and firm area characteristics). We obtain the fitted values for each observation from this regression.

3. We divide the fitted values into four groups separated by quartile thresholds.

4. The observed firms characteristics in every employment spell are then divided into four groups based on their score, ranked from lowest (Q1) to the highest (Q4), and we create a corresponding set of indicator variables for 4 different *firm classes*.

Once each employment spell has been classified, we add an Ordered Probit equation to the model, explaining the probability of each firm class in each period, for each worker, conditional on his type, cohort and education level. Finally, we include the firm-class dummies, interacted with the latent type, the cohort and education dummies in the wage equation. We then re-estimate the whole model. In this variant of the model, a decrease in returns to education is purged from the changes in sorting between workers and firms over time. On the one hand, the model takes care of changes in the demand for skilled workers. It is a way of controlling for variations in the demand for specific categories of workers. On the other hand, the reduced chances of finding a job in a "good firm" can itself be the consequence of an excess supply of graduates: this is one of the forms taken by the devaluation of degrees.

3.5 Estimation

The model is estimated by Maximum Likelihood. We typically use the sequential EM algorithm to obtain preliminary estimates, and then use a standard likelihood maximization routine. The model's likelihood is derived in Appendix C. It is important to note that we pool all cohorts together to estimate the model and the distribution of types. Our estimation method leads to a system of types that is robust and parsimonious. To be more precise, the types are estimated on a dataset that stacks several cohorts of students. If we reestimate the model on the cohorts taken separately, we find types and posterior probabilities of these types that are highly correlated with the original ones (see section 7.1). We find that three types do a good job at approximating unobserved student diversity. We discuss the choice of the number of types, using information and entropy criteria, in Appendix E. We discuss the identification of our model in Appendix D. 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 $K = 3$. An important output of the estimation algorithm is the *posterior* probability that individual i is of type k, that is,

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

The probability p_{ik}, can be expressed with the help of Bayes' rule and individual contributions to likelihood (see Appendix C). The posterior probabilities are a crucial ingredient in many useful computations. We also use the prior probability of type k, denoted p_k. This prior probability can be interpreted as the frequency of type k in the whole population. If we know the posterior probabilities p_{ik}, the prior p_k can be estimated as follows (again see Appendix C). For all k, we have,

$$p_k = \frac{1}{N} \sum_{i=1}^{N} p_{ik}. \tag{9}$$

4 Policy-relevant Parameters

We will use the model, the estimated values of parameters and the posterior probabilities of types p_{ik}, to compute policy-relevant parameters. In particular, we can study ATTs and ATEs. It is also possible to study the heterogeneity of treatment effects: we can compute ATEs and ATTs *conditional on type k*.

4.1 Treatment Effects

Let $y_t(z)$ denote the potential value of any outcome, at time t, for individuals with observable characteristics z.[25] 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,

$$ATE(h,k,t) = \mathbb{E}[y_t(h)|k] - \mathbb{E}[y_t(0)|k]. \tag{10}$$

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

$$ATE(h,t) = \sum_k p_k ATE(h,k,t), \tag{11}$$

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,

$$\hat{\mathbb{E}}[y_t(h)|k] = \frac{\sum_i y_{it}\hat{p}_{ik}\chi_h(i)}{\sum_i \hat{p}_{ik}\chi_h(i)} = \frac{\sum_{\{i|h_i=h\}} y_{it}\hat{p}_{ik}}{\sum_{\{i|h_i=h\}} \hat{p}_{ik}}, \tag{12}$$

where $\hat{p}_{ik}$ is the estimated posterior probability that i belongs to group k, computed by Bayes's law. Statistic (12) is an estimation of $\mathbb{E}(y|h,k)$, using the sample. In a similar fashion, we define,

$$\widehat{ATE}(h,t) = \sum_k \hat{p}_k \widehat{ATE}(h,k,t), \tag{13}$$

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

$$\widehat{ATE}(h,k,t) = \hat{\mathbb{E}}[y_t(h)|k] - \hat{\mathbb{E}}[y_t(0)|k]. \tag{14}$$

[25] 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.

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

$$\mathbb{E}[y_t(h)|h', k] = \mathbb{E}[y_t(h)|k] \quad \text{for all } (h', h). \tag{15}$$

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

$$ATT(h, k, t) = ATE(h, k, t), \tag{16}$$

and it is easy to show that,

$$ATT(h, t) = \sum_k p(k|h)ATE(h, k, t), \quad \text{where} \quad p(k|h) = \frac{p(h, k)}{p(h)}. \tag{17}$$

Now, 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[26]

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

and

$$\hat{p}(h) = \sum_k \hat{p}(h, k) = \frac{1}{N} \sum_i \chi_h(i). \tag{19}$$

Finally, $\widehat{ATT}$ is just obtained by putting hats on p and ATE in equation 17.

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

[26]This definition corresponds to the standard notion of ATT. Note in addition that $\hat{p}(h, k) = \frac{1}{N} \sum_i \Pr(k|h, i) \Pr(h|i)$. It is easy to check that $\sum_h \sum_k \hat{p}(h, k) = 1$.

$$\hat{p}(h|k,z) = \frac{\hat{p}(h,k,z)}{\hat{p}(k,z)} = \frac{\sum_i \hat{p}_{ik}\chi_{hz}(i)}{\sum_i \hat{p}_{ik}\chi_z(i)}, \tag{20}$$

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

4.2 Discounted Sums of Earnings.

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 sequences of employment rates and wage rates $(\tilde{e}_{itk}, \tilde{w}_{itk})$ for each individual i in the sample. Then, using weights p_{ik}, we averaged the expected-discounted fictitious sequence of earnings of each i during the periods $t \in \{1, \ldots, T\}$. We choose a discount factor $\delta = .99$ (per month) and for every (i,k), we compute,

$$\tilde{W}_{ik} = \frac{(1-\delta)}{(1-\delta^T)} \sum_{t=1}^{T} \delta^{t-1} \tilde{e}_{itk} \exp(\tilde{w}_{itk}). \tag{21}$$

$\tilde{W}_{ik}$ 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,

$$H_k = \frac{\sum_{i=1}^{N} \tilde{W}_{ik}\hat{p}_{ik}}{\sum_{i=1}^{N} \hat{p}_{ik}}. \tag{22}$$

See Appendix F 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.

5 Estimation Results

We can now present our estimation results. The wage equation has been estimated with a sample including $N = 15{,}841$ individuals with full-time jobs and 105,496 observations. For the the Multinomial Logit, we use 16,404 individuals (including part-time workers) and the Ordered Probit is estimated with 178,412 observations (and 16,404 individuals). We start by describing the distribution of types and how they impact labor market outcomes. We will show that the types identified by our model present a clear hierarchy.

5.1 Distribution of Types

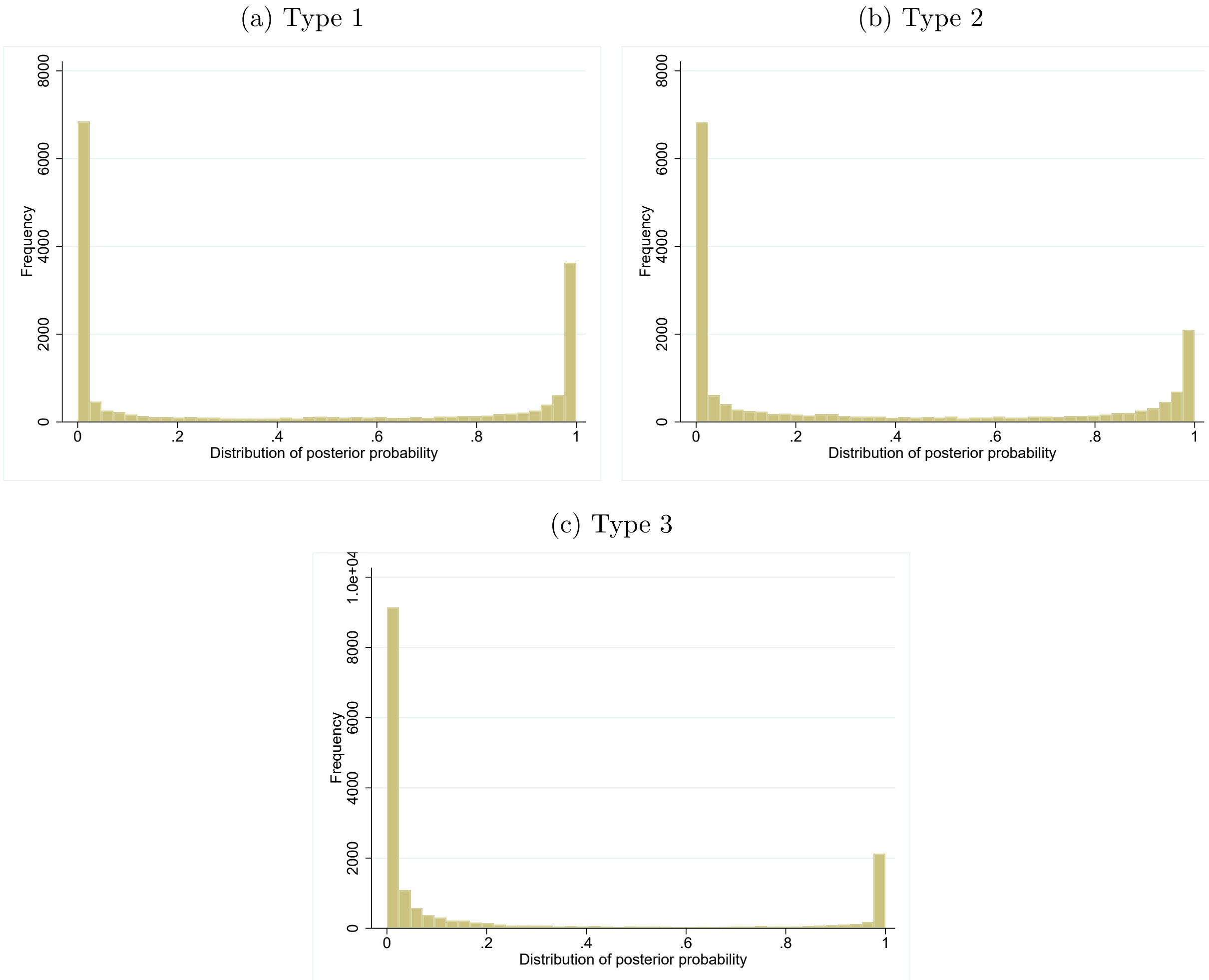
Table 4 presents the estimated probabilities of types when $K = 3$. In our sample 42% of young men are of type 1, 36% of type 2 and 22% of type 3. The type frequencies are very precisely estimated. Before we provide an interpretation of these types, it is important to check if these types generate a good classification of the population. The quality of classification is 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 $p_{ik} \simeq 1/K$. In Figure 1, the histograms of the estimated values p_{ik} for each k immediately shows that the classification is very good. Most individuals have values of p_{ik} close to 0 or 1. 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. In the coming subsections, we will see how individual wages and employment rates depend on types.

Table 4: Estimated Probability of Types

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

Note. The estimates are obtained by Maximum Likelihood. We have $p_3 = 1 - p_1 - p_2$.

Figure 1: Empirical Distribution of Posterior Type Probabilities



Note. Panels (a), (b) and (c) give the histograms of estimated posterior probabilities $\hat{p}_{ik}$ for $k = 1, 2, 3$, resp. The probabilities are concentrated around 1 and 0, showing the good quality of classification.

5.2 Parameter Estimates

To understand how types are affecting labour market outcomes, we now consider in turn the ML estimation results of the three building blocks of our model: the wage equation, the employment equation and the education choice equation.

Wage equation. The ML estimates of the wage equation are presented in Appendix G, in Tables 11, 12 and 13. Table 11 presents the wage returns to experience by cohort, type and education level (β_{0kch}). Table 12 presents the returns to education by cohort and type (γ_{0kch}). Table 13 presents the other parameters of the wage equation ($\alpha_{0k}, \delta_{0kc}, \eta_{0k}$). A glance at Figure 2 will show the main insights that can be drawn from the wage equation. The first column of Figure 2 (Fig 2.a, 2.c and 2.e) shows the returns to education levels in the three cohorts, while the second column shows the returns to experience (*i.e.*, Fig 2.b, 2.d and 2.f). In each sub-figure, we observe the returns for the three types (represented by three confidence intervals with different colors) and the 5 education levels.

The returns to education or returns to degrees can also be viewed as returns at zero experience determining average starting salaries. The education levels (5 groups of three intervals) 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: 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 and engineering schools). The returns to experience are presented as average percentages of wage growth by month. 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; thirdly, returns to experience have tended to decrease with time, between 1998 and 2017. 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 and engineering school graduates, that has increased in the most recent cohort.
Impact of some control variables. Table 13, in Appendix G, provides the estimated coefficients of some important control variables. We control the wage equation for the macroeconomic rate of unemployment. 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

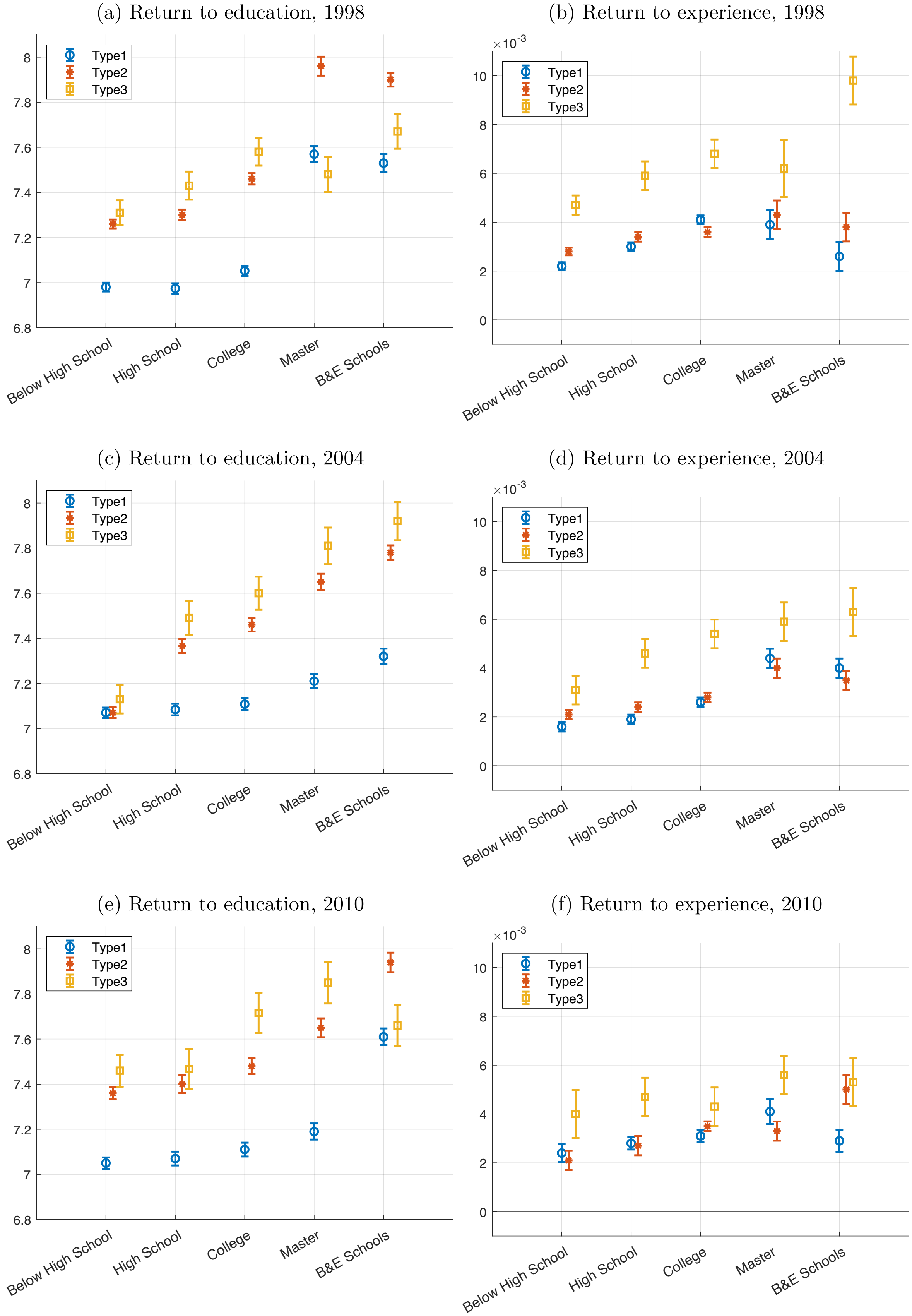
to reach education levels at which wages are protected by the minimum wage legislation. But the impact of overall unemployment is clearly negative for Types 2 and 3, as expected: we find mildly procyclical real wages. Secondly, we find a significant and positive effect of the *father is a professional* dummy. This latter effect is much stronger for Type 3 (five times larger than the effect on Type 1). This dummy indicates individuals whose father's occupation requires higher-education degrees: executives, doctors, lawyers, engineers, teachers, etc. Indications of geographic origin are significant too: the rural and peri-urban individuals earn (slightly) smaller wages.[27]

Employment equation. Estimates of the Ordered-Probit parameters are presented in Table 14 in Appendix G. The Ordered Probit shows a striking feature of Type-2 individuals. This is visible if we look at the Ordered-Probit cuts. Figure 3 gives a representation of these cuts. To be more precise, the table 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[28] has an employment rate e smaller than x. Figure 3 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. These results give the impression that Type 2 finds a job quickly and stays in this job: the matching of Type 2 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.

[27] The Peri-urban is a heterogeneous category including suburban and small-town France. Note that, unlike in America, the French suburban dweller generally does not have a well-to-do background. The urban, city-center resident is more likely to come from a privileged background.

[28] This individual of reference is a High-School dropout in 1998, who lives in urban area whose father has a non-executive occupation and where there is no unemployment.

Figure 2: Monthly Returns to Experience and Education by Type, Education Level and Cohort



(a) Return to education, 1998

(b) Return to experience, 1998

(c) Return to education, 2004

(d) Return to experience, 2004

(e) Return to education, 2010

(f) Return to experience, 2010

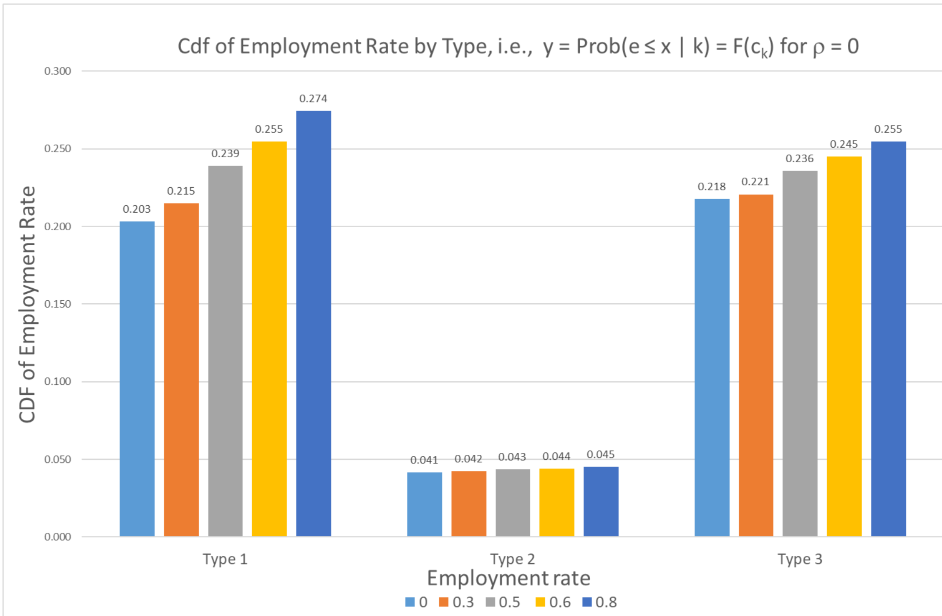
Our estimates of the employment equation exhibit a few other properties. The probability of full-employment generally decreases with the cohort, and particularly for Types 1 and 2. In a certain sense, this contributes to the devaluation of degrees. The probability of a higher rate of employment typically 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 is also stronger for Type 2 than for Types 1 and 3. In Appendix I, we show the average number of employment spells by type on Table 18, which confirms the greater job stability of Type 2.[29] Table 18 also shows that Type 2 individuals are more likely to work in the public sector where jobs are more stable and that Type 1 tends to work in smaller firms, while Type 2 and Type 3 work in bigger firms (medium and large).

Education choices. How do types affect education choices? The estimated parameters of the Multinomial Logit, describing education choices, are presented in Table 15 in Appendix G. Table 5 presents the conditional probabilities $\hat{p}(h|k,c)$ of choosing level h, given the type k and cohort c. This table shows that types are far from being completely predicted by education levels. As time passes, types seem to specialize more but all of them are characterized by shifts towards longer studies. Table 5 shows that 60% of Type-1 students end up with a high-school degree or less in the last cohort, *Some college* and Master 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 5 shows that there has been a rush on Master programs and schools, and a certain flight from the lowest levels and the 'some college' category. Types differ only in the intensity of changes. To conclude, Type 1

[29] Similarly, Figure 8 shows that the average employment rate, by month, since the individual left the educational system, is higher for Type 2 than for the two other types.

populates the lower levels of education more than Type 2 and 3 in every cohort, and this has increased over time. The multinomial logit confirms that Type 1 is in a certain sense the "weak type".

Figure 3: Employment Rates by Type: Analysis of Ordered Probit Cuts



Note. The figure plots the cumulated probabilities of the Ordered Probit cuts, $F(c_k)$ by type $k = 1, 2, 3$ and by employment rate e, for employment rates $e < 1$, when all controls are set equal to zero. The probability of $e = 1$ is $1 - .274 = .726$ for Type 1; $1 - 0.045 = .955$ For Type 2 and $1 - .255 = .745$ for Type 3. It follows that the Type 2 are strikingly different in terms of their estimated probability of being fully employed.

Table 5: 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
High-school Degree and Less	0.69	0.62	0.65	0.63	0.49	0.46	0.60	0.44	0.43
Among which									
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 and School Degrees	0.05	0.08	0.09	0.09	0.20	0.22	0.16	0.22	0.34
Among which									
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

Note. In 1998, 65% of Type 3 students who left education had reached a High School degree or less, 26% had some college education and 9% of had a Master or a B & E School degree (among which 4% with a Master and 5% with a B & E School degree).

5.3 Interpretation of the Types

The previous results have helped us characterizing the three types.

Type 1 has smaller returns to experience and smaller returns to education than other types. This type also studies less than other types. It seems that it represents individuals with a smaller ability.

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. Type 2 is also characterized by a median position in terms of educational achievement. 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.

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%. Type 3 also have higher educational achievement. The lower employment of Type 3 as compared to Type 2 is associated with higher returns to experience, suggesting a higher job mobility behind the causes of less employment. In contrast, the lower level of employment of Type 1 is associated with low return to experience, suggesting a high probability of remaining unemployed.

Prediction of types. Regression of type probabilities on observable variables. By construction, types are supposed to be orthogonal to the observable characteristics included in our model. However, are they correlated with some 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, we used an *elastic net* method to select the variables correlated with types 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 (see Appendix H). We first assign his most likely type to each individual i (*i.e.*, the type k with the highest ex-post probability $\hat{p}_{ik}$). Then, we estimate which variables are correlated with the types. Table 16 gives the results of the elastic net procedure. A first reassuring result is that observable variables do not help predicting correctly the types. The confusion matrix shows very poor prediction results. Among the selected variables, the results indicate that Type 2 and Type 3 individuals are more prevalent among those who did not repeat a grade before junior high school, whereas Type 1 is more common among those who repeated a grade before high school. Additionally, Type 3 is less prevalent among individuals who grew up in rural areas, while Types 1 and 2 are more common in these areas. Parental occupation also plays a role: Types 2 and 3 are less prevalent among individuals whose parents work in agriculture, whereas Type 1 is more common in this group. Geographically, Type 1 is more frequent among individuals living in the south of France, while Type 3 is more concentrated in Paris and its surrounding region. Furthermore, Type 3 is more prevalent among individuals whose parents hold a university degree.

Overall, these findings suggest that types are correlated with certain observable characteristics but are not merely proxies for omitted observable traits. Notably, family background alone does not predict type membership, as all three types are present across various family backgrounds. It is reasonable to infer that types are associated with individuals' underlying abilities. As further evidence, Table 17, in Appendix H, shows that type membership is correlated with students' grades on the National High School Exam (*i.e.*, the *baccalauréat*), which marks the completion of high school (available only for the 2010 cohort). Notably, Type 3 has higher grades, Type 2 intermediate grades and Type 1 the smallest grades.

6 Evolution of Returns to Degrees and Experience

We now use our model to analyze the evolution of returns to education, using the ex post probabilities of types to compute ATTs and ATEs. A key question is whether this evolution stems from a composition effect, that is, did the expansion of higher education lead to a lower ability threshold for obtaining a degree, thereby reducing the graduates' productivity? In addition, are changes in returns to education heterogenous, and more precisely, type-dependent? The identification of types allows us to address these questions.

6.1 Treatment Effects on Wages after 7 Years

For each individual i in the sample, we compute the arithmetic average $\bar{w}_i$ of the log-wages observed during the seventh year of career (more precisely, when potential experience is between 78 and 90 months). Then, using notations similar to those of section 4, we compute,

$$\widehat{\mathbb{E}}(\bar{w}_i|c,h,k) = \frac{\sum_i \bar{w}_i \hat{p}_{ik} \chi_{ch}(i)}{\sum_i \hat{p}_{ik} \chi_{ch}(i)}, \tag{23}$$

where $\chi_{ch}(i) = 1$ if individual i is in cohort c and has education level h, and 0 otherwise. Similarly, we use χ_c with the meaning that $\chi_c(i) = 1$ if i is in cohort c and 0 otherwise. Next, we need two different estimated conditional probabilities of types, $\hat{P}(k|c)$, the frequency of type k in cohort c and $\hat{P}(k|c,h)$ the frequency of type k in the subset of individuals of cohort c with education level h. More precisely, we compute,

$$\hat{P}(k|c) = \frac{\sum_i \hat{p}_{ik} \chi_c(i)}{\sum_i \chi_c(i)} \quad \text{and} \quad \hat{P}(k|c,h) = \frac{\sum_i \hat{p}_{ik} \chi_{ch}(i)}{\sum_i \chi_{ch}(i)}, \tag{24}$$

Now we redefine ATT and ATE as being simply different ways of computing expected wages in the seventh year of career, given education h and cohort c. More precisely, we take the exponential of the conditional log-wage averages, to obtain an expression of wages as monthly salaries, in euros, as follows,

$$ATE(c,h) = \exp\left(\sum_k \widehat{\mathbb{E}}(\bar{w}_i|c,h,k)\hat{P}(k|c)\right) \text{ and } ATT(c,h) = \exp\left(\sum_k \widehat{\mathbb{E}}(\bar{w}_i|c,h,k)\hat{P}(k|c,h)\right). \tag{25}$$

$ATT(c,h)$ is the realized average wage of individuals who actually achieved education level h in cohort c. In contrast, $ATE(c,h)$ is the average wage that we would have observed, had a random and representative sample of types (drawn in the whole population) reached education level h, in cohort c. Finally, we define a measure of selection as a percentage variation, that is,

$$\text{Selection}(c,h) = \frac{ATT(c,h) - ATE(c,h)}{ATT(c,h)}. \tag{26}$$

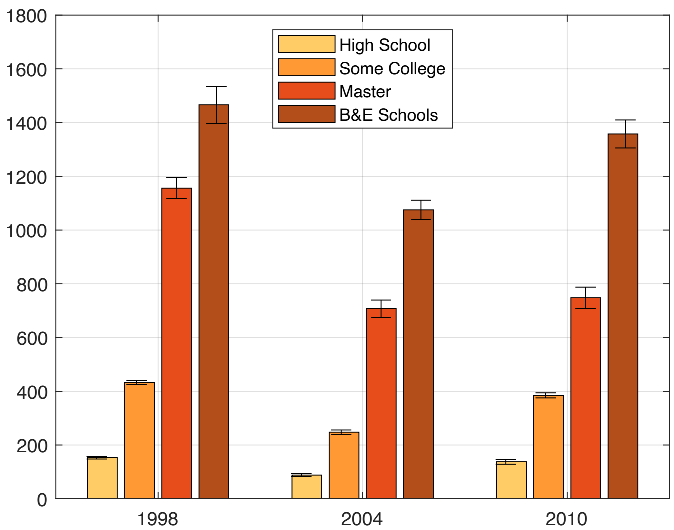
The idea is to express the difference between ATT and ATE as a fraction of the actual return, ATT. If this difference is positive, then, the individuals who self-select in education h earn higher wages on average than what would have been the average earnings of a random sample of the population: we define this as *positive selection*. In the opposite case, we observe a negative selection. This type of selectivity measure has been used in other studies (Belzil and Hansen, 2020).

Figure 4 gives: (a) the difference between the ATE of education level h and the ATE of level education 0 (less than a high-school degree), and (b) the difference between the ATT of education level h and the ATT of level 0, for full-time wages observed during the seventh year of career as defined above. Figure 4(c) gives the measure of selection defined above. The ATE of an education degree consistently increases with the level of education. Compared to individuals with less than a high school degree, the ATE values in 1998 are as follows: High school degree: +150 euros/month, Bachelor's degree: +400 euros/month, Master's degree: +1,150 euros/month Business and Engineering school degree (B&E): +1,450 euros/month. While the ranking of ATE values remains stable across the three cohorts, a notable trend is the sharp decline in the ATE

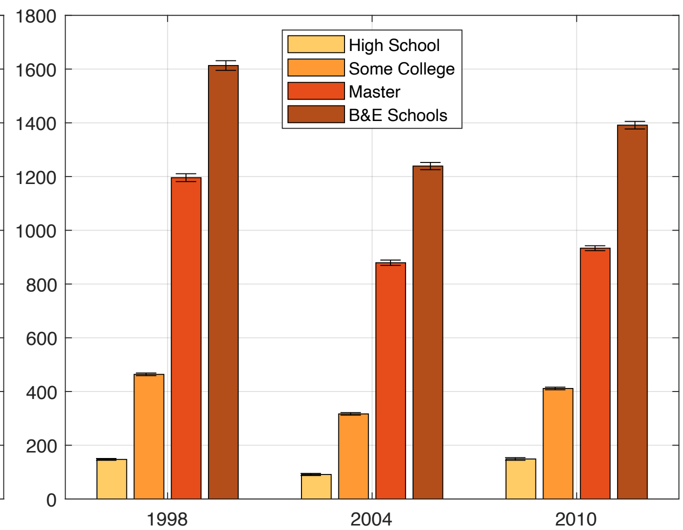
of Master's degrees in 2010. Specifically, the ATE of Master's degree holders drops by 35%, from 1,150 to 750 euros per month, whereas the ATE decreases by 7% for B&E school graduates and remains stable for individuals with some college education. The decrease in the ATE is statistically significant.[30]

Figure 4: ATE and ATT of Education on Full-time Wages are the Outcome

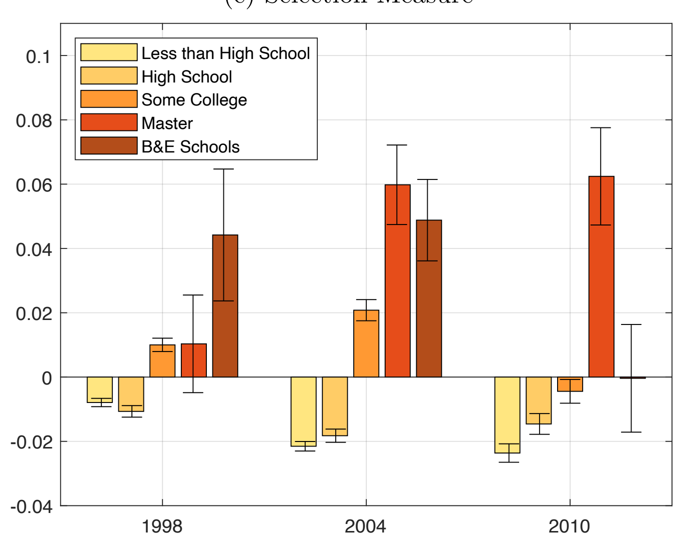
(a) ATE(c,h) - ATE(c,0)



(b) ATT(c,h) - ATT(c,0)



(c) Selection Measure



Note. ATE and ATT are computed using the average full-time wages observed after 7 years of career. More precisely $ATT(c,h)$ is the average return of the individuals who effectively reached level h in cohort c. $ATE(c,h)$ is a counterfactual estimation of the average return obtained at level h if a random representative sample of the population in cohort c had achieved level h. ATT and ATE are expressed as monthly salaries in euros. The measure of selection is the percentage variation $(ATT - ATE)/ATT$. A positive (negative) value means positive (negative) selection.

[30] Standard errors have been obtained using 50 replications of a standard bootstrap procedure.

We now examine the ATT and, more importantly, the difference between ATT and ATE, which provides insights into selection effects. We find that the ATE is slightly larger than the ATT for less than high school and high-school graduates, *i.e.*, individuals whose highest level of education is a high-school diploma. This confirms the intuition that high-school completion with no further education does not select the most productive students. However, the ATE is smaller than the ATT for higher levels of education.

Panel (c) of Figure 4 shows that over time, the selection of Master's graduates has improved. This finding is surprising, as one might have expected a decline in selection due to the rapid expansion of Master's programs. Hence, this result suggests that the degree of devaluation observed is not primarily due to adverse selection but to an excess supply of graduates. In contrast, the opposite trend is observed for business and engineering school graduates. When it comes to B&E schools, the difference between ATT and ATE decreased between the 1998 and 2010 cohorts, indicating a decline in selection quality for these schools. This is consistent with the rapid expansion of business schools in France in the recent years. It is possible that this growth was in some cases achieved at the expense of selectivity.[31] Overall, our results suggest a statistically significant divergence in selection quality between universities and business schools. Our results seem to indicate that, despite increasing enrollment, university Master's programs have improved their selectivity. In contrast, business schools seem to have expanded at the cost of reduced selectivity, as we will see below, when we discuss composition effects. In fact, when excluding business and engineering schools, as well as doctoral programs, the Master's degree has emerged as the only truly selective pathway within French universities. We can view these results as evidence that a form of Spence's signaling theory is true, since higher education levels are more positively selected than lower levels.

[31] Recently, French business schools have introduced new teaching programs, such as "Bachelor's" degrees. However, as previously mentioned, in this study, individuals in the business-school category have completed five years of post-secondary education and can be compared to Master's graduates.

The changes in ATE on wages after 7 years encompass changes in the starting wage, the return to experience, and the amount of experience accumulated over the first 7 years of a career. It is possible to decompose the changes in discounted earnings using the model coefficients estimated via maximum likelihood and the posterior probabilities, to compute the ATE. In particular, we compute changes in wages at zero experience and changes in the returns to experience. To compute changes in the employment rate for a random sample of the population, we calculate the average employment rate by type and cohort, using the posterior probabilities of types as weights. Results are displayed in Table 19 in Appendix J. The first column and Figure 9 show the returns to experience. The second column shows the ATE of education on employment, the third column shows ATE of education on starting wages. For a random individual in the population graduating with a Master's degree, the expected wage at zero experience in the 1998 cohort is 2026 euro/month, while in 2010 it is 1785 euro/month (a 12% drop). The ATE of effective experience decreased by 9% over the period—from 5.6% per year in 1998 to 5.1% in 2010. The average employment rate was 76% in 1998 and declined to 69% in 2010, representing a 7 percentage point drop. Figures for the other education levels can be found in the above-mentioned appendix. Notably, we observe a drop in returns to experience for all education levels and conclude that the recent years witnessed a remarkable decrease in the return to experience of young French workers. The drop in returns to experience of the lowest education levels creates a polarization of the latter returns. Indeed, when it comes to returns to experience, the difference between the lowest and highest levels of education increased over time. In the next section, we show the changes in discounted earnings to provide a synthetic measure of the change in the returns to degrees, accounting for all these different components.

6.2 Simulations of Discounted Earnings

To study the devaluation of degrees and the changes in selectivity, we also computed measures of "human capital", that is, more precisely, sums of expected and discounted future wages, conditional on type, education and cohort. To obtain these discounted sums, we simulated employment levels $\tilde{e}_{itk}$ and wages $\tilde{w}_{itk}$ for each individual i, each period t and each possible avatar k of i, as indicated in subsection 4.2. Future expected wages are discounted using the monthly discount factor $\delta = .9987$. Technical details on the simulations are presented in Appendix F. With the help of simulated careers, we compute discounted sums $H_k(c, h)$ for each type k and each education level h in cohort c. From these values we can derive the *actual* discounted earnings, based on averaging $H_k(c, h)$ over types k with probabilities $\hat{p}(k|c, h)$ and to obtain the *counterfactual* expected value of discounted earnings at level h, we average over types k with probabilities $\hat{p}(k|c)$ (the equivalent of ATE in this context). The discounted earnings so obtained are synthetic measures: they have the advantage of summarizing the impact of all the variables that are related with observable choices and unobservable types: not only wages, but also employment rates and accumulated experience.

The first striking result displayed in Table 6 is the devaluation of Master degrees in terms of discounted earnings. Columns called Actual and Counterfactual give the expected discounted wages defined above. The values are in real euros per month. We see that the counterfactual value of a Master falls from 1689 in 1998 to 1261 in 2010, a loss of 400 euros (a 25% decline). The corresponding decline in discounted expected real earnings based on observed outcomes (Actual, ATT) is about 11% over the same period: from 1589 euro in 1998 to 1414 in 2010.

It is common to evaluate the ATE of education relative to the high school level, and to compute the ATE of one additional year of schooling. The difference in discounted expected real earnings between a Master's degree and high school declined by 37 percentage points between 1998 (82%)

and 2010 (46%)—corresponding to a 7.2 percentage point decrease in the return to an additional year of education over the period (i.e., from 16.4% in 1998 to 9.2% in 2010). The corresponding figure based on observed (*actual*) data yields a decline of 3 percentage points, from 74% in 1998 to 71% in 2010.

The column called 'Difference' 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 Difference column are negative in the three cohorts, but the difference between Actual and Counterfactual is small. These education levels seem to suffer from more adverse selection as time passes. In contrast, the selection of Master program graduates seems to improve with time. This confirms the results of Figure 4, based on wages in the seventh year. The latter results show a significant increase in the selection measure for masters, and a significantly positive selection in 2004 and 2010. The number of students enrolled in master programs has increased, but in spite of this growth, these university degrees have selected students that seem better than the average in a certain sense: they tend to have a higher type.

Next, we see on Table 6 that the discounted earnings of engineering and business-school graduates has followed a completely different path. It seems that the quality of the selection of B&E 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. We will now discuss the composition effects underlying the estimated changes.

Table 6: 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	$+9$	$+0.8\%$
	Master	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	$+21$	$+1.8\%$
	Master	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	$+17$	$+1.5\%$
	Master	1414	1261	$+153$	$+10.8\%$
	Bus. & Eng. Schools	1862	1954	-92	-4.9%

Note. The column called 'Actual' gives the values of $\sum_k \hat{p}(k|c, h)H_k(c, h)$. The column called 'Counterfactual' gives $\sum_k \hat{p}(k|c)H_k(c, h)$. 'Difference' is just Actual$-$Counterfactual. All entries of the middle three columns are in expected euros per month.

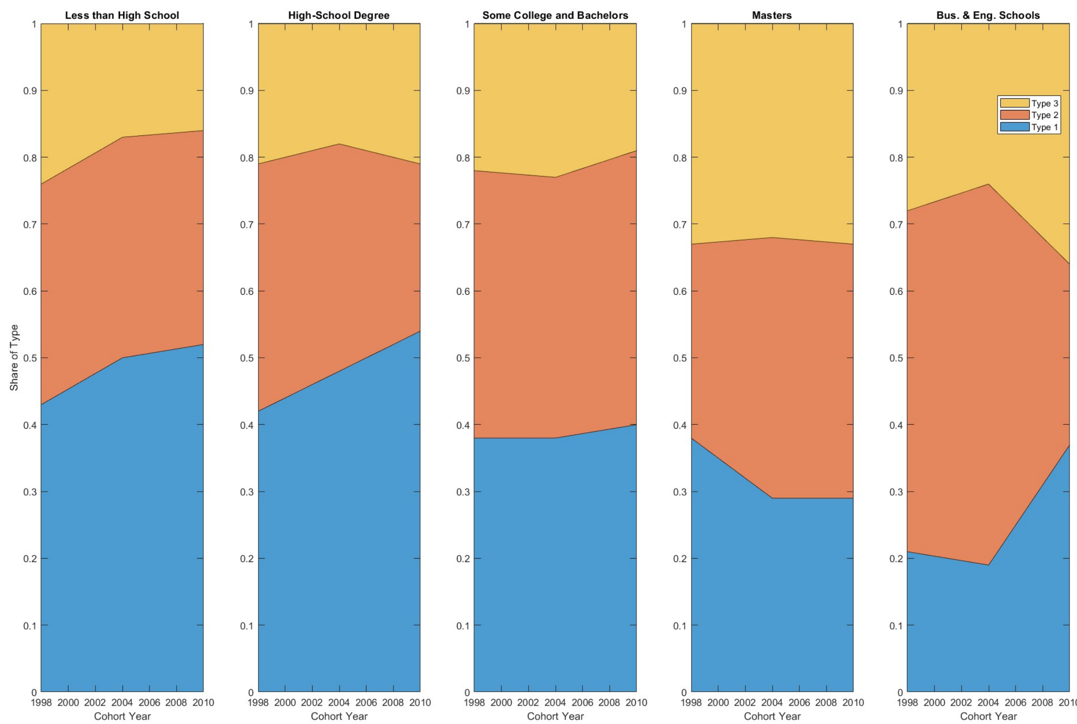
6.3 Composition Effects

The distribution of unobserved student types has evolved within each education level. This gives rise to composition effects. Figure 5 displays the empirical values of conditional probabilities $\hat{p}(k|h, c)$, using the estimated posterior probabilities. At the lower education levels (*i.e.*, less-than-high-school and high-school degree), we observe that selection deteriorated to the extent that the share of Type 1 increased. Selection at the level of 'Some College and Bachelors' improved slightly as the increase in the share of Type 2 compensates for the increase in the share of Type 1 and the decrease in that of Type 3. The selection into Master's improves due to a clear drop in the probability of Type 1, the latter being replaced by Type 2. The Master programs were able to sort out Type-1 students in the most recent cohort and to select a higher share of Type 2. Whereas B&E Schools were able to starkly increase the share of Type 3, there has been also a strong increase of the Type 1 frequency so that selection deteriorated on average between cohorts 1998 and 2010.

Overall, we find that the French educational system has become increasingly selective over time.

In 1998, the distribution of lower education levels closely resembled that of the general population (up to 'Some College', see Table 4), and Master's degree programs were suffering from adverse selection, meaning that talent was not the main factor in selection. In more recent cohorts, lower education levels are characterized by adverse selection, while higher education levels benefit from a more positive selection. Business and engineering schools are an exception, since some of them became significantly more selective, while others increased enrollment while becoming less selective.

Figure 5: Mix of Types by Education Level and Cohort



There has been a change in the importance of family background for the selection into higher levels of education. Table 20 in Appendix K gives the probability of reaching level h conditional on type k and cohort c. The table shows a general increase in the probability of reaching the highest levels of education (*i.e.*, Master's degrees or B & E schools) between 1998 and 2010, across all groups, regardless of parental background. This increase is especially pronounced for high-ability individuals from disadvantaged backgrounds. We distinguish two coarse categories of father occupations: the professional and 'non-professional' fathers.[32] For the sons of a 'non-professional' father, the probability of reaching a Master's degree tripled for Type 1 individuals, increased elevenfold for Type 2, and nearly fivefold for Type 3. Among sons of professional fathers, the gains are more modest: the probability doubles for Type 1, and increases by nearly four times for both Type 2 and Type 3. This suggests improved selection at the Master's level, allowing more high-ability students from modest backgrounds to succeed. In terms of access to B & E schools, the probability also rose significantly for disadvantaged students — multiplying by five the frequency of the Type-1 sons of non-professional fathers, and doubling or quadrupling this frequency for Types 2 and 3. Interestingly, low-ability individuals saw their chances of reaching B & E schools increase more than high-ability ones, regardless of parental background.

Table 21 (in Appendix K) focuses on the distribution of types in each education level and cohort. In 2010, 71% of Master's graduates with a non-professional father were classified as high-ability (*i.e.*, Type 2 or 3), up from 59% in 1998. The same proportion (71%) was observed among graduates with a professional father in 2010, compared to 65% in 1998. This convergence suggests that, at the Master's level, high-ability students from disadvantaged and privileged backgrounds are now equally represented. However, the trend is reversed for B & E schools: in 1998, only 18% of the B & E school graduates from privileged backgrounds were low-ability, whereas by 2010, this figure had doubled to 38%. This indicates that B & E schools are increasingly enrolling low-talent individuals from wealthy families, while Master's programs have expanded opportunities for talented students from disadvantaged backgrounds (on this theme, see Ichino et al. (2024)).

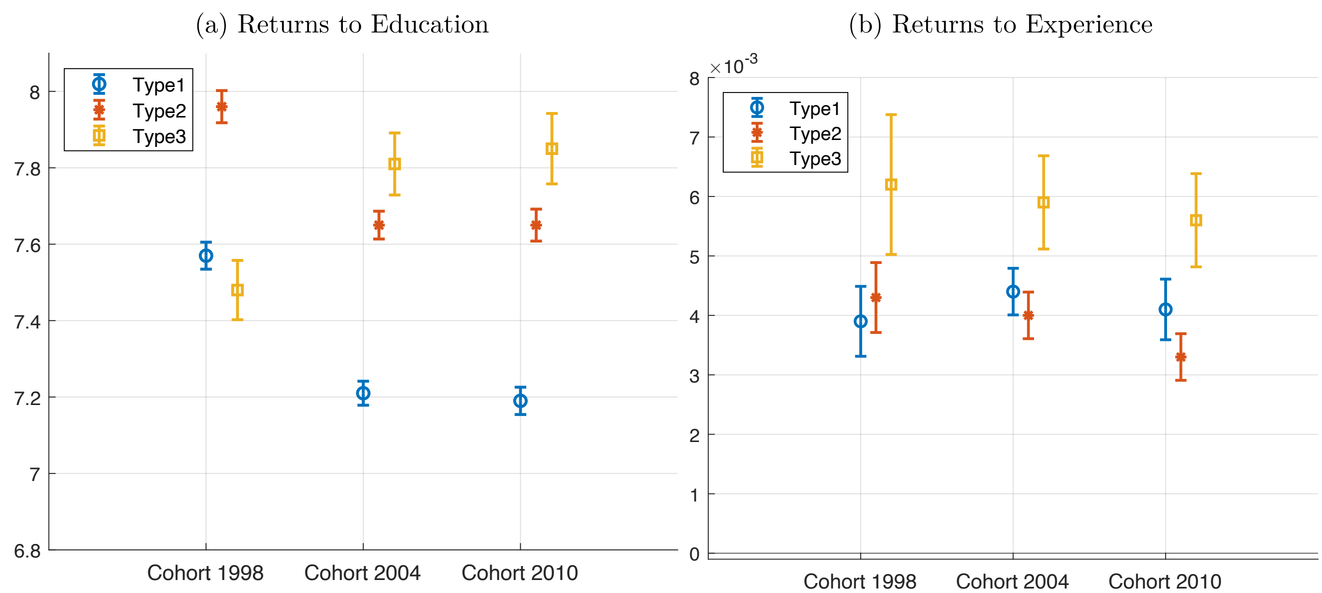
[32]Professional fathers include in fact executives, doctors, lawyers, engineers, teachers, etc., all occupations requiring higher education.

6.4 Heterogeneity: Type-Dependent Effects

Our methodology allows us to uncover that treatment effects vary with the unobserved type. Changes in entry wages and returns to experience vary even within the same education level.

Figure 6 illustrates the evolution of returns to education and experience for Master's degree holders across cohorts. The results reveal that entry wages, conditional on type and a Master's degree, have declined for Types 1 and 2 but increased for Type 3. Additionally, returns to experience dropped for Type 2, while they remain stable for Types 1 and 3. This confirms the devaluation of Master's degrees, but also highlights that the effect is heterogeneous across types.

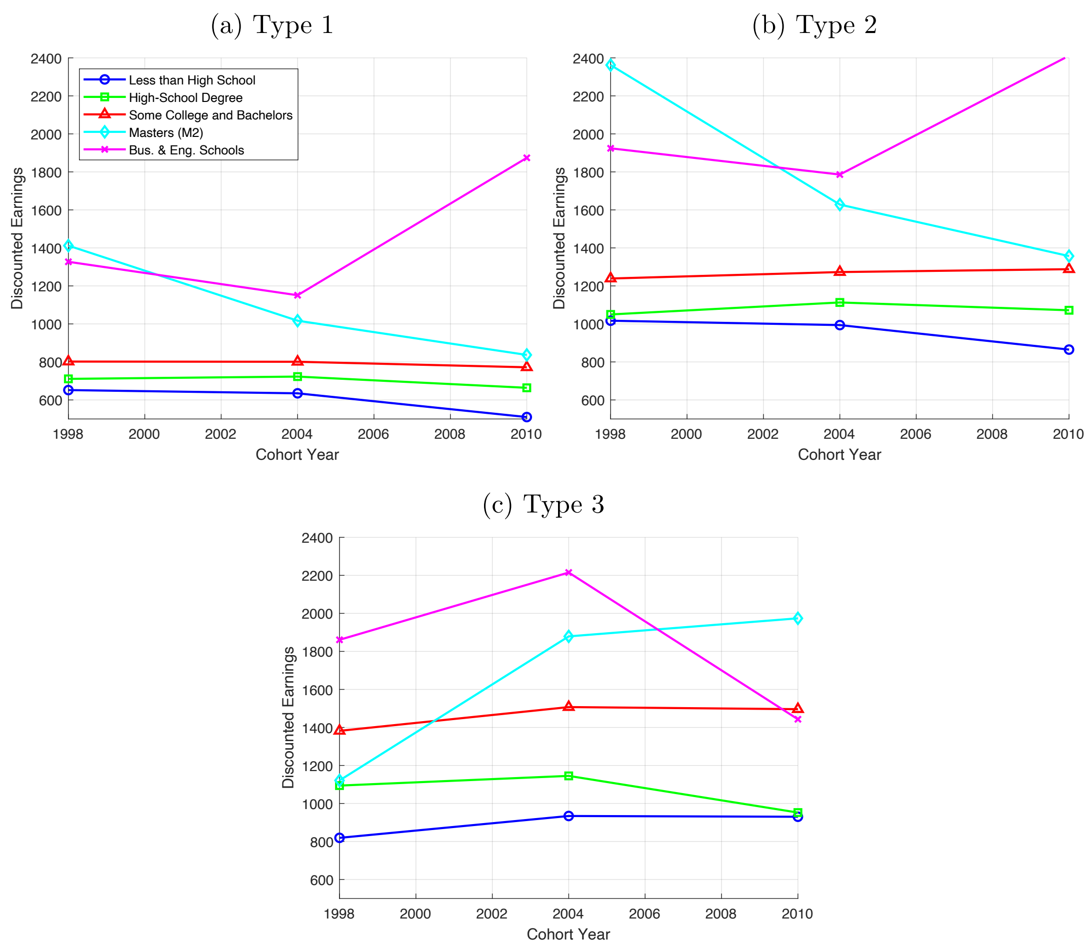
Figure 6: Evolution of the Returns to Education and Experience for the Master Level



Note. The figure displays estimated coefficients of the model and associated confidence intervals, focusing on the students holding a Master's degree. Panel (a) gives the evolution of wages at zero experience for the three types. Panel (b) shows the evolution of returns to experience for the three types.

To give a synthetic picture of the heterogeneity of changes in the values of degrees, we present in Figure 7 the dynamics over time of the discounted wages as computed in sub-section 6.2 for every education level. Figure 7 shows the average discounted earnings by education level, cohort and type, obtained over the first seven years of career. Again we see a clear hierarchy of types. The figures are rather low for Type 1: this is due, not only to smaller monthly wages, but also to a high unemployment rate. In terms of discounted earnings, devaluation took place for the less-than-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. The devaluation of Masters is confirmed for Types 1 and 2, but not for Type 3 (this confirms the results illustrated by Fig. 6). 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).

Figure 7: Average Discounted Earnings by Type, Degree and Cohort, *i.e.*, $H_k(c, h)$



Note. The figure displays the discounted sums of expected wages $H_k(c, h)$ conditional on cohort, education and unobservable type. Computations are based on simulated careers, as indicated in Appendix F. Panel (a), (b) and (c) give the evolution over time of the discounted expected wages conditional on Types 1, 2 and 3 respectively.

7 Robustness and Discussion of Firm Effects

7.1 Cohort-by-Cohort Estimation of the Model

The model presented above has been estimated with a sample stacking three cohorts of males. We find very similar results when the model is estimated with three types on each of the three cohorts, taken separately. Table 22 in Appendix L 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 with a single cohort. The table 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.[33] 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, obtained with each of the three cohorts, taken separately, do not show big differences.[34] This is reassuring, because one could have suspected that the structure of the economy has changed with time in a manner that the observed variables do not explain well. Yet, the three-cohort model is very flexible, most coefficients depending on the cohort: this very flexibility probably explains that subsample estimation does not lead to markedly different results.

7.2 Matching of Workers to Firms. Congestion Effects.

Given the results presented until now, we could suspect that our latent type system may be capturing firm effects, or wage premia related to firm-worker matching. To make sure that this is not the case, we estimated a more complicated model, called 'the variant', with an additional equation. As described above, in sub-section 3.4, we use a score valuing observed firm characteristics, like

[33]We relabeled the types to match the most similar types obtained when all cohorts are pooled. Types are identified only up to relabeling.

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

firm size. Then, this score is attached to each employment spell and we generate a set dummies indicating 4 firm classes separated by quartiles. Next, a type-dependent Ordered Probit explains the firm class associated to the worker at each period. Finally, firm-class indicators, interacted with the education level, the cohort and the type, are included in the log-wage equation, yielding wage premia related to firm classes.

The conclusions are twofold: *(i)*, the results obtained with the base model are robust; our latent-type system essentially survives; *(ii)*, the firm effects are significant, sizeable and we observe a type-dependent drop in the wage premium commanded by a job in the top firm-class.

Labor-market theory à la Rosen (1986) sees jobs as imperfect substitutes from the point of view of workers, and workers as imperfectly substitutable inputs from the point of view of firms. We observe that 'top types' tend to be matched with 'top firms'. Jobs in top firm classes command a premium because these firms compete for talent. If large firms were different because they offer nice offices and amenities, they would be in a position to pay less for graduates. As a consequence, it must be that the net effect of nice amenities and competition for talent is a positive wage premium. Conversely, if the growth in the number of graduates looking for a job is larger than the growth in the demand for this type of labor, a congestion effect appears in top firm classes, in fact, this is a form of excess supply, resulting in a decrease of wage premia (combined with some occupational downgrading). The drop in firm effects that we find therefore confirms the hypothesis of a devaluation of degrees due to an excess supply of graduates. In Appendix M , we detail the results obtained with this variant.

By construction, our firm classes strongly impact wages. A simple OLS regression shows, on Table 23, that, as compared to class 1, working in a class 2 firm increases wages by 5%, working in a class 3 firm by 10%, and in a class 4 firm by 20%. We observe that education levels correlate with firms classes. Table 24 indicates that in the 1998 cohort, 64% of Master's graduates and 79% of B & E school graduates worked in class 3 or 4 firms. In the 2010 cohort, these figures remained relatively stable at 61% and 75% respectively, despite the strong increase in the number of graduates and the stability of the distribution of firm classes each year. This means that the share of highly educated individuals has strongly increased in firms of class 3 and 4, more than in the total population. In these firms, increasing competition between educated individuals may be pushing wages downward.

As mobility across firm classes might explain higher returns to experience, we study how workers move across firm classes overtime and whether this mobility has evolved across cohorts. In Table 25, we show that over the first 7 years of career, 32% of workers have changed their firm class; 21% of them have changed once; 7% have changed twice and 3.5% three times or more. Moreover, 21% of the workers have moved at least once upward and 19% of them have moved at least once downward. Besides, mobility across firm classes has slightly decreased over time. In total, 36% of the individuals had moved at least once in the 1998 cohort, but only 31% in the 2010 cohort. Mobility has decreased in both directions. 24% had moved upward at least once in the 1998 cohort, but only 19% in the 2010 cohort; 22% had moved down in the 1998 cohort, but only 19% in the 2010 cohort.

Mobility across firm classes is heterogeneous across education levels (see Table 26 in Appendix M). In the 1998 cohort, we see that the higher the degree, the lower the mobility rate, both for upward moves and downward moves. In the 2010 cohort, mobility rates are distributed more uniformly both for upward and downward moves. In particular, mobility rates have strongly decreased for lower education levels and have slightly increased for higher education levels. The fact that mobility across firm classes has declined provides a partial explanation for the decreasing returns to experience. The latter may be in part a consequence of a decrease in sorting between firms and workers.

In the model variant, mobility is modeled to study how the sorting of workers across firms impacts the ATE and ATT of education. Figure 10 shows the ATE and ATT, based on wages observed in the seventh year of career, as a function of degree and cohort. The treatment effects are recomputed using the posterior probabilities of types of the model variant. A comparison with Figure 4 shows that the results of the base model and the variant are very close, leading to the same general conclusions. The probabilities of types $p(k|c)$ obtained in the variant are also very close to those obtained in the base model (see Table 27). Table 28 gives the estimated coefficients of the Ordered Probit determining firm class.

While the ATE and ATT are similar, we can provide further explanations for the decrease in the returns to higher-education degrees. Figure 11 shows the effects of firms classes in the wage equation of the model variant where these effects are allowed to vary with types and cohorts. We observe that *(i)* the effect of firm class strongly increases with the type, *i.e.*, best types are more productive in better firms and *(ii)* there is a clear decline in wage premia in class 4 overtime as shown by Figure 11(d). This drop mainly affects the top types.

With respect to mobility across firm classes, we observe in Table 29 that Type 3 is the most mobile, and Type 2 is always the least mobile. In 1998, over the observed 7 years of career, 36% of workers have experienced a change of firm class.[35] In 2010, the mobility gap between types has increased.[36] In 1998, over the first 7 years of career, 24% of workers have moved up the firm-class ladder.[37] In

[35] Among them, 31% were of Type 2, 36% of Type 1 and 43% of Type 3.

[36] In the 2010 cohort, 24% of the Type-2 workers have changed their firm-class; resp. 31% of Type 1 and 41% of Type 3.

[37] Among them, we find 22% of Type 2; 23% of Type 1 and 30% of Type 3.

2010 again, the gap between types has strongly increased,[38] yet, it seems that upward mobility has decreased with time.

The estimation results of the model variant with firm effects yield comparable returns to education and experience across cohorts and types as shown in Figure 12. Returns to experience for Master's degree holders remain stable for Types 1 and 3 but decline for Type 2, albeit to a lesser extent than in the main model. This suggests that part of the observed decline in the main model was due to sorting effects. Additionally, while the zero-experience wage (*i.e.*, return to education) decreases for Type 1; it slightly increases for Types 2 and 3, whereas in the main model, it was decreasing for Type 2. Hence, a significant portion of the decline in the zero-experience wage we found with the main model can be attributed to a drop in Type 2's return to working in a class-4 firm, a large part of Type-2 individuals being employed in these firms. Table 30 reveals that in the 1998 cohort, 83% of Type 1 individuals start to work in firms of class 1 or 2, whereas 53% of Type 2 individuals are employed in class 4 firms. Type 3 individuals are more evenly distributed across firm classes. Over time, both Types 2 and 3 have increasingly moved to class-4 firms. Tables 31 shows that both in the 1998 and 2010 cohorts, 95% of Type 2 students with either a Master's degree or a B & E school degree worked in class 4 firms, despite a significant increase in their numbers. Only 5% of them worked in class 3 firms, while 0% were employed in class 1 or class 2 firms. Considering now Type 3 individuals with a Master's degree, 67% worked in class 3 or 4 firms, while for those with a B & E school degree, 71% did so. These proportions have remained stable across cohorts.

These results are consistent with the congestion (or excess supply) effect, as described in Corblet (2024). The growing prevalence of Type 2 individuals pursuing Master's degrees has led them to enter the same firms as before, intensifying competition and thereby reducing both firm-class wage

[38] 14% of Type 2 workers have moved upward, resp., 20% of Type 1 and 25% of Type 3.

premia and returns to experience.

8 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. 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. From these results we compute policy-relevant parameters, such as the ATT and ATE of various education levels. 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 400 euros per month, if we define the high-school dropouts as the untreated. 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. 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.

References

Daron Acemoglu. Changes in unemployment and wage inequality: An alternative theory and some evidence. *American Economic Review*, 89(5):1259–1278, 1999.

Daron Acemoglu and David Autor. Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, volume 4, pages 1043–1171. Elsevier, 2011.

Elisabeth S. Allman, Catherine Matias, and John A. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics*, 37(6A):3099–30132, 2009. doi: 10.1214/09-AOS689.

Peter Arcidiacono and John Bailey Jones. Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, 71(3):933–946, 2003. doi: https://doi.org/10.1111/1468-0262.00431. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00431.

Peter Arcidiacono, V. Joseph Hotz, Arnaud Maurel, and Teresa Romano. Ex ante returns and occupational choice. *Journal of Political Economy*, 128(12):4475–4522, 2020. doi: 10.1086/710559. URL https://doi.org/10.1086/710559.

Peter Arcidiacono, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. College attrition and the dynamics of information revelation. *Journal of Political Economy*, 133(1):53–110, 2025.

Damiano Argan and Robert Gary-Bobo. Les diplômes français se sont-ils dévalorisés? *Revue économique*, 74(2):95–137, 2023.

Jared Ashworth, V Joseph Hotz, Arnaud Maurel, and Tyler Ransom. Changes across cohorts in wage returns to schooling and early work experiences. *Journal of Labor Economics*, 39(4):931–964, 2021.

David H Autor, Lawrence F Katz, and Melissa S Kearney. Trends in us wage inequality: Revising the revisionists. *The Review of Economics and Statistics*, 90(2):300–323, 2008.

José Azar and Ioana Marinescu. Monopsony power in the labor market: From theory to policy. *Annual Review of Economics*, 16(Volume 16, 2024):491–518, 2024. ISSN 1941-1391. doi: https://doi.org/10.1146/annurev-economics-072823-030431. URL https://www.annualreviews.org/content/journals/10.1146/annurev-economics-072823-030431.

Thi Hang Banh, Trang Hong Dao, Paul Glewwe, and Giang Thai. An investigation of the decline in the returns to higher education in vietnam. *Education Economics*, 32(5):665–685, 2024. doi: 10.1080/09645292.2024.2362262. URL https://doi.org/10.1080/09645292.2024.2362262.

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

Paul Beaudry, David A Green, and Benjamin M Sand. The great reversal in the demand for skill and cognitive tasks. *Journal of Labor Economics*, 34(S1):S199–S247, 2016.

Magali Beffy, Denis Fougère, and Arnaud Maurel. Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter? *The Review of Economics and Statistics*, 94(1):334–347, 02 2012. ISSN 0034-6535. doi: 10.1162/REST_a_00212. URL https://doi.org/10.1162/REST_a_00212.

Christian Belzil and Jorgen Hansen. Reconciling changes in wage inequality with changes in college selectivity, using a behavioral model. Institut Polytechnique de Paris, 2020.

Richard Blundell, David A. Green, and Wenchao Jin. The u.k. as a technological follower: Higher education expansion and the college wage premium. *Review of Economic Studies*, 89:142–180, 2022.

Stéphane Bonhomme, Koen Jochmans, and Jean-Marc Robin. Estimating multivariate latent-structure models. *Annals of Statistics*, 44(2):540–563, April 2016. ISSN 0090-5364, 2168-8966. doi: 10.1214/15-AOS1376. URL https://projecteuclid.org/journals/annals-of-statistics/volume-44/issue-2/Estimating-multivariate-latent-structure-models/10.1214/15-AOS1376.full.

Stéphane Bonhomme and Elena Manresa. Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184, 2015. doi: https://doi.org/10.3982/ECTA11319. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA11319.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. A Distributional Framework for Matched Employer Employee Data. *Econometrica*, 87(3):699–739, May 2019. doi: 10.3982/ECTA15722. URL https://ideas.repec.org/a/wly/emetrp/v87y2019i3p699-739.html.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. Discretizing unobserved heterogeneity. *Econometrica*, 90(2):625–643, 2022. doi: https://doi.org/10.3982/ECTA15238. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA15238.

Charles Bouveyron, Gilles Celeux, T Brendan Murphy, and Adrian E Raftery. *Model-based clustering and classification for data science: with applications in R*, volume 50. Cambridge University Press, 2019.

Thomas Brodaty, Robert J. Gary-Bobo, and Ana Prieto. Do risk aversion and wages explain educational choices? *Journal of Public Economics*, 117:125–148, 2014. ISSN 0047-2727. doi: https://doi.org/10.1016/j.jpubeco.2014.05.003. URL https://www.sciencedirect.com/science/article/pii/S0047272714001108.

David Card. Who set your wage? *American Economic Review*, 112(4):1075–90, April 2022. doi: 10.1257/aer.112.4.1075. URL https://www.aeaweb.org/articles?id=10.1257/aer.112.4.1075.

David Card and Thomas Lemieux. 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, 2001.

Pedro Carneiro and Sokbae Lee. Trends in quality-adjusted skill premia in the united states, 1960-2000. *American Economic Review*, 101(6):2309–49, 2011.

Pedro Carneiro, Karsten T. Hansen, and James J. Heckman. 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, 2003. doi:

https://doi.org/10.1111/1468-2354.t01-1-00074. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-2354.t01-1-00074.

Pedro Carneiro, James J. Heckman, and Edward J. Vytlacil. Estimating marginal returns to education. *American Economic Review*, 101(6):2754–81, October 2011. doi: 10.1257/aer.101.6.2754. URL https://www.aeaweb.org/articles?id=10.1257/aer.101.6.2754.

Oliver Cassagneau-Francis. The returns to higher education by cognitive and non-cognitive abilities. Technical report, UCL, London, November 2021.

Oliver Cassagneau-Francis, Robert Gary-Bobo, Julie Pernaudet, and Jean-Marc Robin. A non-parametric finite mixture approach to difference-in-difference estimation, with an application to on-the-job trainig and wages. Technical report, Sciences Po, Paris, 2021.

Pauline Corblet. Higher-educated workers' flattening returns to experience. *Available at SSRN 5018246*, 2024.

Flavio Cunha and James J. Heckman. Identifying and estimating the distributions of ex post and ex ante returns to schooling. *Labour Economics*, 14(6):870–893, 2007. ISSN 0927-5371. doi: https://doi.org/10.1016/j.labeco.2007.06.002. URL https://www.sciencedirect.com/science/article/pii/S0927537107000383. Education and Risk.

Flavio Cunha, James Heckman, and Salvador Navarro. Separating uncertainty from heterogeneity in life cycle earnings. *Oxford Economic Papers*, 57(2):191–261, 04 2005. ISSN 0030-7653. doi: 10.1093/oep/gpi019. URL https://doi.org/10.1093/oep/gpi019.

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

William R Emmons, Ana HernÃ Kent, Lowell Ricketts, et al. Is college still worth it? the new calculus of falling returns. *Federal Reserve Bank of St Louis Review*, 101(4):297–329, 2019.

Robert J Gary-Bobo, Marion Goussé, and Jean-Marc Robin. Grade retention and unobserved heterogeneity. *Quantitative Economics*, 7(3):781–820, 2016.

Claudia Goldin and Lawrence F Katz. *The race between education and technology.* Harvard University Press, 2008.

Fatih Guvenen. Learning your earning: Are labor income shocks really very persistent? *American Economic Review*, 97(3):687–712, June 2007. doi: 10.1257/aer.97.3.687. URL https://www.aeaweb.org/articles?id=10.1257/aer.97.3.687.

Fatih Guvenen, Fatih Karahan, Serdar Ozkan, and Jae Song. What do data on millions of u.s. workers reveal about lifecycle earnings dynamics? *Econometrica*, 89(5):2303–2339, 2021. doi: https://doi.org/10.3982/ECTA14603. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA14603.

Peter Hall and Xiao-Hua Zhou. Nonparametric estimation of component distributions in a multivariate mixture. *The Annals of Statistics*, 31(1):201 – 224, 2003. doi: 10.1214/aos/1046294462. URL https://doi.org/10.1214/aos/1046294462.

Eric A Hanushek, Yuan Wang, and Lei Zhang. Understanding trends in chinese skill premiums, 2007-2018. Working Paper 31367, National Bureau of Economic Research, June 2023. URL http://www.nber.org/papers/w31367.

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

James J. Heckman and Edward Vytlacil. Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73(3):669–738, May 2005. URL https://ideas.repec.org/a/ecm/emetrp/v73y2005i3p669-738.html.

James J. Heckman, John Eric Humphries, and Gregory Veramendi. Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy*, 126(S1):S197–S246, 2018. doi: 10.1086/698760. URL https://doi.org/10.1086/698760.

Mark Huggett, Gustavo Ventura, and Amir Yaron. Sources of lifetime inequality. *American Economic Review*, 101(7):2923–2954, 2011. ISSN 00028282. URL http://www.jstor.org/stable/41408726.

Andrea Ichino, Aldo Rustichini, and Giulio Zanella. College, cognitive ability, and socioeconomic disadvantage: policy lessons from the uk in 1960-2004. Technical report, 2024.

Hyeok Jeong, Yong Kim, and Iourii Manovskii. The price of experience. *American Economic Review*, 105 (2):784–815, 2015.

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

Hiroyuki Kasahara and Katsumi Shimotsu. Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices. *Econometrica*, 77:135–175, 2009.

Lawrence F Katz and Kevin M Murphy. Changes in relative wages, 1963–1987: supply and demand factors. *The Quarterly Journal of Economics*, 107(1):35–78, 1992.

Rasmus Lentz, Suphanit Piyapromdee, and Jean-Marc Robin. The anatomy of sorting—evidence from danish data. *Econometrica*, 91(6):2409–2455, 2023.

Lance Lochner, Youngmin Park, and Youngki Shin. The evolution of unobserved skill returns in the u.s.: A new approach using panel data. Working Paper 33407, National Bureau of Economic Research, January 2025. URL http://www.nber.org/papers/w33407.

Thierry Magnac and David Thesmar. Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816, 2002. doi: https://doi.org/10.1111/1468-0262.00306. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00306.

Thierry Magnac, Nicolas Pistolesi, and Sébastien Roux. Post-schooling human capital investments and the life cycle of earnings. *Journal of Political Economy*, 126(3):1219–1249, 2018. doi: 10.1086/697206. URL https://doi.org/10.1086/697206.

G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley series in probability and statistics: Applied probability and statistics. Wiley, 2000. ISBN 9780471654063. URL https://books.google.fr/books?id=7M5vK8OpXZ4C.

Elio Nimier-David. Local human capital and firm creation: Evidence from the massification of higher education in france. *Unpublished Manuscript*, 2023.

OECD. Education gps. *http://gpseducation.oecd.org*, 2025.

Sherwin Rosen. The theory of equalizing differences. volume 1 of *Handbook of Labor Economics*, chapter 12, pages 641–692. Elsevier, 1986. doi: https://doi.org/10.1016/S1573-4463(86)01015-5. URL https://www.sciencedirect.com/science/article/pii/S1573446386010155.

SIES. L'état de l'enseignement supérieur, de la recherche et de l'innovation en france. Technical report, MESR, Paris, May 2022. URL https://www.enseignementsup-recherche.gouv.fr/fr/l-etat-de-l-enseignement-superieur-de-la-recherche-et-de-l-innovation-en-france-2022-84954.

UNESCO. Higher education figures at a glance. *UNESCO, Institute for statistics*, 2022. URL https://uis.unesco.org/sites/default/files/documents/f_unesco1015_brochure_web_en.pdf.

Robert G Valletta. 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*, pages 313–342. University of Chicago Press, 2018.

Gregory Verdugo. The great compression of the french wage structure, 1969–2008. *Labour Economics*, 28:131–144, 2014.

A Appendix. Construction of the Sample

In this work we exploit the CEREQ surveys called *Enquête Generation 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 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. 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: 60,917 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 7 lists the degree types that have been aggregated in each of the categories used for estimation, the definition of the Urban, Peri-Urban and Rural areas used to construct the corresponding indicators and the definition we use for the profession of the father. 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 7: Definition of variables

Education level	
	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, 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

Continues on the next page...

	other Bac+2
	Bac+2/3,
	Licence pro
	L3, other Bac+3
	M1, Bac+4
Masters	M2 Humanities Business adm. Law
	M2 Maths Sciences Technology
	other Bac+5
Bus. and Engin. Sch. Deg.	Bac+5 Business Schools, Engineering schools
Residence area	
	1998 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
Occupation of the Father	
Not a "professional"	Farmer, Craftsman, Storekeeper, Entrepreneur,

Continues on the next page...

	Technician, Foreman, Salesman, Associate professional, White collar worker, Blue collar worker, unknown
Father is a "professional"	Executive, Engineer, Learned profession, Professor

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

B Appendix. Descriptive Statistics

Table 8: Descriptive Statistics

Cohort	Education	Individuals	% (Weighted)	Observ.	Average Rate of Employ^t	Average Full-Time Real Log-Wage
1998	Less than High School	3026	0.42	35285	0.65	7.16
	High-School Degree	1815	0.25	19216	0.64	7.23
	Some College, Bachelors	2038	0.24	20902	0.64	7.39
	Masters	203	0.04	1826	0.64	7.72
	Bus. and Engin. Sch.	301	0.05	2777	0.69	7.85
	All	7383	100	80006	0.65	7.28
2004	Less than High School	1601	0.41	20475	0.61	7.23
	High-School Degree	1406	0.23	16038	0.65	7.27
	Some College, Bachelors	1628	0.25	16580	0.68	7.39
	Masters	470	0.05	4363	0.70	7.65
	Bus. and Engin. Sch.	395	0.06	3461	0.73	7.80
	All	5500	100	60917	0.65	7.36
2010	Less than High School	870	0.36	10855	0.50	7.22
	High-School Degree	917	0.28	10538	0.58	7.27
	Some College, Bachelors	964	0.22	9554	0.65	7.40
	Masters	419	0.08	3869	0.67	7.62
	Bus. and Engin. Sch.	351	0.06	2673	0.76	7.78
	All	3521	100	37489	0.60	7.39

Note. Column '% Weighted' gives the weighted share of individuals in their respective cohorts, 1998, 2004 or 2010. We used the CEREQ *Generation* survey weights to compute these percentages.

Table 9: EVOLUTION OF THE NUMBER OF GRADUATES BY COHORT (*GENERATION SURVEYS*)

Number of graduates (highest degree granted)				
	1992	1998	2004	2010
Dropouts	171 145	145 937	123 098	118 839
Secondary Vocational	107 184	118 923	127 450	102 315
High-School degree	70 767	109 172	176 965	204 064
2 years of College	99 550	228 611	149 156	127 463
3 years of College	17 686	24 328	36 189	27 517
Master 1	18 605	30 867	36 224	9 087
Master 2	26 076	39 910	43 426	69 387
Engineering schools	9 003	8 494	7 836	10 385
Business schools	11 274	16 178	22 796	20 846

Note. Each column gives the weighted total of individuals in their respective cohorts. Each line corresponds to an education level. The cohorts are the Generation surveys 1992, 1998, 2004 or 2010. Generation 1992 appears here to give a broader picture of higher education in France but is not used to estimate our model. We used the CEREQ *Generation* survey weights to compute these totals. Survey weights are such that weighted totals are equal to the effective count of graduates in the cohort, as computed by the Ministry for Higher education and Research, *i.e.*, the MESR. *Secondary Vocational* includes CAP and BEP; High-school degree is the count of *baccalauréats*; *2 years of college* includes the equivalent of Associate's, including DUT and BTS; *3 years of College* includes the French bachelors (the L3 or *Licence*). The increase in the figures of the Master 2 line (5 years of higher education) from one cohort to the next, is particularly striking.

C Appendix. 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,

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

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

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

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:

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

where $Q_{itg} = 1$ if $e_{it} = \mathbf{e}_g$ and 0 otherwise 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 that are non time-varying $\bar{Z}_i$ in Z_{it} and latent type k as follows,

$$\Lambda_k(h|\bar{Z}_i) = \frac{\exp(\mathbf{v}_{ihk})}{\sum_{j=1}^{H} \exp(\mathbf{v}_{ijk})}.$$

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. We denote E_i the subset of dates t at which e_{it} is observed. This model is estimated with observations e_{it} at the beginning and the end of each employment spell of individual i.[39] 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,

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

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,

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

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

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

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

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

The *posterior* probability that individual i is of type k is denoted p_{ik}; it can be expressed with the

[39]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.

help of Bayes' rule and the likelihood, as follows,

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

Variant with firm effects. The variant of our model with firm effects and the likelihood function of this model are relatively straightforward extensions of the above model and likelihood. Let $\Pr(j_{it} \in Q_f | c, h_i, k)$ be the probability that firm j_{it} is in the firm class Q_f, with $f = 1, \ldots 4$, conditional on cohort c, education h_i and type k, where j indexes firms and j_{it} is the employer of i at time t. Experience is not included in the list of variables. Hence, the matching with firm classes crucially depends on education, type and their interaction. The indicators of subsets Q_f, $f = 1, \ldots, 4$, interacted with education, cohort and type, are now variables explaining wages. To obtain the likelihood function of the variant with firm effects, under the same conditional independence assumptions, we add the product of the probabilities of choosing the observed firm class, conditional on k, on all employment spells, as an additional factor in the expression for L_{ik}.

D Appendix. 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 et al. (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 $G(.|k)$, $k = 1, \ldots, K$, are linearly independent.[40] The latter condition is reasonable if types are

[40]See particularly Theorem 8 in Allman et al. (2009). On this topic, further results are proved and estimation

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.

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[41] 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

methods are provided in Bonhomme et al. (2016).

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

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.

E Appendix. Choice of the Number of Types K

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 $K + 1$.[42] AIC tends to overestimate the correct number of components (AIC pushes towards over-fitting). BIC corrects for these difficulties but tends to underestimate K. These criteria are useful, but they do not 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.

We estimated the model for different values of K and looked at different criteria, including entropy,

[42]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 $BIC = q \ln(N) - 2\mathcal{L}$.

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 10 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,

$$\mathcal{E}(K) = -\sum_{i=1}^{N}\sum_{k=1}^{K} \hat{p}_{ik}\ln(\hat{p}_{ik}), \tag{27}$$

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.[43] Entropy is minimal (and equal to zero) when partitioning is perfect.[44] 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,

$$\text{NEC}(K) = \frac{\mathcal{E}(K)}{\mathcal{L}(K) - \mathcal{L}(1)}. \tag{28}$$

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

$$H(K) = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} \hat{p}_{ik}^2 \tag{29}$$

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. A normalized index can be 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.

[43]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 $1/K$.

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

Table 10: 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 10 are derived from EM estimations of the full model with $K = 1, 2, 3$ and 4 types.

The most important information shown by Table 10 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.

F Appendix. Simulations

The simulations can be decomposed in a few steps.

Step 1. We first recursively simulate the employment level $\tilde{e}_{itk}$ for each (i, t), and each $k = 1, ..., K$, $t = 1, \ldots, T$ with $K = 3$ and $T = 90$. 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 $e_g \in \{0, .3, .5, .8, 1\}$. We then 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),

$$\tilde{w}_{itk} = \mathbb{E}[w_{itk}|\tilde{x}_{itk}, X_i] = \alpha_{0k} + \beta_{0k}\tilde{x}_{it} + \gamma_{0kh} + Z_i\eta_{0k}.$$

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 $t \in \{1, \dots, T\}$. 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[45]

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.

G Appendix. Full tables. Parameter Estimates

[45] 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 $(\delta^{t-1}(1-\delta))/(1-\delta^T)$.

Table 11: Wage Equation I. 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)

Note. The table gives the ML estimated values of the model's wage-equation coefficients (ML standard deviations in parentheses) that are related to experience. All the coefficients are interactions of effective experience with degrees, cohorts and types. Each column corresponds to a different type. Almost all of these coefficients are estimated with precision, with p-values below 1%. The other parameters of the wage equation are displayed in the next tables. Estimation is based on a sample including 15,841 individuals with full-time jobs and 105,496 observations in total.

Table 12: Wage Equation II. 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)

Note. The table gives the coefficients of the log-wage equation, estimated by ML (with ML standard deviations in parentheses) that are related to returns to education, *i.e.*, the zero-experience wages, for each education level, cohort and type. 'Dropouts' (*i.e.*, *less than high school degree* individuals are the reference). Each of the three columns corresponds to a different type. With a few exceptions, these coefficients are estimated with good precision and significant at the 1% level. The only non-significant coefficients are Type 1×1998×High-School degree; Type 3×2004×High-School degree; and Type 3×2010×High-School degree, meaning that these categories do not earn more than the reference. Estimation is based on a sample including 15,841 individuals with full-time jobs and 105,496 observations in total.

Table 13: 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)

Note. The table gives the ML estimated coefficients of controls in the log-wage equation (ML standard deviations in parentheses). 1998 urban students whose fathers are not professionals are the reference. The coefficients associated with the three types appear in three different columns. The unemployment rate affects Type 2 and Type 3 negatively. Estimation is based on a sample including 15,841 individuals with full-time jobs and 105,496 observations in total.

Note of Table 14. The table gives the coefficients of the type-dependent linear combinations of variables and their standard deviations, estimated by the Maximum Likelihood algorithm. The three columns give the coefficients of the linear combination of variables for each of the three types. The variables are cohort indicators and the cohort indicators interacting with education level dummies, plus a number of controls. There are 6 discrete values of the employment rate, $\{0, .3, .5, .6, .8, 1\}$, and therefore 5 estimated cuts for each type.

Note of Table 15. The table gives the estimated values and the standard deviations of the Multinomial Logit coefficients explaining the choice of individual education levels, estimated by the ML algorithm, simultaneously with the wage equation and the Ordered Probit explaining employment. All coefficients are type-dependent; the three columns give the coefficients for each of the three types. The choice of each education level is explained by cohort indicators, the 'father-is-a-professional' dummy, and indicators for residence in urban or rural zones. The 'below high-school degree' category is the reference choice.

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)

Table 15: 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)

H Appendix. Results obtained with the Elastic Net Method

The Elastic Net is a regularization technique used in linear regression that blends the strengths of Lasso (L1 regularization) and Ridge (L2 regularization) to improve model performance, especially when dealing with high-dimensional data or correlated features. It achieves this by combining both L1 and L2 penalties in its loss function, encouraging sparse solutions while preserving groupings of correlated variables. The objective function for the Elastic Net minimizes the residual sum of squares subject to the combination of L1 and L2 penalties, expressed as:

$$\text{minimize } \frac{1}{2N}\sum_{i=1}^{N}(y_i - \mathbf{X}_i \cdot \beta)^2 + \lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p}\beta_j^2$$

Here, y_i are the observed values, $\mathbf{X}_i$ are the predictors, β represents the coefficients, λ_1 controls the L1 regularization (Lasso), and λ_2 controls the L2 regularization (Ridge). By tuning these parameters, Elastic Net finds a balance between the two penalties, offering flexibility and robustness in feature selection and model accuracy.

Table 16 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); 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.

Table 16: 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			

Table 17: Grade at Final High-school Exam by Type in the 2010 Cohort

	$\leq$ Grade D	Grade C	Grade B	Grade A
Type 2	-0.0560***	0.0195	0.0287**	0.00776
	(0.0137)	(0.0127)	(0.00877)	(0.00398)
Type 3	-0.0807***	0.00378	0.0487***	0.0282***
	(0.0150)	(0.0139)	(0.00961)	(0.00436)
Const	0.605***	0.292***	0.0912***	0.0115***
	(0.00917)	(0.00849)	(0.00585)	(0.00266)

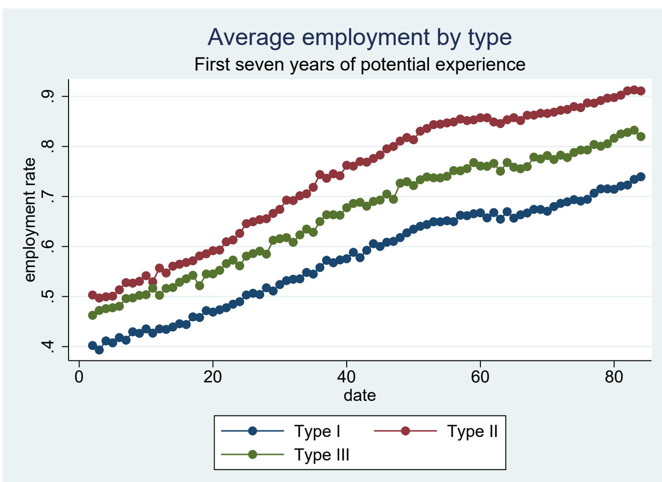
Note. We observe the *baccalauréat* grades only in the 2010 cohorts. N = 2322. The table presents the regressions of grades on types, Type 1 being the reference.

I Appendix. Employment characteristics by type.

Table 18: Employment characteristics by Type and Cohort

		Cohort	Overall	Type 1	Type 2	Type 3
Average Number of Employment Spells		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
Share in Public Sector		1998	.19	.18	.22	.15
		2004	.17	.19	.17	.11
		2010	.20	.22	.22	.15
Firm size	Small Firms $[1, 9]$	1998	.26	.34	.22	.19
	Medium Firms $[10, 49]$		.27	.31	.24	.24
	Large Firms ≥ 50		.47	.36	.54	.57
	Small Firms $[1, 9]$	2004	.26	.33	.23	.21
	Medium Firms $[10, 49]$		.28	.30	.28	.25
	Large Firms ≥ 50		.46	.37	.49	.54
	Small Firms $[1, 9]$	2010	.25	.32	.21	.20
	Medium Firms $[10, 49]$		.24	.25	.24	.21
	Large Firms ≥ 50		.51	.43	.55	.59

Figure 8: Employment Rates by Type: Simulations



Note. The simulations of expected wages are also giving simulations of the employment rates e_{it} as a by-product. The Figure gives the average employment rates conditional on type, as a function of time in months.

J Appendix. ATE of Experience and Education.

Figure 9: ATE of Experience by Cohort and Level of Education

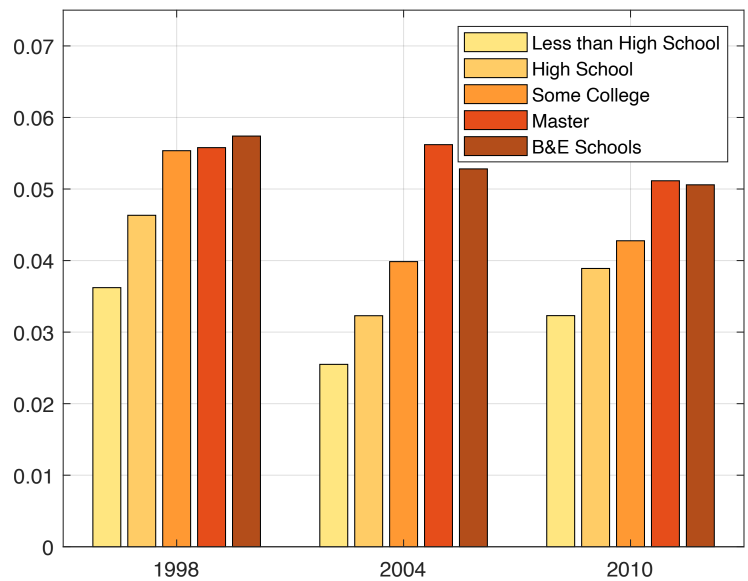


Table 19: ATE of experience and education on wages, employment and starting wage.

Cohort	Education	ATE of Effective Experience	ATE of education on employment	on starting wage
1998	Less than High School	0.036	0.73	1196
	High School	0.046	0.74	1245
	Some College	0.055	0.75	1408
	Master	0.056	0.76	2026
	B&E Schools	0.057	0.80	2039
2004	Less than High School	0.025	0.68	1416
	High School	0.032	0.73	1442
	Some College	0.040	0.75	1543
	Master	0.056	0.75	1806
	B&E Schools	0.053	0.80	2031
2010	Less than High School	0.032	0.54	1398
	High School	0.039	0.64	1431
	Some College	0.043	0.70	1580
	Master	0.051	0.69	1785
	B&E Schools	0.051	0.81	2275

Note. To obtain the ATE of effective experience in this table, we compute the weighted average of the type-dependent yearly returns to experience for each education level, using the coefficients from the log-wage equation. The weights are the estimated prior probabilities of each type, reflecting the distribution of types in the overall population of the cohort. The ATEs are expressed as yearly returns. To obtain the ATE of education on employment, we use the same method described in Section 6.1. but instead of the average wage in the seventh year we use the average employment rate over the seventh year of an individual. To obtain the ATE of education at zero experience in this table, we compute the weighted average of the type-dependent returns to education, using the coefficients from the log-wage equation. The weights are the estimated prior probabilities of each type, reflecting the distribution of types in the overall population of the cohort. The ATEs are expressed as exponentials of log wages.

K Appendix. Education Level and Father Occupation

Table 20: Probability of Reaching Education Level h by Type k.

		Non Professional Father			Professional Father		
		Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
1998	< High School	0.47	0.42	0.50	0.23	0.14	0.13
2004		0.40	0.31	0.30	0.16	0.10	0.08
2010		0.34	0.29	0.24	0.12	0.09	0.06
1998	High School	0.25	0.26	0.22	0.28	0.19	0.22
2004		0.30	0.26	0.25	0.27	0.15	0.14
2010		0.32	0.21	0.29	0.29	0.16	0.16
1998	Some College	0.25	0.28	0.22	0.35	0.41	0.44
2004		0.24	0.29	0.30	0.36	0.37	0.38
2010		0.23	0.33	0.22	0.28	0.38	0.26
1998	Master	0.02	0.01	0.03	0.07	0.06	0.07
2004		0.03	0.07	0.10	0.14	0.15	0.22
2010		0.06	0.11	0.14	0.14	0.22	0.27
1998	B & E schools	0.01	0.03	0.03	0.07	0.20	0.14
2004		0.02	0.08	0.05	0.07	0.22	0.18
2010		0.05	0.06	0.12	0.17	0.14	0.26

Note. The table gives the estimated values of $p(h|k, c, Z)$ where $Z = 1$ when the father is a professional (and $Z = 0$ otherwise).

Table 21: Distribution of Type by Education and Cohort

		Non Professional Father			Professional Father		
		Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
1998	< High School	0.42	0.33	0.25	0.46	0.32	0.22
2004		0.49	0.33	0.18	0.54	0.30	0.16
2010		0.52	0.32	0.16	0.54	0.32	0.15
1998	High School	0.42	0.38	0.20	0.42	0.31	0.27
2004		0.46	0.35	0.19	0.56	0.28	0.16
2010		0.53	0.25	0.22	0.57	0.25	0.19
1998	Some College	0.40	0.40	0.20	0.30	0.39	0.31
2004		0.38	0.39	0.23	0.40	0.36	0.24
2010		0.41	0.41	0.18	0.38	0.41	0.21
1998	Master	0.41	0.24	0.35	0.35	0.36	0.29
2004		0.23	0.43	0.34	0.36	0.34	0.30
2010		0.28	0.39	0.32	0.29	0.36	0.35
1998	B & E schools	0.24	0.46	0.29	0.18	0.54	0.28
2004		0.19	0.60	0.21	0.19	0.53	0.28
2010		0.35	0.29	0.36	0.38	0.25	0.36

Note. The table gives the estimated values of $p(k|h, c, Z)$ where $Z = 1$ when the father is a professional (and $Z = 0$ otherwise).

L Appendix. Cohort-by-Cohort Estimation

Table 22: 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

Note. The table permits a comparison of the main model, estimated with the three cohorts stacked, and the same model, estimated separately with the subsamples of individuals belonging to each of the three cohorts. Up to type relabeling, we always find correlation matrices with a strong positive correlation on the diagonal and negative values for off-diagonal correlations, when we consider in turn the models estimated with the three subsamples.

M Appendix. Worker-firm Matching

Table 23: Log-Wage Equations with Firm Classes and Types

	(1)	(2)	(3)	(4)
High School	0.0545***	0.0548***	0.0551***	0.0573***
	(0.0012)	(0.0011)	(0.0011)	(0.0018)
Some college	0.1906***	0.1782***	0.1833***	0.1677***
	(0.0012)	(0.0012)	(0.0011)	(0.0018)
Master	0.4272***	0.3827***	0.3974***	0.3484***
	(0.0021)	(0.0021)	(0.0020)	(0.0032)
B & E schools	0.5716***	0.4978***	0.5184***	0.4910***
	(0.0021)	(0.0021)	(0.0020)	(0.0032)
Experience	0.0037***	0.0036***	0.0037***	0.0036***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Professional Father	0.0334***	0.0283***	0.0302***	0.0223***
	(0.0012)	(0.0011)	(0.0011)	(0.0018)
Peri-urban	-0.0192***	-0.0053***	-0.0080***	-0.0083***
	(0.0012)	(0.0011)	(0.0011)	(0.0018)
Rural	-0.0292***	-0.0049***	-0.0103***	-0.0084***
	(0.0011)	(0.0011)	(0.0010)	(0.0017)
Cohort 2004	0.0262***	0.0308***	0.0308***	0.0393***
	(0.0010)	(0.0010)	(0.0010)	(0.0016)
Cohort 2010	0.0421***	0.0403***	0.0446***	0.0557***
	(0.0014)	(0.0014)	(0.0013)	(0.0022)
Unemployment	-0.0113***	-0.0109***	-0.0111***	-0.0115***
	(0.0006)	(0.0006)	(0.0006)	(0.0009)
Class 2			0.0462***	0.0109***
			(0.0012)	(0.0020)
Class 3			0.0993***	0.0418***
			(0.0012)	(0.0022)
Class 4			0.1938***	0.1617***
			(0.0013)	(0.0025)
Type 1				-0.0403***
				(0.0021)
Type 3				0.2067***
				(0.0020)
Constant	7.1632***	7.0744***	7.0764***	7.0867***
	(0.0053)	(0.0087)	(0.0052)	(0.0084)
R-squared	0.38	0.44	0.43	0.52

Note: Log-wages is the dependent variable. Standard errors in parentheses. Significance is indicated by: *** p<0.01, ** p<0.05, * p<0.1. Column 2 differs from column 3 because firm characteristics are added as controls. Column 3 uses our firm classes as controls instead of observed firm characteristics. In column (4), individuals are replicated three times, each with a different type and the regression is weighted by the posterior probabilities of types. The number of observations is 106,009.

Table 24: Distribution of Education Levels by Firm Class and Cohort, in %

	Cohort 1998						Cohort 2010					
	≤ HS	HS	College	Master	School	All	≤ HS	HS	College	Master	School	All
Class 1	26.59	29.77	22.17	13.44	7.40	25.16	27.44	29.33	24.86	19.03	8.55	24.37
Class 2	25.80	22.70	28.90	22.29	13.00	25.30	24.88	23.93	22.53	19.84	16.30	22.52
Class 3	26.76	25.09	22.61	21.16	21.14	24.93	27.83	23.64	25.76	22.11	21.10	24.79
Class 4	20.86	22.44	26.32	43.10	58.46	24.61	19.86	23.10	26.86	39.03	54.05	28.32

Note. School means Business or Engineering School. HS means high school degree. ≤ HS means less than high school. College means 'Some College and Bachelors'.

Table 25: Mobility Rates I. Firm-Class Transitions. Comparison of Cohorts

	1998	2004	2010	All
No move	64.2	73.5	69.1	68.4
1 move	23.8	17.9	22.2	21.5
2 moves	7.9	5.1	6.5	6.7
3 moves or more	4.1	3.5	2.2	3.4
At least 1 move up	24.2	17.3	19.2	21.0
At least 1 move down	21.6	16.4	18.5	19.2

Table 26: Mobility rates II. Firm-Class Transitions by Education Level

	1998					2010				
	≤ HS	HS	College	Master	School	≤ HS	HS	College	Master	School
At least 1 move	39.79	34.10	33.27	28.57	26.91	28.05	31.84	31.43	32.70	32.19
At least 1 move up	28.35	22.59	21.30	17.73	16.94	19.77	20.39	18.05	17.66	19.37
At least 1 move down	24.42	20.77	19.73	15.76	15.95	16.32	19.19	18.78	19.81	19.94

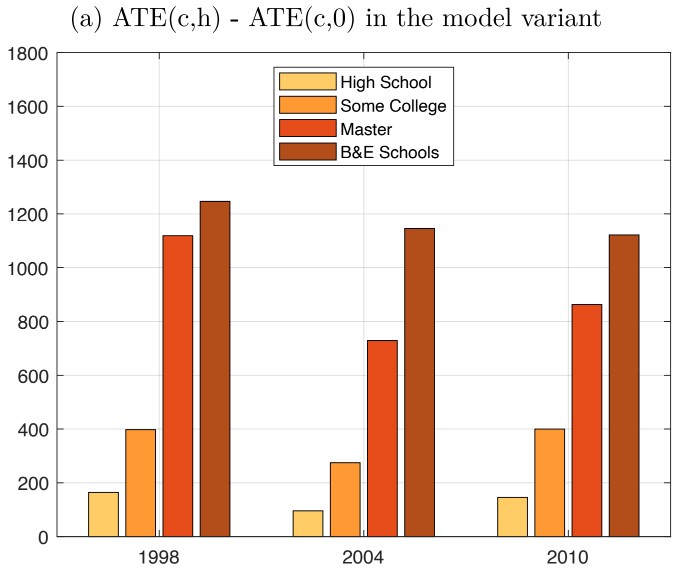
Note. School means Business or Engineering School. HS means high school degree. ≤ HS means less than high school. College means 'Some College and Bachelors'.

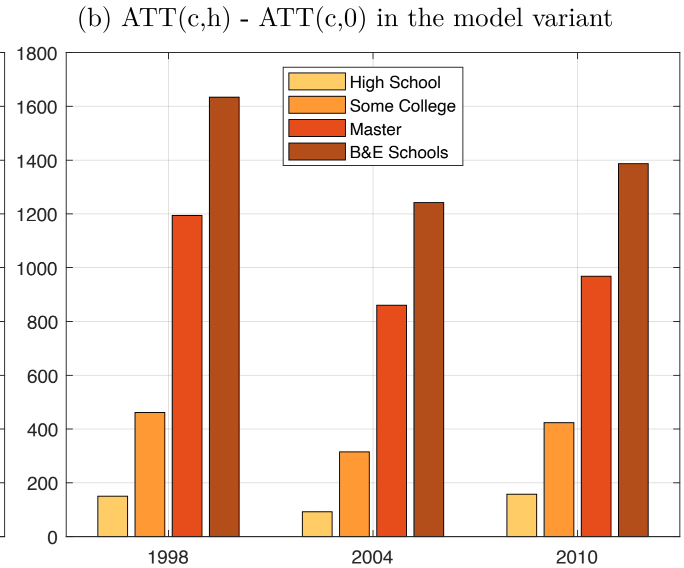
Table 27: Distribution of Types across Cohorts

	Base Model				Variant Model			
	1998	2004	2010	All	1998	2004	2010	All
Type 1	0.40	0.42	0.45	0.42	0.42	0.43	0.44	0.43
Type 2	0.36	0.37	0.33	0.36	0.34	0.34	0.34	0.34
Type 3	0.23	0.21	0.22	0.22	0.24	0.23	0.22	0.23

Note. Estimated frequency of types $\hat{p}(k|c)$, by cohort, in the main model and in the variant model.

Figure 10: Robustness Check: ATE and ATT when Education is the Treatment and Full-time Wage is the Outcome, in the Variant with Firm Effects





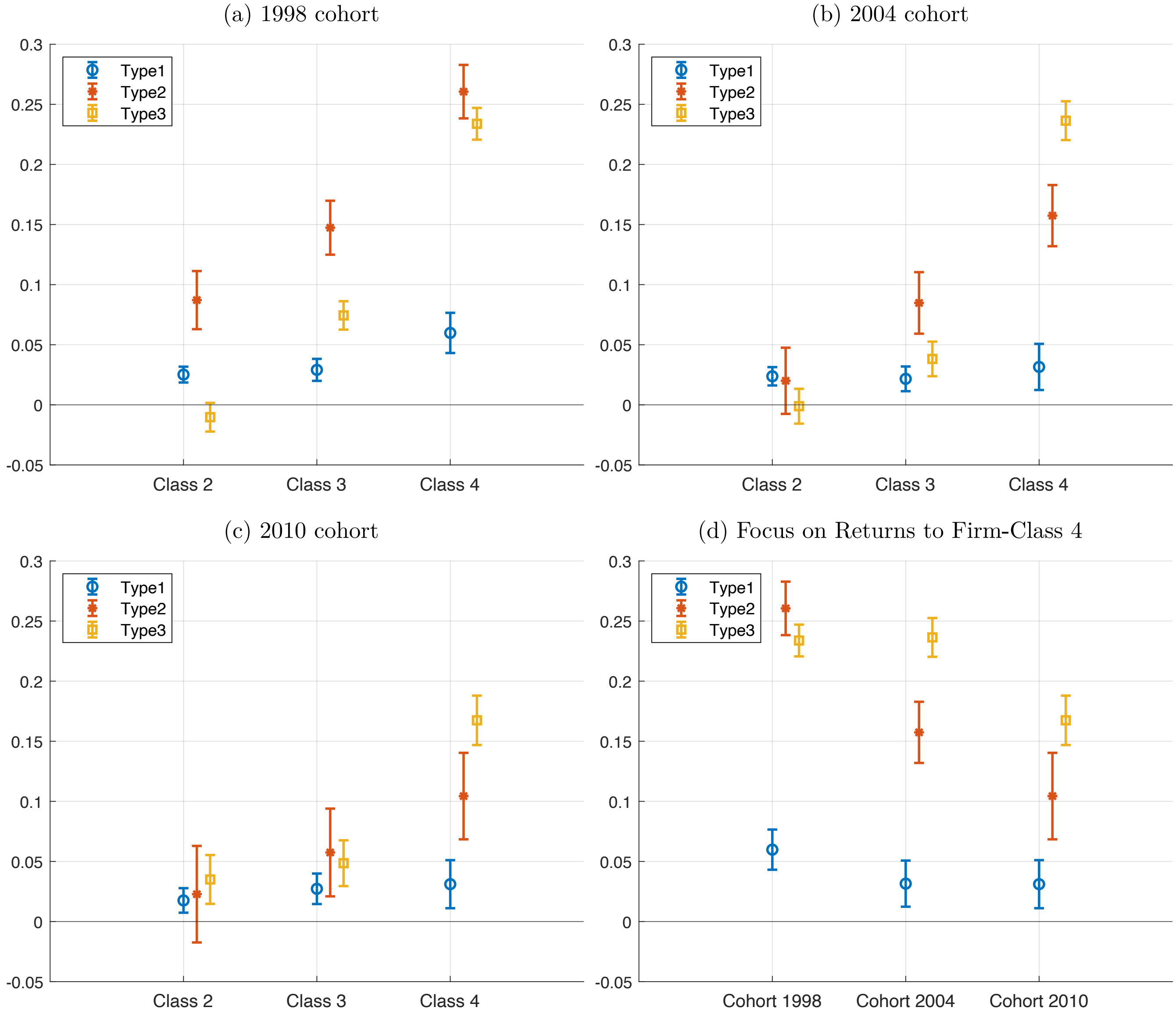
Note. The ATEs and ATTs, when wages observed in the seventh year of career are the outcome and education is the treatment, are recomputed with the new posterior probabilities obtained in the model variant. This variant has firm effects in the wage equation, 4 firm classes, and a type-dependent equation determining the firm class of each employment spell. Results must be compared with the corresponding base model values presented in Figure 4. The conclusions are essentially the same; the differences are small.

Table 28: Ordered Probit for the Choice of Firm Class

2004 cohort	-0.0349
	(0.0079)
2010 cohort	0.0659
	(0.0096)
High-school Degree	-0.0463
	(0.0090)
Some College and Bachelors	0.1292
	(0.0088)
Masters	0.5370
	(0.0159)
Bus. Engin. School Degrees	0.8202
	(0.0167)
Type 2	2.0063
	(0.0091)
Type 3	0.9127
	(0.0095)
cut1	-1.9992
	(0.0099)
cut2	-1.0571
	(0.0090)
cut3	-0.0723
	(0.0084)

Note. The table gives the estimation results of the Ordered Probit determining the choice of firm class.

Figure 11: Returns to Firm Class by Type and Cohort



Note. The figure depicts the estimated coefficients and confidence intervals of firm-class indicators (*i.e.*, firm effects) in the variant of our model including firm effects. Class 1 is the reference.

Table 29: Mobility Rates by Type and Cohort, in %

	Cohort 1998			Cohort 2010		
	Type 2	Type 1	Type 3	Type 2	Type 1	Type 3
At least 1 move	30.86	35.73	42.85	24.04	31.12	41.02
At least 1 move up	22.24	22.93	29.36	14.38	20.12	24.55
At least 1 move down	16.30	23.50	26.12	14.05	18.11	26.13

Note. In this table, types are ordered from the smallest to the largest mobility. The ranking is the same in all cohorts and for the three measures. Type 2 less than Type 1 less than Type 3.

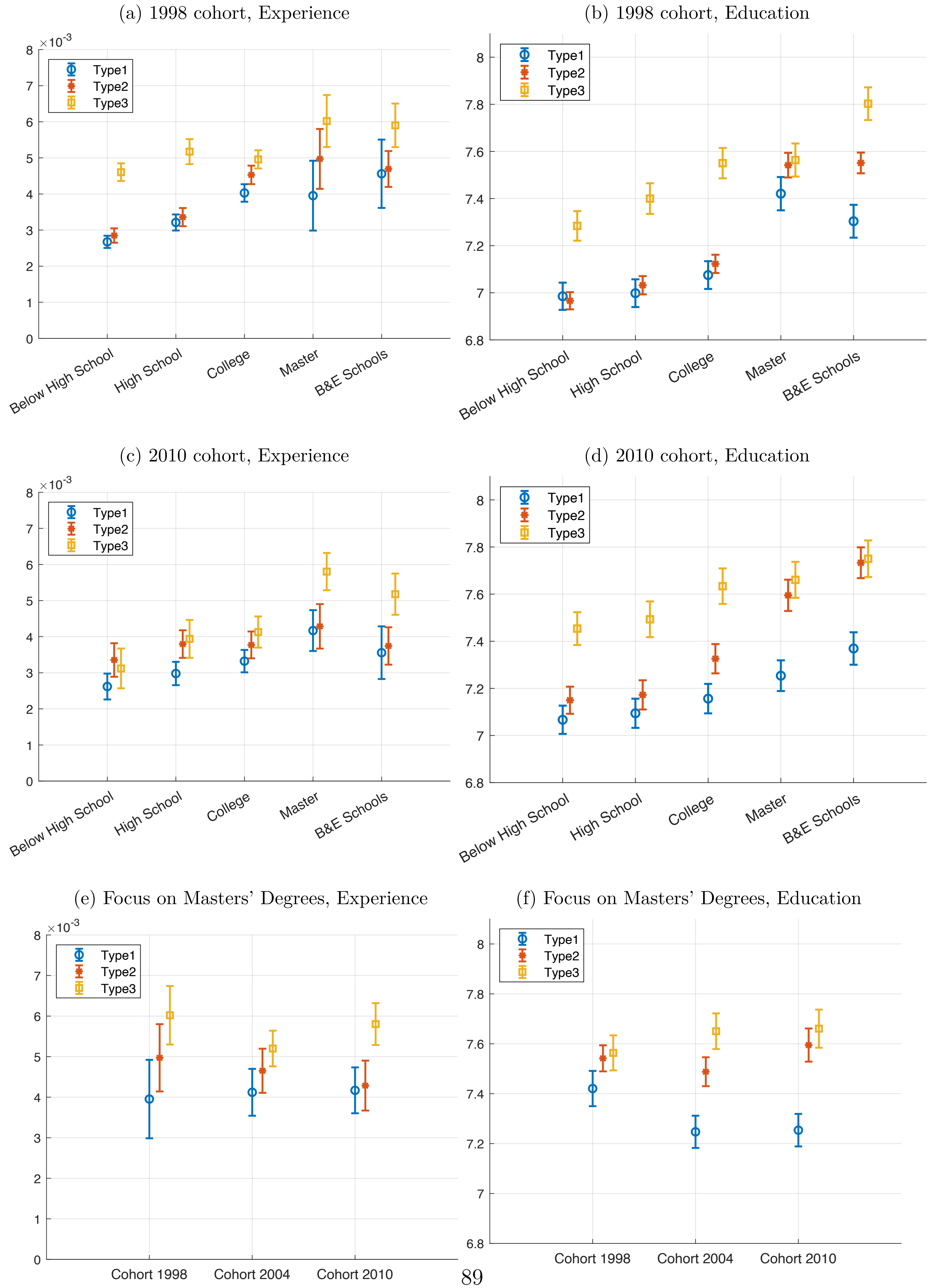
Table 30: Distribution of Types by Firm Class (First Job), in %

	Cohort 1998			Cohort 2010			
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3	All
Class 1	49.1	3.2	17.1	48.2	3.1	13.8	24.9
Class 2	34.0	9.6	33.2	30.4	6.5	25.8	25.0
Class 3	12.7	34.3	31.7	15.9	24.6	35.2	24.6
Class 4	4.2	52.8	18.1	5.6	65.8	25.2	25.5

Table 31: Distribution of Types by Firm Class and Cohort.

	Cohort 1998			Cohort 2010		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Master Degree Student						
Class 1	41.65	0.00	7.49	42.20	0.02	9.89
Class 2	43.50	0.58	25.79	34.39	0.54	18.87
Class 3	11.38	4.42	37.95	17.20	6.58	36.70
Class 4	3.48	94.99	28.77	6.21	92.86	34.54
Business and Engineer schools' students						
Class 1	25.98	0.06	8.87	20.32	0.04	10.43
Class 2	35.25	0.42	19.95	40.56	0.33	18.44
Class 3	25.38	7.30	38.83	23.96	6.09	36.13
Class 4	13.40	92.22	33.01	15.15	93.54	34.99

Figure 12: Monthly Returns to Experience and to Education by Type, Educational Attainment and Cohort in the Variant with Firm Effects



Note. Estimated coefficients and confidence intervals of returns to experience in the variant of our model including firm effects and firm classes.