Heterogeneous effects of dropout on labor market outcomes : the French higher education case

Gaspard Tissandier - Université Paris 1 Panthéon Sorbonne

20/05/2022

Abstract:

In the sense of Spence, education sends a signal to the labor market which helps to reveal the ability of former students. Dropout is usually perceived as a negative signal, leading to lower wages and employment rates. This paper tests for the presence of heterogeneous effects of dropout in french higher education (post-high school diploma) on labor market outcomes from 2010 to 2013. These effects are measured on the time in employment and average wages at an individual's entry into the labor market. I analyze if the heterogeneous structure of the effects is conditional on individual characteristics such as diploma, social origin, or gender. I use the Generalized Random Forest algorithm with the distance to the closest higher education institution at 6th grade as an instrument to estimate Local Average Treatment Effect. In line with the literature, I find a negative effect of dropping out on the time in employment. However, this effect is heterogeneous across individuals, ranging from -55% to no effect. I find two subgroups with a significant and negative effect of dropping out on the average wages, with effects ranging from -5.1% per month to a -1.2% on average monthly wage. The gender, study field, study duration, and the social origin of parents, especially the diploma, have an active role in shaping the heterogeneity of the dropout effect.

JEL code : J01, J24, I2, I24

Keywords : dropout, higher education, grf, instrumental variable, labor market outcomes

1 Introduction

For a large part of the workforce, higher education is a crucial phase for accumulating competencies, knowledge, and skills that will be later valued in the labor market. While it is decisive for students to acquire diplomas to testify to their abilities, dropping out is a recurrent event in the French higher education system, whether voluntary or involuntary (Aina et al.). Dropping out has a strong effect on labor market outcomes and can penalize new entrants in the long run (Schnepf). Dropouts are numerous in the French higher education system: in 2018, 23.9% of students enrolled in their first year of higher education dropped out. This phenomenon is persistent trough time as 4.1% of students who began their study in 2014 dropped out at the end of the second year, and 10.4% at the end of the third year[1].

Following Spence, dropping out send a negative signal to the labor market about the ability of the individual. Ability is defined as an underlying variable reflecting the capacity of a worker to perform well in a task, or a set of tasks. Indeed, the negative signal of dropping out on the ability competes with the positive signal brought by the diploma, affecting the wage, the probability of being employed, or the path of the worker in the labor market. Dropout has a negative effect on labor market outcomes, either on salary, opportunity or rate of employment (Flores-Lagunes and Light, Matkovic and Kogan ,Reisel, Schnepf) . It has been observed that the negative effect of dropout can be heterogeneous conditional on the motivation of the dropout (Bjerk) or even on the diploma or the age of the individual (Brodaty, Gary-Bobo, and Prieto, Navarro, Fruehwirth, and Takahashi, Scholten and Tieben).

It has been shown that the social origin and individual characteristics of students are highly determinant in the dropout process (Aina et al., Vignoles and Powdthavee), and that specific variables such as the academic path or gender have an impact on the structure of the effect of dropout. However, no study tried to analyze the heterogeneity of the dropout effect on labor market outcomes concerning a large set of individual characteristics. Acknowledging that the consequences of dropout are not the same depending on individuals is crucial to understand educational choices from students, and to design better policies against student dropout.

This paper asks whether the effect of dropping out on labor market outcomes, such as wages and probabilities of being employed, follows a heterogeneous distribution. Then, the distribution of this effect is studied to test if it is conditional on individual characteristics and more specifically if variables such as the social origin or the gender play a more important role than endogenous variables such as diplomas. Finally, an additional analysis on the time to the first job and the first salary is proposed in order to explain the mechanisms generating the heterogeneity of the dropout effect across diplomas and social origins.

According to both fundamental models of education (Becker or Spence), acquiring more year of education bring higher earnings, as detailed in the analysis of Fang or the review of Psacharopoulos and Patrinos. The role of higher education as a signaling method is widely

[1]Repères et références statistiques 2019 - Direction de l'évaluation de la prospective et de la performance

explored in the literature. The study of Arcidiacono, Bayer, and Hizmo shows that college diplomas act as an ability signaling method, as college graduates get wages matching their measured ability quicker in their carrier than high school graduates. They use an external individual measure of ability (an ability test matched with the students' database) to prove that the revealed ability is not only by diplomas levels but that education conveys a very precise signal of ability allowing the companies to find better fitting individuals more quickly.

However, dropout raises less consensus about its effects on labor market performance. As highlighted by Schnepf, the literature about labor market performance of dropout is scarce. Bjerk studies the effect of dropping out on criminal activity and labor market outcomes. The author finds that dropping out has a strong negative effect on both indicators. However, one of the main findings lies in the heterogeneity of the dropout effect: students who drop out for "passive reasons" have lowest performances than those who drop out with plans, or on purpose. In Schnepf, the author finds that in most of the European countries, dropouts are benefiting from their study time, compared to students who didn't enter higher education. In this paper, Schnepf uses a propensity score matching model on data from the 2011 Programme for the Internationale Assessment of Adult Competencies to pursue the study on many different European countries. This conclusion is similar to Reisel, where the author shows that in the United States, it is beneficial to integrate higher education even without graduating, compared to individuals without any higher education experience. Similarly, Matkovic and Kogan compares the effect of dropping out on labor market performances in Croatia and Serbia and corroborates the finding of the overall negative effect. They also find that the longer a student stays in higher education, the smoother the transition in the labor market is, especially in Serbia. This result is similar to the one of Flores-Lagunes and Light in the United States, where the sheepskin effect (the premium of having graduated from a diploma) is highly conditional on the number of years of schooling. In France, the study from Brodaty, Gary-Bobo, and Prieto covers the effect of delayed graduation (of which dropout is a special case) on labor market performance. The authors find a negative effect of delayed graduation, with significant differences between the effects conditional on the highest diploma. In Norway and the United States from 1989 to 1999, Reisel finds heterogeneity in the return to education due to the distribution of women and minorities across the income distribution, while Scholten and Tieben finds that in Germany, for individuals born between 1944 and 1986, the dropout effect is mostly conditional on the previous diploma, which acts as a "safety net".

The main issue in estimating the effect of dropout or delayed graduation is the endogeneity of the event with the underlying ability of the student. As seen before, the propensity score matching is used to solve this issue, as in Schnepf. On the other hand, recent papers like Mahjoub use the period of birth as an instrument, inspired by Angrist and Krueger. An alternative instrument is a distance to the closest higher education institution, as proposed by Card. In Brodaty, Gary-Bobo, and Prieto, the authors use a dense system of geographical IV with the distance to the closest university in 6th grade, and the number of openings of higher education institutions in the geographical area during secondary education.

To allow the estimation of heterogeneous treatment effect, I apply the Generalized Random Forest (GRF) methods, developed by Athey, Tibshirani, and Wager, on a French database of 24000 young workers who finished their education in 2010. Their work records are surveyed from 2010 to 2013, which helps us to construct two indicators of the average of wages and the time in employment for every individuals. The GRF algorithm, based on the Random Forest structure (Breiman), allows us to estimate individual Conditional Average Treatment Effect (CATE). Individuals are then gathered in subgroups following their CATE magnitude to compute the Average (Conditional) Local Average Treatment Effect (A(C)LATE) on these subgroups (Imbens and Angrist, Athey and Wager). The asymptotic property of the A(C)LATE allows to do inference on the quality of the estimators and to test for the presence of heterogeneity in the CATE distribution. To assess if the shape of the CATE distribution is conditional on individual characteristics, I estimate a LOGIT model for being less penalized by dropout. By estimating the likelihood of belonging in the top 50% as the dependent variable, we can study which individual characteristics are the most important in shaping the heterogeneous effect of dropout, and who are the most penalized students with respect to their diplomas, social origin, gender and other characteristics.

Finally, in order to propose a causal mechanism in the results, I execute the same analysis on two additional indicator : the time to the first full time job and the first salary. It is indicative of the behavior of the former students on the labor market, especially shortly after the end of their education.

The endogeneity of dropout is tackled with an instrumental variable setting adapted to the Random Forest structure of the GRF. I use the square of the distance to the closest Higher Education institution at 6th grade as an instrument. Paired with a vector of controls to estimate the predicted probabilities of dropout, I obtain an efficient instrumental variable setting allowing me to identify heterogeneous causal effects of dropout on labor market outcomes. The distance is measured at 6th grade to avoid the endogeneity due to the use of the distance between higher education institutions and the high school diploma city (Brodaty, Gary-Bobo, and Prieto).

I succeed to find two groups with heterogeneous treatment effects for both indicators. After estimating the individual Conditional Average Treatment Effect (CATE - the individual effect), the sample is split around the CATE median, and the Average (Conditional) Local Average Treatment Effect (A(C)LATE) is estimated on both subsamples. This methodology, proposed by Athey and Wager and Chernozhukov et al., allows to estimate a doubly robust treatment effect on the treated compatible with the Instrumental Forest method. According to this estimation, dropping out has a negative effect of -55% of the observed period for the most penalized 50% of the sample, and no effect for the other 50%, on average. On average wages, the effect goes from -5% on monthly wage for the most penalized group to -1% for the less penalized group. The difference between the A(C)LATE of these subgroups is tested with a Student test, and these two tests confirm the heterogeneous distribution of dropout effect on both labor market outcomes.

To understand which individual characteristics are the most important in shaping the individual's CATE distributions and who are the most penalized individuals, I estimate a LOGIT model for the likelihood of having a dropout effect above the median of CATE.

For the CATE distribution of dropping out on the time in employment, the students who dropout from longer degree such as Master are less penalized when they drop out. The social origin (ranked from disadvantaged to highly advantaged, on 4 levels) and the parents' highest diploma are highly determinant in shaping the heterogeneity of the CATE. The most penalized individuals are those coming from disadvantaged or highly advantaged background, while having parents with at least an high school diploma helps to reduce the dropout penalty. Academic achievements such as having done an internship greatly helps to reduce the dropout penalty.

Regarding the dropout CATE on average wages, the parents' education is far more preponderant than the parents' social category, and having educated parents still helps to reduce the dropout penalty. Students dropping out from long degree are no longer less penalized, except for those who dropout from business or engineering school. Academic achievements are still helpful to minimize the dropout penalty

Finally, I proposed a causal mechanism explaining to explain these results. I find that the time to the first full time job is crucial when shaping the effect on the time in employment and the average wage, and students with a disadvantaged background or parents with no diploma are the most penalized when entering the labor market. They also tend to accept lower entry salary, thus affecting durably the labor market path.

This paper sheds new light on the integration of french tertiary education dropouts in the labor market. The main contribution is the application of machine learning techniques that helps to account for individuals' characteristics and unfold the heterogeneous structure of the dropout effect on labor market outcomes. These results will help to understand better the path of higher education dropouts and to design a policy that prioritized students who could benefit from it the most.

2 Data

To identify the effect of dropping out on former students' labor market outcomes, I use "Génération 2010", a longitudinal survey provided by the CEREQ (Centre d'Etudes et de Recherches sur les Qualifications) [2]. This survey is conducted on individuals who have finished their education in 2010 (between October 2009 and October 2010), without any interruption before. Individuals are surveyed in 2013, three years after they left the educational system. The resulting database consists of a panel gathering information about former students' background, education, and a detailed schedule of employment from 2010 to 2013. The survey covers 33547 individuals with a wide range of education, social background variables, and professional records. I restrain this data set to individuals who at least, tried to obtain a higher education diploma. This array goes from high school diploma holders who tried one year of higher edu-

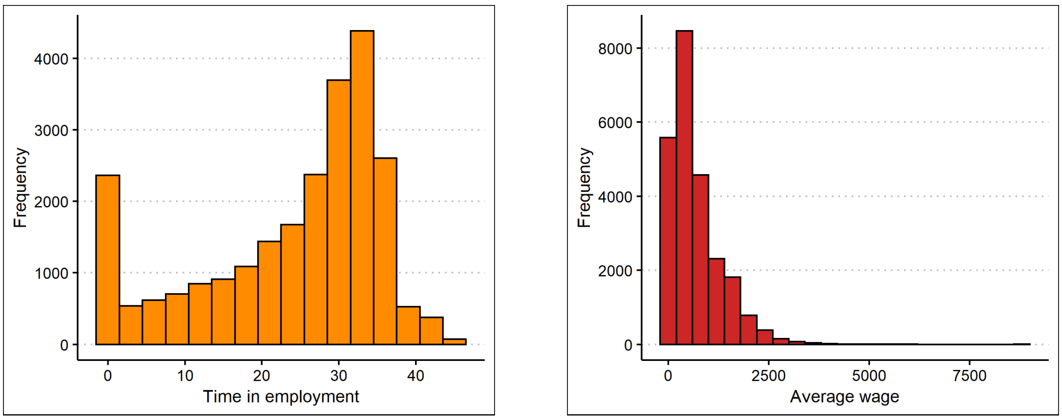
[2] Génération 2010 – Interrogation à 3 ans – 2013 (2013, CEREQ)

cation to Ph.D. graduates. This represents a data set of 24201 individuals.

I create indicators variables for dropout, the number of months worked, and the average wages. Dropping out is defined here as not having validated a diploma in 2010, or exiting the educational system before the last year of said diploma. For example, if a student didn't graduate of her Master 2 because she didn't pass the exams, she will be considered as a dropout. A student who interrupted her study in the second year of undergraduate, out of the three required years will also be considered as a dropout. According to this definition, the database consists of 5844 individuals who dropped out, and 18357 who didn't (24% of dropout).

Each individual's employment curriculum is entered in a side database where employment and unemployment periods are filled in. For each working sequence, the beginning and ending salaries are specified, as the duration in months. This setting allows us to create two variables in order to test our hypothesis.

The first variable, *Time in employment*, consists in the number of months worked full time over 36 months (3 years), expressed as months (to ease the understanding, I will often express this indicator in percentage). The second variable, *Average Wages*, is an average of the wages on the whole period (36 months). Then, if an individual works 12 months with a salary of 1200€, *Average Wages* will equal 400€. In order to work on percentage differential between individuals, and not in monthly salary directly, I use the logarithm of the average wage as a dependent variable.

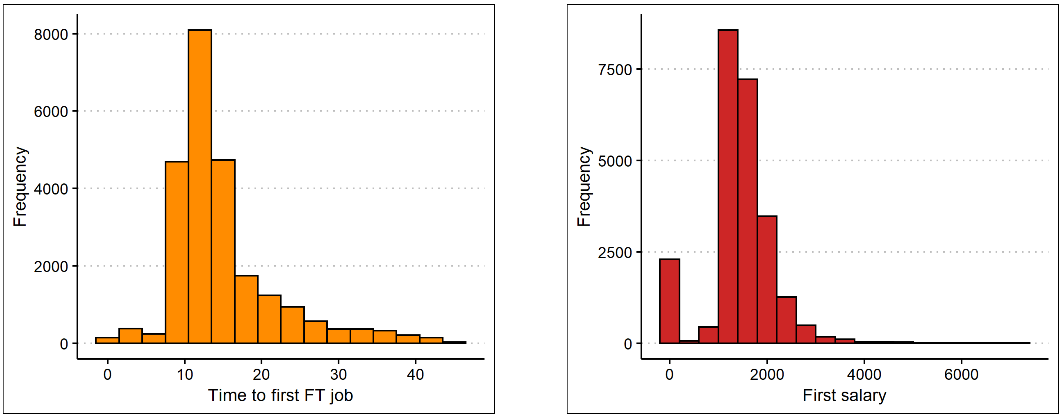


(a) Time in Employment

(b) Average wages

Figure 1: Distribution of dependent variables

Finally, I include the time to the first full time job and the first salary in the database. The former is expressed in months while the second is also included as a logarithm of the original indicator. Their distribution are presented in table 2.



(a) Time in Employment

(b) Average wages

Figure 2: Distribution of dependent variables

The following variables are kept : highest diploma tried on 32 levels and 12 levels, if foreign study travel or internship have been done by the student, geographical location in 6th grade, high school, and in 2010, when the individual left the education system. I also keep the gender of the individual, the professional occupation, professional sector (public or private) and diplomas of both parents, the number of siblings, and information about past education such as an indicator of presenting delay in 6th grade, at the high school diploma and the discretized grade of the high school diploma. Finally, I keep information on the type of high school diploma acquired (general or professional, and also scientific, economics, literature). The descriptive statistics are presented in table 1. For commodity reasons, the diploma is presented on 12 levels and the social origin is presented only for the mother.
The time in employment and average wages conditional on dropping out and individual characteristics are presented in Appendix.

HSD corresponds to the High School Diploma (or Baccalaureate). The time needed to acquire the diploma is counted as "+y" : HSD +2 corresponds to two years or study after the HSD. "Licence Pro" corresponds to vocational degrees, needing 3 years to complete. The short degrees (two and three years) lead to a precise field, and are considered as "professional degrees". The rest of the notations are self-explanatory.

Concerning parents' occupation, the levels are defined using the type of occupation. Disadvantaged social category corresponds to factory worker and unemployed individuals. The intermediate category gathers employee and farmer, the advantaged category gathers intermediary profession, craftsman and independent while the highly advantaged gathers CEO, managers and executives. The parents' diploma are self explanatory, except for long degree which gathers individuals with 5 years or more of higher education. I use the maximum of these variables among the both parents in order to account for the global family environment, and not only the father's or the mother's background.

In table 1, we can also observe the distribution of dropout by highest diploma tried. The percentage presented for the dropout columns is the distribution of dropout for the considered diploma, while it is the percentage of the considered diploma among the whole population in

	No dropout	Dropout	No dropout (in %)	Dropout (in %)	Total	Percentage
			Gender			
Male	7914	3180	71.3 %	28.7 %	11094	45.8 %
Female	10443	2664	79.7 %	20.3 %	13107	54.2 %
			Highest diploma tried			
HSD +2	7454	3868	65.8 %	34.2 %	11322	46.8 %
HSD +3/4	3370	1397	70.7 %	29.3 %	4767	19.7 %
HSD +5 University	3296	206	94.1 %	5.9 %	3502	14.5 %
Grande Ecole	1791	58	96.9 %	3.1 %	1849	7.6 %
PhD	2446	315	88.6 %	11.4 %	2761	11.4 %
			Parents' highest social category			
Disadvantaged	2510	1019	71.1 %	28.9 %	3529	14.6 %
Intermediate	4625	1702	73.1 %	26.9 %	6327	26.1 %
Advantaged	3526	1164	75.2 %	24.8 %	4690	19.4 %
Highly Advantaged	7696	1959	79.7 %	20.3 %	9655	39.9 %
			Parents' highest diploma			
No diploma	3009	1328	69.4 %	30.6 %	4337	17.9 %
HSD or below	6517	2391	73.2 %	26.8 %	8908	36.8 %
Short degree	5151	1361	79.1 %	20.9 %	6512	26.9 %
Long degree	3680	764	82.8 %	17.2 %	4444	18.4 %

Table 1: Summary statistics by dropout status

the total column. We can observe that the dropout rate decrease with the accumulation of years of education. However, the PhD students present a dropout percentage of 11.4%, which is quite high for the longest degree possible.

The dropout rate is decreasing with the increase of the parents' highest social category or highest diploma. These results are fully in line with the literature documenting the heterogeneity of the dropout rate among different social origin.

Finally, I base my instrumental variable setting on the distance to the closest higher education institution from the student's 6th-grade city. This distance is computed by using GPS coordinates and the distance between both points on the geodesic[3]. The geographical unit is the *zone d'emploi*, dividing France into around 310 areas. If there is a university or a school in the *zone d'emploi* of the 6th-grade city, the distance is then 0. The density function of this variable is presented in figure 3. I use the square of the distance as an instrument.

[3]For computation methodology, see : C.F.F. Karney, 2013. Algorithms for geodesics, J. Geodesy 87: 43-55. doi: 10.1007/s00190-012-0578-z. Addenda: https://geographiclib.sourceforge.io/geod-addenda.html. Also, see https://geographiclib.sourceforge.io/

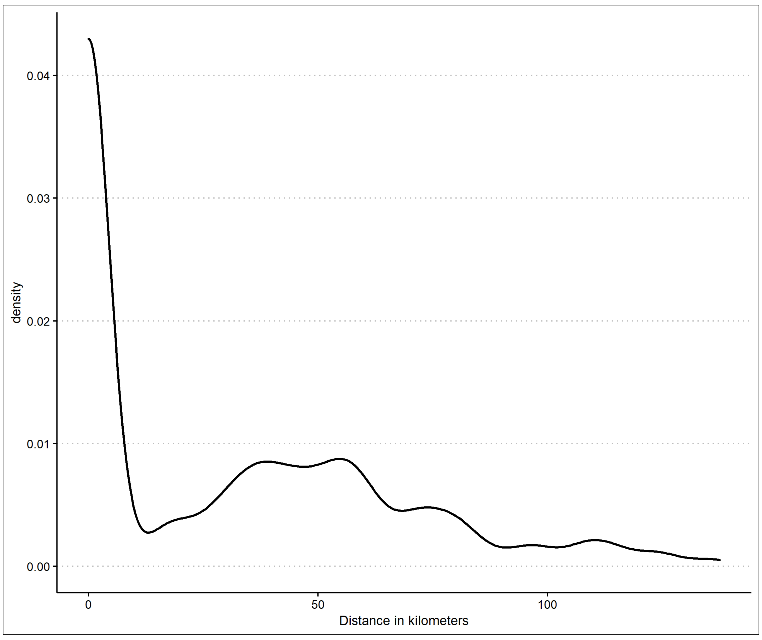


Figure 3: Distribution of the distance from 6th grade home to the closest university

While widely used as an instrument for educational attainment, the distance to the closest university is also a valid instrument for dropout or delay in graduation .With a hypothesis of rational individuals seeing education as a sequential process, the distance to the closest university will have a negative effect on educational attainment : for two similar students except that one is living farther than the other from the local university, the first one will have a higher cost of education, thus have a higher probability of not engage in the next year.

The distance to the closest university also affects negatively the dropout probability in two main ways. First, the instrument affects negatively the years spent to acquire education. Thus, if the distance reduces the time in education, it also reduces the occasion of dropping out. Second, dropping out of a degree leads to sunk costs. If a student quits a two-year degree, even if this experience can be valued in a certain manner, some benefits of having acquired the diploma will be lost. For two students who attained the same degree, the one which had to move from her parents' place will have a higher sunk cost, and then will be less likely to drop out.

3 Methodology

The objective of this analysis is to identify subgroups with different treatment effects of dropout, conditional on a vector of covariates X. If we want to test for every interaction that this vector allows, the number of interaction terms could be gigantic and will obviously detect spurious correlation. To avoid this pitfall, I rely on the Generalized Random Forest developed by Athey, Tibshirani, and Wager, and use the data structure to identify heterogeneous treatment effects. This method allows us to compute Conditional Average Treatment Effect (CATE), the individual treatment effect, and the corresponding standard error, and then to average these effects on selected partitions of the population as Average (Conditional) Local Average Treatment Effects

(A(C)LATE). This method relies on regression trees to estimate the CATE, and average the estimated CATE across all trees. This methodology is called the Random Forest (Breiman) and allows to account for large possibilities of interactions between covariates, without risking over-fitting. To avoid spurious correlations due to using similar data to construct the trees and estimating the treatment effect, the authors rely on the "honest methodology". Finally, the endogeneity of dropout forces the use of an adapted instrumental setting. In this section, we will develop the Generalized Random Forest algorithm, the instrumental variable setting, and then the Average (Conditional) Local Average Treatment Effect estimation.

The objective of our paper is to evaluate the causal effect of dropout. As defined by Rubin, we want to compute the individual difference in potential outcome $\tau_i = Y_i(1) - Y_i(0)$ with $Y_i(W_i)$ the outcome depending on the treatment status W_i. Since we do not observe both $Y_i(1)$ and $Y_i(0)$, alternative estimators are needed. Thus, we focus on the estimation of the Conditional Average Treatment Effects (CATE) defined as $\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x]$. This estimator is used as a subsample average treatment effect on the individuals sharing $X_i = x$. Thus, for a combination of the vector $X_i = x$, we will be able to compute the treatment effect on this combination x, corresponding to individuals showing similar characteristics with i. In order to estimate the individual CATE, I rely on the Generalized Random Forest Algorithm (GRF).

In this section, we will avoid to use too much technical explanations and try to focus on the general idea of the methodology. See Athey, Tibshirani, and Wager for all the technical details of the Generalized Random Forest.

3.1 The Generalized Random Forest algorithm

(All the notation are taken from either Hastie, Tibshirani, and Friedman, Athey and Imbens, Wager and Athey or Athey, Tibshirani, and Wager).

The GRF is based on the regression tree algorithm developed by Breiman et al. (called the CART for Classification and Regression Trees) and adapted as causal honest tree by Athey and Imbens. I will proceed by first describe the honest causal tree, and then the adaptation to the Generalized Random Forest.

The objective of an honest causal tree is to create subgroups in the population on which the Conditional Average Treatment Effects (CATE) are evaluated. For a given dataset, we observe (Y_i, W_i, X_i), for $i = 1, ..., N$, with Y_i the outcome, X_i a vector of covariates and W_i the treatment status. In our example, if the considered individual has dropped out, she shows $W_i = 1$ and $W_i = 0$ if she didn't.

In the initial paper by Breiman et al., the regression tree use a training sample $\mathcal{S}^{tr}$ for which we know (Y_i, X_i) and a target sample for which we know only X_i. By fitting a tree model on $\mathcal{S}^{tr}$, the objective is to predict correctly the outcomes for the target sample. To do so, the algorithm first search for a splitting point s on a splitting variable X_j in order to create two subsample $R_1(j,s) = [X \mid X_j \leq s]$ and $R_1(j,s) = [X \mid X_j > s]$. In this setting, s is found by minimizing the mean squared error defined as :

$$MSE = \left[\sum_{x_i \in R_1} (y_i - \bar{y_1}(j,s))^2 + \sum_{x_i \in R_2} (y_i - \bar{y_2}(j,s))^2 \right] \tag{1}$$

Finally, the algorithm repeat this method until a stopping point (minimum number of individuals in subsamples, maximum number of subsamples, for example). To compute prediction for another sample, the algorithm fit new observations into its corresponding subsample, and then assign the mean of this subsample as the predicted outcome $\hat{Y_i}$. Compared to linear regression or similar methods, the CART allow us to account for high dimensional interactions between all covariates in X_i and help to build strong predictive models.

If the CART is efficient to produce prediction on a target sample, it is not yet suitable to estimate CATE. For this aim, we need two modifications of the original algorithm : introduce an "honest" design and use an modified splitting rule.

The honest design, firstly applied to regression tree by Athey and Imbens, help to solve the over fitting problem. Over fitting arise when a model match too closely the data and then present no generalization power. Indeed, if we use the same sample to build the regression tree and to estimate the CATE in every created subsamples, wee will obtain completely biased estimators. In the honest design, we use two different and randomly drawn subsamples to build the tree with the first one, and then to estimate effect in the subsamples build by the regression tree in the second one.

In order to account for the second stage estimations, we need to adapt the objective function. We will focus on the Expected Mean Square Error, an adapted estimator of the Mean Squared Error.

We introduce here the estimated Conditional Average Treatment Effect, the estimated expression of the CATE presented below. With $\hat{\mu}^2$ the conditional mean of a subsample, it is defined as :

$$\hat{\tau}(x; \mathcal{S}) = \hat{\mu}(w_i = 1, x, \mathcal{S}) - \hat{\mu}(w_i = 0, x, \mathcal{S})$$

This expression estimate the CATE on individuals with $X_i = x$ as the difference between the both treated and non treated conditional mean on the given subsample. With an adapted estimate of the CATE, it is possible to design an objective function which suit our need. With N^{tr} the size of the training sample (made equal to the size of the estimation sample), l a subsample, $S^2_{\mathcal{S}^{tr}}(l)$ the subsample estimated variance of $\hat{\tau}$ and p the probability of being treated , the adapted expected Mean Squared Error is defined as :

$$\widehat{EMSE}_\tau(\mathcal{S}^{tr}) = \frac{1}{N^{tr}} \sum_{i \in \mathcal{S}^{tr}} \hat{\tau}^2(X_i; \mathcal{S}^{tr}) - \frac{2}{N^{tr}} \sum_{l} \left(\frac{S^2_{\mathcal{S}^{tr}_{treated}}(l)}{p} + \frac{S^2_{\mathcal{S}^{tr}_{control}}(l)}{1-p} \right) \tag{2}$$

This estimator of the Expected Mean Squared Error is almost composed as the MSE, but add a negative effect of within subsample variance of the CATE. This allow the algorithm to take into account that finer partition generate greater variances. Then, with this objective function, the algorithm will search for split that maximize treatment heterogeneity in treatment effect while avoid generating too much in-partition variance. For more details on the construction of this objective function, please refer to Appendix.

Since we have a efficient splitting criterion, one problem remain : due to the initial honest design, the built tree will greatly depend of the initial random splitting. To solve this issue, we apply the Random Forest algorithm first developed by Breiman and applied to causal inference by Wager and Athey. The objective of the causal Random Forest is to create causal honest trees on subsamples of the whole population. For example, we draw a partition α of the initial population, and build the honest causal tree on this partition as described below. Then, the algorithm average all the individual CATE given by all trees to compute the individual CATE. This method provide estimates of individuals treatment effects with the associated standard error. One of the main assumption of this model is the unconfoundedness i.e $W_i \perp (Y_i(0)), Y_i(1), X_i)$. This assumption is satisfied in a random treatment assignment setting such as Random Control Trials. Since it is almost impossible to randomize the dropout, I have to include a instrumental variable setting in the framework.

The Generalized Random Forest developed by Athey, Tibshirani, and Wager propose a general framework to estimate CATE with methods such as causal Random Forest and Instrumental Forest. The main divergence from the initial causal Random Forest come from the usage of a gradient-based loss criterion rather than the exact loss criterion (2). The gradient-base criterion is an approximation of (2) build with gradient-based approximations of $\hat{\tau}$ for each subsamples. This method, designed as a general framework for estimation in non-linear setting, help to use IV and is less costly in computation.

In this paper, I use the GRF to build individual CATE by using the following variables : the highest diploma tried on 12 levels, if the student made internship or international travel, the higher education institution region, the high school region, the type of high school diploma (general, technical or professional) on three variables, a categorical variable for the grade at the high school diploma, the delay at the high school diploma, the professional occupation and diploma of both parents, the number of siblings and the gender.

3.2 The instrumental variable setting

In this setting, I want to identify the effect of dropout on wages and time in employment. Unfortunately , these two variables have (at least) a common generating variable usually defined as the ability. If I believe our treatment and outcome variable to be link by the model $Y_i = \mu(X_i) + \tau(X_i)W_i + \epsilon_i$, a clear endogeneity problem arise as the treatment is correlated to the error term via the individual ability. To solve this issue, I need to find an instrument Z_i, correlated with the treatment W_i (having dropped out) but not with the error term i.e the ability.

In the case of the GRF algorithm, estimating a Instrumental Forest is equivalent to apply the Wald formula for individuals with $X_i = x$. The interactions terms generated by the GRF, change for every tree and then help us to account for high dimension heterogeneity. Since I have at our disposal an instrument Z_i satisfying all the IV assumptions, I can estimate the treatment effect as :

$$\tau(x) = \frac{Cov[Y_i, Z_i \mid X_i = x]}{Cov[W_i, Z_i \mid X_i = x]} \tag{3}$$

In this setting, I can implement an IV setting with a binary or continuous instrumental variable. However, since the A(C)LATE is used to estimate treatment effect on group of former students, we need a dichotomous instrumental variable. We need to compute the CATE as a doubly robust score as proposed by Athey and Wager, and average them over subsamples to get unbiased results for every A(C)LATE. As precised in Athey and Wager, we need a binary instrument to compute the doubly robust score for the LATE.

To propose a binary instrumental variable with an high correlation, I adapt the Procedure 18.1 from Wooldridge with a logit model. This procedure, established for endogenous treatment, use a logit model to predict the probability of treatment, including all exogenous control and instrumental variable in the model.

In first step I estimate the following logit model :

$$P(W = 1 \mid X, Z) \tag{4}$$

With w the treatment status, x a vector of control variables and z a vector of instrumental variables described in section 3. The fitted probabilities $\hat{W}_i$ can be used as instrument in the GRF setting. However, since the LATE estimator need a dichotomous instrumental variable, I transform $\hat{W}_i$ as follows :

$$\begin{cases} z_{GRF} = 1 \quad \text{if} \quad \hat{w}_i > p(\alpha) \\ z_{GRF} = 0 \quad \text{if} \quad \hat{w}_i \leq p(\alpha) \end{cases}$$

With $p(\alpha)$ the value corresponding to the α^{th} percentile and $\hat{w}_i$ the estimated probability of dropping out. My choice of α is motivated by the Local Average Treatment Effect estimation step. The LATE is the average treatment effect on the compliers i.e individual who respond positively to the instrument. Since the LATE is computed by averaging the treatment effect times a weighting function which is divided by product of compliance scores, we need to keep the compliance scores as high as possible. The compliance score is defined as the individual propensity to dropout conditional on (x, z). The threshold which maximize the product of the scores is $(\alpha) = 0.80$.

$$X = (\text{geographical location in 6th grade, professional situation of parents,}$$
$$\text{diplomas of parents, gender, delay in 6th grade}$$
$$Z = (\text{distance to the closest university in 6th grade})^2$$

With the described setting, I can estimate CATE for every individuals conditional on their characteristics. Since I build an instrumental setting, the obtained $\hat{\tau}(x)$ are the causal effect of dropout, conditional on individuals' covariates.

3.3 Doubly robust estimation and Average Conditional Local Average Treatment Effect

The instrumental forest described previously generate individual Conditional Average Treatment Effect (A(C)LATE), formally $\tau(X) = \frac{Cov[Y,Z|X=x]}{Cov[W,Z|X=x]}$. In their paper Athey and Wager, the authors propose a method inspired from Chernozhukov et al. to estimate doubly robust score of $\tau(X)$. To assess potential heterogeneity in the estimated treatment effects, we average the doubly robust scores to obtain the Average Conditional Local Average Treatment Effect. The A(C)LATE is asymptotically normally distributed, thus we can interpret it as an estimator of the doubly robust treatment effect on the compliers for a chosen subgroup.

The method chosen to assess CATE heterogeneity is to used the estimated treatment effect value generated by the Instrumental Forest built with the GRF methodology, to split the sample around the median of estimated CATE and to compute the A(C)LATE on each subsample. Since the A(C)LATE is a asymptotically normal, we can test if each subsample groups individuals with a significantly different from 0 treatment effect, and if the difference between the both groups A(C)LATE is significant.

The doubly robust score is computed as the sum of the estimate CATE by the Instrumental Forest and the multiplication of the Y residuals multiplied by a debiasing weight :

$$\Gamma = \tau(X) + g(X,Z)\,(Y - \mathbb{E}[Y|X] - (W - \mathbb{P}[W = 1|X])\tau(X)) \tag{5}$$

With $g(X, Z)$ the vector of debiasing weight :

$$g(X,Z) = \frac{1}{\Delta(X)} \frac{Z - \mathbb{P}[Z = 1|X]}{\mathbb{P}[Z = 1|X](1 - \mathbb{P}[Z = 1|X])} \tag{6}$$

In (6), $\Delta(X)$ is the vector of compliance score : $\mathbb{P}[W|Z = 1, X]$. It represent the propensity of an individual to dropout if the instrument is positive. The compliance score are computed using a causal forest (see Aronow and Carnegie for detailed explanation around the compliance score). For the practical way of estimating the doubly robust score, see Athey and Wager.

Finally, the A(C)LATE is estimated as the average of all doubly robust scores. The A(C)LATE are computed on each subsample divided around the median of the CATE.

4 Results

4.1 First stage regression

As described before, dropping out is not randomly distributed in the student population and thus in the used sample. We need to rely on an instrument that is correlated with the dropout indicator, but excluded from the outcome equation. To assess the quality of our instrument, the first linear stage is presented in table 2. The first stage regression is run in the first model without controls, then with controls i.e presenting delay before 6th grade, diplomas, and professional position and work sector (public, private) of both parents, gender, and geographical area of birth (on 12 levels for the whole world).

Dropout	(1)	(2)
($\min^2$ /100)	-0.00403	-0.00458
	(0.00096)	(0.00095)
Constant	0.26052	0.69521
	(0.00388)	(0.14470)
Control variables	No	Yes
R Squared	0.001	0.036
F-Stat for the instrument	17.68	23.36

Table 2: Linear first stage

The initial correlation of the instrument with dropout is strong, and even better after controlling. With a final F-Statistics of the instrument of 23.36, I can rule out the weak instrument bias. As described in the Data section, the minimum distance to the closest higher education establishment is negatively correlated with the probability of dropout.

The effects of the minimum distance on the probability to dropout conditional on the parents' highest diploma or social category are presented in Appendix, figure 10 and 11.

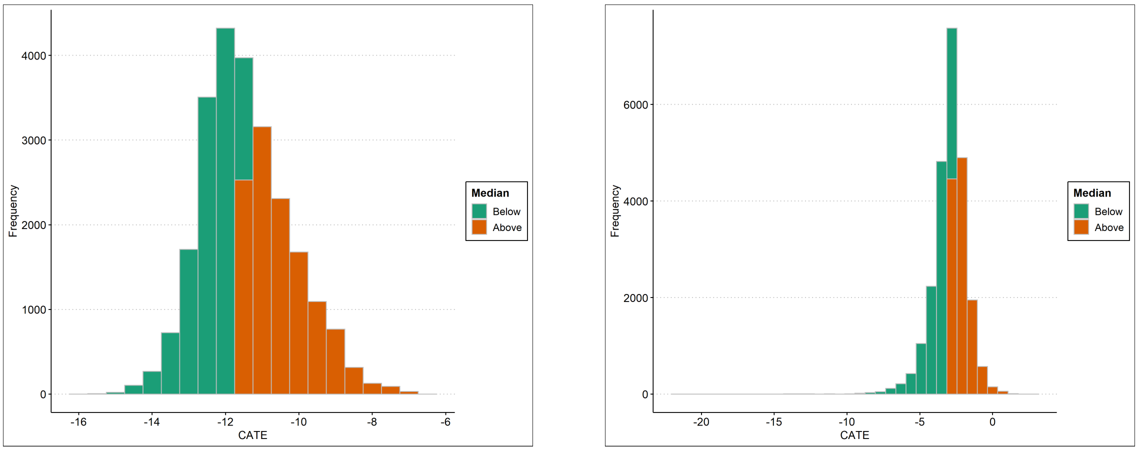
4.2 Assessing treatment heterogeneity

As described in the methodology part, the constructed distributions of CATE are the results of a data-mining process used to discover heterogeneity in the effect of a treatment. Since the whole process is made to maximize the heterogeneity in $\hat{\tau}$, it is entirely possible that the distributions are the results of noise in the data. Then, we need to rely on statistical estimators and their properties to test for the presence of heterogeneity in the distribution.

The CATE distributions for the time in employment and average wages are presented in figure 4. Simple statistics are presented in table 3.

In this section, I split the dropout effect distributions on the time in employment and the average wages around the median, compute the A(C)LATE on both subsamples and test for the statistical significance of their differences.

The estimated CATE follow the distributions showed in figure 4 and table 3.



(a) Time in Employment

(b) log(Average wages)

Figure 4: Conditional Average Treatment Effect of dropout

	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Time in Employment	-16.2	-12.3	-11.6	-11.4	-10.7	-6.7
	(-45.0%)	(-34.1%)	(-32.1%)	(-31.8%)	(-29.7%)	(-18.5%)
log(Average Wages)	-21.71	-3.52	-2.87	-2.96	-2.27	2.48

Table 3: Distribution of CATE of dropout on the Time in Employment and the Average Wages

The CATE on the time of employment range from -45% to -18.5% of the potential employment duration (36 months), with the median equal to -32.1% and the mean equal to -31.8%. This range indicates that individuals show CATE with a slight effect, to very important effects with more than a year of unemployment. This wide gap on individual effect needs to be validated through a A(C)LATE groups difference student test. Half of the individuals show a negative effect of -31.8% or less, which means that around 50% of the sample could undergo a negative effect of almost a year of unemployment, out of three years.

The CATE on the logarithm of the average wage range from an effect -21.7% to 2.5%, with the median effect at -2.87% and mean at -2.96 % on monthly wage. This first step results seriously denote with precedent findings, as the distribution show individuals effects going from strong negative effects to null or positive effects for the less disadvantaged individuals.

As explained in the Methodology section, even if the CATE distribution can be indicative of the dropout effect, we need to rely on the doubly robust score average presented in the methodology part. As proposed by (Athey and Wager), the individuals are split following their individual CATE, around the Median. I define $\hat{\tau}_1 \leq Median(\hat{\tau})$ the group of individuals showing a CATE below the median of estimated CATE and $\hat{\tau}_2 > Median(\hat{\tau})$ the group of individuals showing a CATE equal or above the median of estimated CATE. The $\hat{\tau}_1$ corresponds to a group showing a strong negative effect of dropout, while the $\hat{\tau}_2$ correspond to a group showing a slighter negative effect. I also present the A(C)LATE for the first and fourth quarters ($\hat{\tau}_1 \leq Q1(\hat{\tau})$ and $\hat{\tau}_2 > Q3(\hat{\tau})$. A student test for the difference between the A(C)LATE of both groups is performed and the T-Statistic is given in the table.

The A(C)LATE on each subgroups are presented in Table 4 for A(C)LATE on time in employment and on average wages.

	Time in Employment		log(Average wage)	
	Median	Q1/Q4	Median	Q1/Q4
$\hat{\tau}_1^{roe/aw}$	-20.2***	-22.27***	-5.14***	-6.48***
	(2.09)	(3.11)	(0.42)	(0.69)
$\hat{\tau}_2^{roe/aw}$	-2.3	3.28	-1.15***	-0.08
	(2.07)	(3.10)	(0.35)	(0.46)
T-Statistic	-6.09***	-5.82***	-7.34***	-7.69***

Table 4: Average Local Average Treatment Effect of dropping out Labor Market Outcomes

Reading : the 50% of the sample which have a CATE below the median (presenting a strong negative effect) have an average effect of dropping out on the time in employment of -20.2 months on the compliers

For the effect of dropping out on the time in employment, measured by the A(C)LATE, we observe one subsample with a strong negative effect on the compliers of -20 months, so a negative difference of 56 % for the dropouts, compared to those who didn't. The subsample grouping individual with a CATE above the median shows an A(C)LATE which is not significantly different from 0. The student test of the difference of A(C)LATE between these two groups indicates that this difference is statistically significant at the 1% level This results indicate that while half of the compliers undergo an effect of 50% in time of employment, the other 50% does not have any effect from dropping out. The results for the first and fourth quarter confirm these results.

The A(C)LATE of dropping out on the average wages indicates that 50% of the population of actual dropout is penalized with an effect of -2.3% per month on average wages for the three observed years. The top 50% of the distribution shows a A(C)LATE of -1.17%. These two A(C)LATE are significantly different from 0, and the difference between both groups is also significant at the 1% level. We can conclude that we correctly identify two subgroups of dropouts with heterogeneous effects of the dropout. While the first group presents a strong negative effect, the second one presents less than half of the effect.

In this section, we correctly assessed the presence of heterogeneity in the estimated CATE for both interest indicators. However, we need to study the social composition of the CATE distribution to understand the social dynamic behind the effect of dropout on labor market outcomes.

(For the distribution of actual dropout around the CATE median, see Appendix 1).

4.3 Composition of subsamples

In this section, we study the social and academic composition of the subsamples built around the median for both indicators. The objective is to understand if the CATE distribution is conditional on individual characteristics and how they are influencing the effect of dropout. To do so, a Logit model of being in the top 50% of the CATE distribution is estimated.

I include all the variables used to build the Generalized Random Forest. The main independent variables of interest are gender, the diplomas, if the student did a foreign trip during her study, or an internship, the highest education and the highest professional occupation among the parents. The objective of this analysis is to understand how these variables are influencing the CATE distribution i.e if having parents with a certain occupation or diploma can increase the chance to be in the less penalized group, or if dropping out from a certain diploma is preferable as another diploma.

4.3.1 Time in employment

The logarithm of the odds ratios for the gender, the highest diploma tried, the academic achievements, the parents' highest social origin and diploma are presented succinctly in figure 5. The reference level is on the vertical line, and the p-value for each estimated parameters is reported in parenthesis. For the complete regression tables, please refer to the appendix. I perform the GVIF analysis in order to detect potential multicolinearity in the variables, however no measure goes above the recommended level. This analysis can be found table **??**.

The results from the LOGIT model are to be interpreted as follow : it is the effect of a variable on the propensity of being less penalized when dropping out. I use a split of the sample in half to study the structure of the effect on the whole sample, and not only on a reduce sample (the top quarter for example).

Being a female drastically reduce the probability of being highly penalized by dropout on time in employment. This effect is strong and highly significant.

Regarding the highest diploma tried, the most penalized student are those from short degrees, followed by PhD student. It means that dropping out from a short degree will lead to a shorter time in employment in the three years following exit of the education system. The PhD dropout are also highly penalized. We observe that Grande Ecole dropout are the least penalized in terms of time in employment, which can be explained either by the valuation of the acquired human capital or by the professional network brought by the Grande Ecole environment. University middle and long degrees are less penalized than short degree but don't benefit the penalty reduction of the Grande Ecole.

Regarding the maximum social category of the parents, the results are puzzling. Children from highly advantaged background are the most penalized. We could expect them to perform well even in case of dropout, thanks to their family environment, however it is not the case. Children coming from disadvantaged background are also penalized, but far less than the first category, and are close from children of intermediate family. Finally, children from advantaged background are the less probe to be penalized.

The children from parents with no diploma have the lowest probability of being in the top 50% of the distribution. The children from parents having a long degree are the second most penalized, while this effect is less than for no diploma family. Finally students with parents with short degrees are the less penalized. Except for the long degree, this structure of the diploma effect is logical, as children from educated parents will have less issue finding a job even after dropping out, once again thanks to their parents network or social capital.

Finally, we observe that other educational achievement such as having done a foreign trip or internship during the study helps to limit the dropout penalty. However this is the case only if the students has done only one such trip, and not many. Doing an internship is by far the most important factor to limit the dropout penalty. This is easily understandable by the guaranty on a students ability that an internship provide, and by the skills acquired in real professional situation.

The most puzzling fact pertaining from this results is that children from highly advantaged or educated parents are not the less penalized when they drop out.

By using the magnitude of the different variables' effect, we can assess that the variable partly of entirely determined by the students are fairly important in shaping the CATE distribution for the time in employment. The strongest effect is attached to the the internship indicator, and the exit diploma exhibit strongest coefficient than the social category of the parents.

4.3.2 Logarithm of the Average wage

While being a female was reducing the potential dropout penalty in terms of month worked over the three observed years, it is not the case anymore. Female have a higher probability of being highly penalized when dropping out regarding the average wage, and this effect is highly

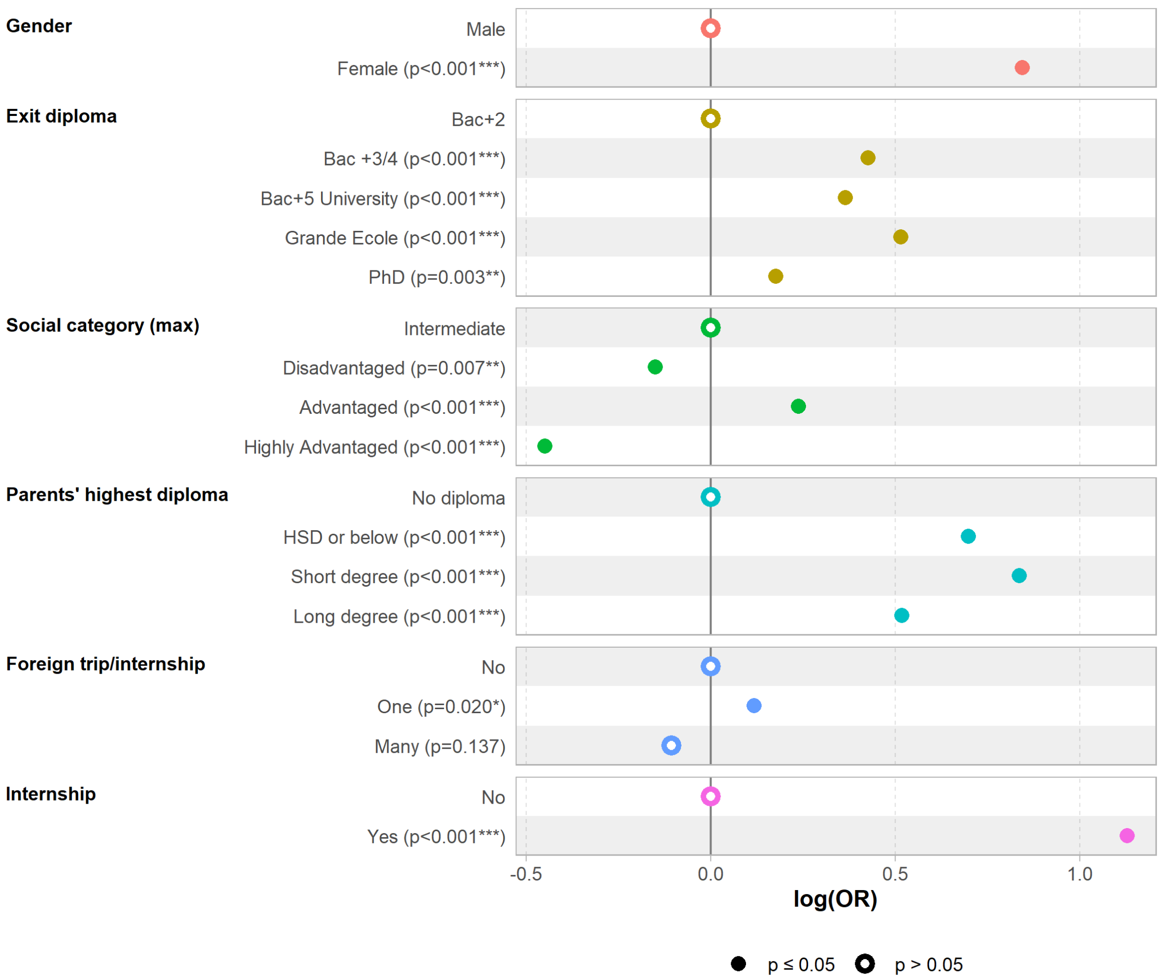


Figure 5: Log(odd ratios) for being in the top 50% of the CATE distribution (Time in employment)

significant.

PhD dropouts are now the most penalized students, followed closely by short university degrees (Bac+3/4) and long university degrees. For PhD students, it means that dropping out will not decrease the time in employment after dropping out, but their salary will highly decrease. The short degree dropout are not the most penalized individuals anymore, as they are the second less penalized level. This means that while these dropouts will take more time to find a job, their salary will be less impacted. This can be attributed to the fact that these degrees lead to precise employment sector, with fixed salary (the health field for example), or with salary close to the minimum wage. Once again the Grande Ecole dropouts are the less penalized, which is easily explained by the networking activity practiced by their schools.

The children of highly advantaged parents were the most penalized by dropout, they are now the less penalized. This could mean that when they dropout, they take more time to find a job but with still a competitive salary. The most penalized students is the advantaged children, while disadvantaged children are close to the reference level.

Regarding the parents' diploma, it is clear that this characteristics is highly determinant

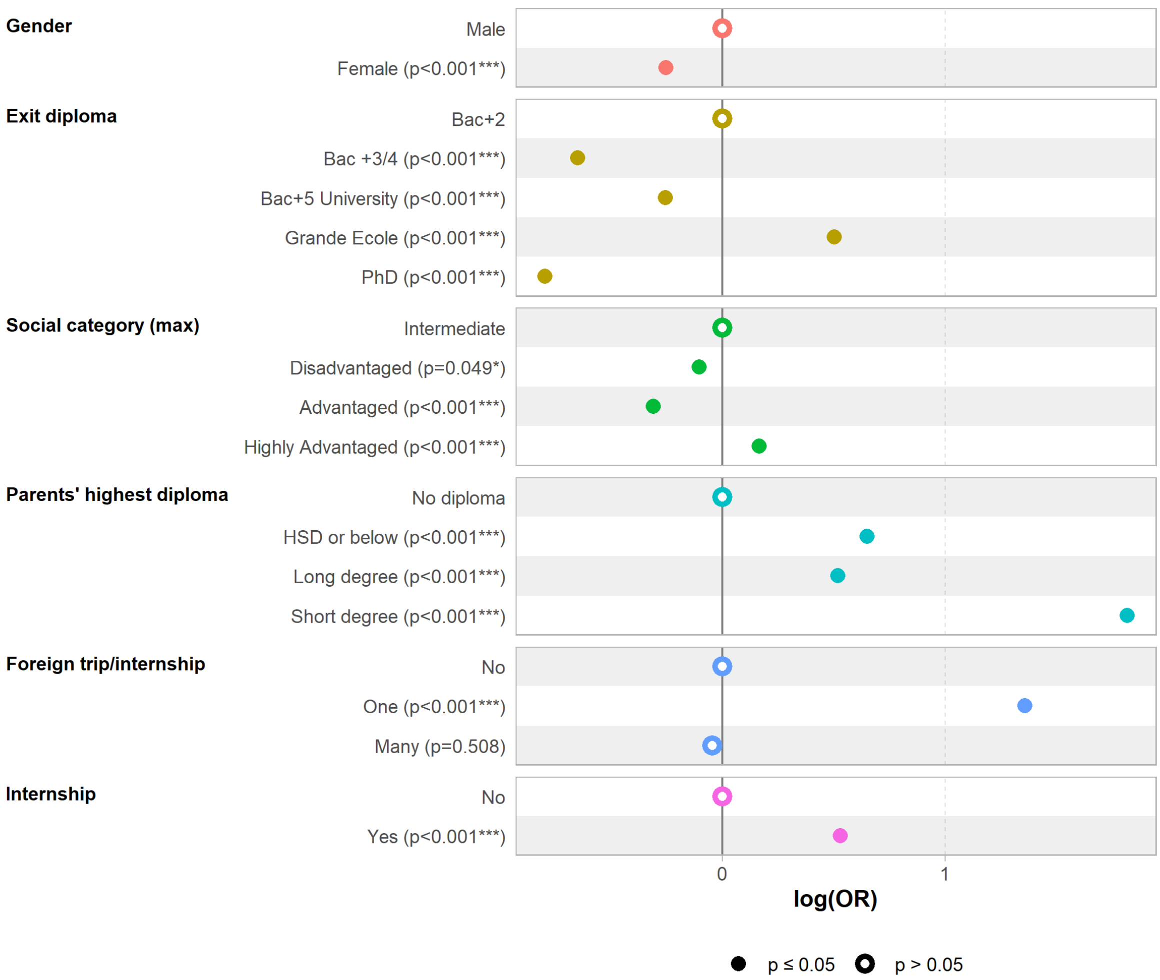


Figure 6: Log(odd ratios) for being in the top 50% of the CATE distribution (Average wage)

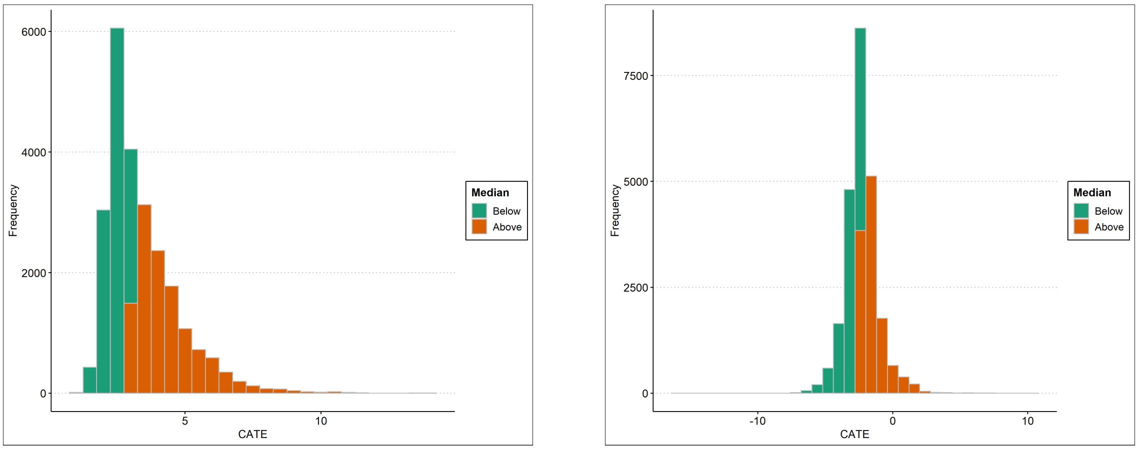
in the magnitude of the effect. Children from long degree graduate are the less penalized by far, while those coming from no diploma family are the most penalized. This indicate that, while the effect of the social occupation is debatable, the cultural and social background brought by the parents is a major determinant of how a dropout will integrate the labor market.

Finally, the others academic achievements such having done a foreign trip or an internship are reducing the dropout penalty on the time in employment and on the average wage.

By analyzing the effect of each variable used to build the Instrumental Forest on the likelihood of being less penalized than average by dropout, we got a better understanding of how this effect is shaped across the population. The social origin is mostly driving the heterogeneity of the dropout effect, especially regarding the average wages. However, the chosen indicator are reflecting the path on the labor market over the three observed year, without any details on the mechanisms behind these results, or the path at the end of their education. To solve these issues, I propose two additional indicators.

4.4 Investigating potential mechanisms

While studying the effect of dropout on time in employment and average wages, these two indicators are not helping to understand fully the structure of the effect of dropping out on labor market outcomes. To provide an explanation for precedent results, I perform the same analysis as in the precedent section on the time to the first full time job, and the first salary. These two indicators will help to spot if certain students tend to find job faster, and if they accept lower entry wage. The CATE distributions for both indicators are presented in figure 7, the A(C)LATE are reported in table 6.



(a) Time to first FTJ (b) log(First Salary)

Figure 7: Conditional Average Treatment Effect of dropout

	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Time in Employment	1.67	-2.71	3.11	3.43	3.95	8.79
log(First Salary)	-15.99	-2.95	-2.37	-2.31	-1.74	10.19

Table 5: Distribution of CATE of dropout on the Time in Employment and the Average Wages

Due to the structure of the indicator, the most penalized individual regarding the time to the first FTJ are above the median : they take more time to find a job than the other. The interpretation for the first salary is the same as for the average wage : the most penalized individual are those with a CATE below the median.

4.4.1 Time to first full time job

Due to the structure of the indicator measuring the time to the first full time job, we estimate the logit for students being in the bottom 50% of the distribution. Individuals with lower or null effect are less penalized because they find a job quicker than those in the top 50% of the distribution. The results are to be interpreted the same as the other models.

	Time to first FTJ	log(First salary)
$\hat{\tau}_1^{ttf/fs} \leq Median(\hat{\tau})$	-0.09	-4.01***
	(1.02)	(0.44)
$\hat{\tau}_2^{ttf/fs} > Median(\hat{\tau})$	6.20***	-0.92*
	(1.34)	(0.35)
T-Statistic	-3.74***	-15.35***

Table 6: Average Local Average Treatment Effect of dropping out on the Time to first Full Time Job and the First Salary

The gender and the diploma don't influence a lot the distribution of the CATE, and the family background is now driving the heterogeneity of the dropout effect.

We observe that disadvantaged students are less likely to find a job soon after their dropout. This result is consistent with the effect on the time in employment : disadvantaged students work less and take more time to find a job. Highly advantaged individuals don't have an effect significantly different from the reference (Intermediate), meaning that even is they spend less time in employment following a dropout, they don't take more time to find a job.

Students with parents without diploma are the most penalized by dropout, in line with the section 4.3.1 results. The probability of being more disadvantaged follow the education time : the longer the parents' degree, the lower the probability of being highly penalized by dropout. The effects described here are really strong, and support the idea that the human, social and cultural capital brought by the family are the most important in order to minimize the dropout penalty.

Finally, having done many foreign trip increase the dropout penalty, while having done a internship reduce it. However, the coefficients are small compared to those of the parents' highest diploma or social category.

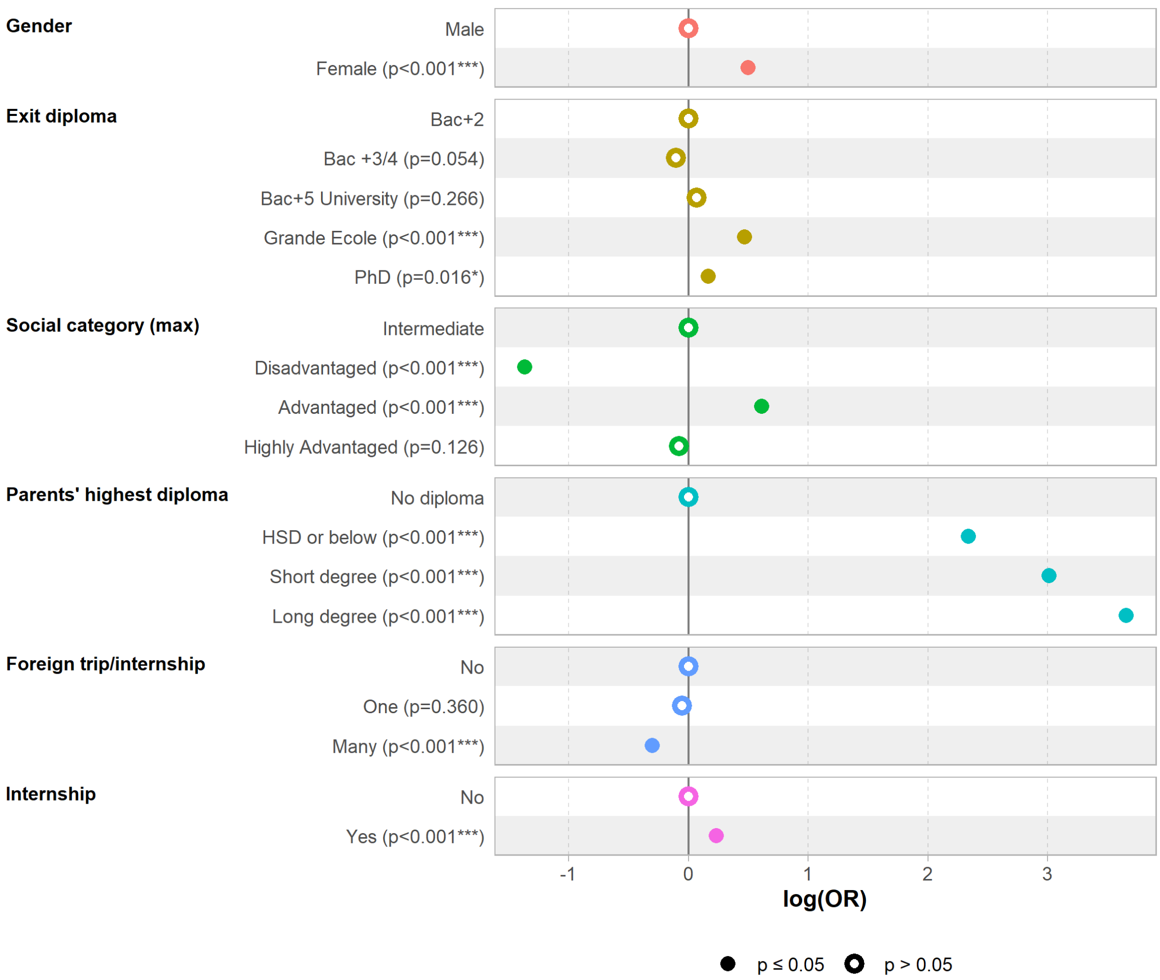


Figure 8: Log(odd ratios) for being in the bottom 50% of the CATE distribution (Time to first full time job)

4.4.2 Logarithm of the First salary

As for the logarithm of the average wage, being a female influence negatively the probability of being less penalized. However the magnitude of the effect is greater for the first salary than for the average wage. This effect is highly significant and in line with the structural wage inequality observe on the french labor market.

The most penalized dropouts are PhD students, followed by the medium length diplomas and the long university degree, with an effect non significantly different from the one of the short degree. Once again the Grande Ecole dropout are the less penalized students, in line with the rest of the results. We can note an advantage for the Grande Ecole against the Bac+5 at University, with the same years of education.

Regarding the social category, we observe that disadvantaged students, while accepting job more quickly than the rest of the sample, also take job with lower entry salary. It is also the case for advantaged students, with almost the same effect. This result confirm that disadvantaged students take more time to find a job in case of dropout, and accept lower entry salary. We can also assume that students from highly advantaged background take more time to find a position in case of dropout, but don't lower to much their salary expectation, however less than the rest of the sample.

The education of the parents plays as a guaranty against lower entry salary in case of dropout : the longer the education, the lower the probability of being penalized in case of dropout, with a real advantage for children of short and long degree holders.

Finally, the only academic achievement which is helping to reduce the dropout penalty is having done a foreign trip. The internship does not play anymore as a determinant of the heterogeneity of the CATE distribution.

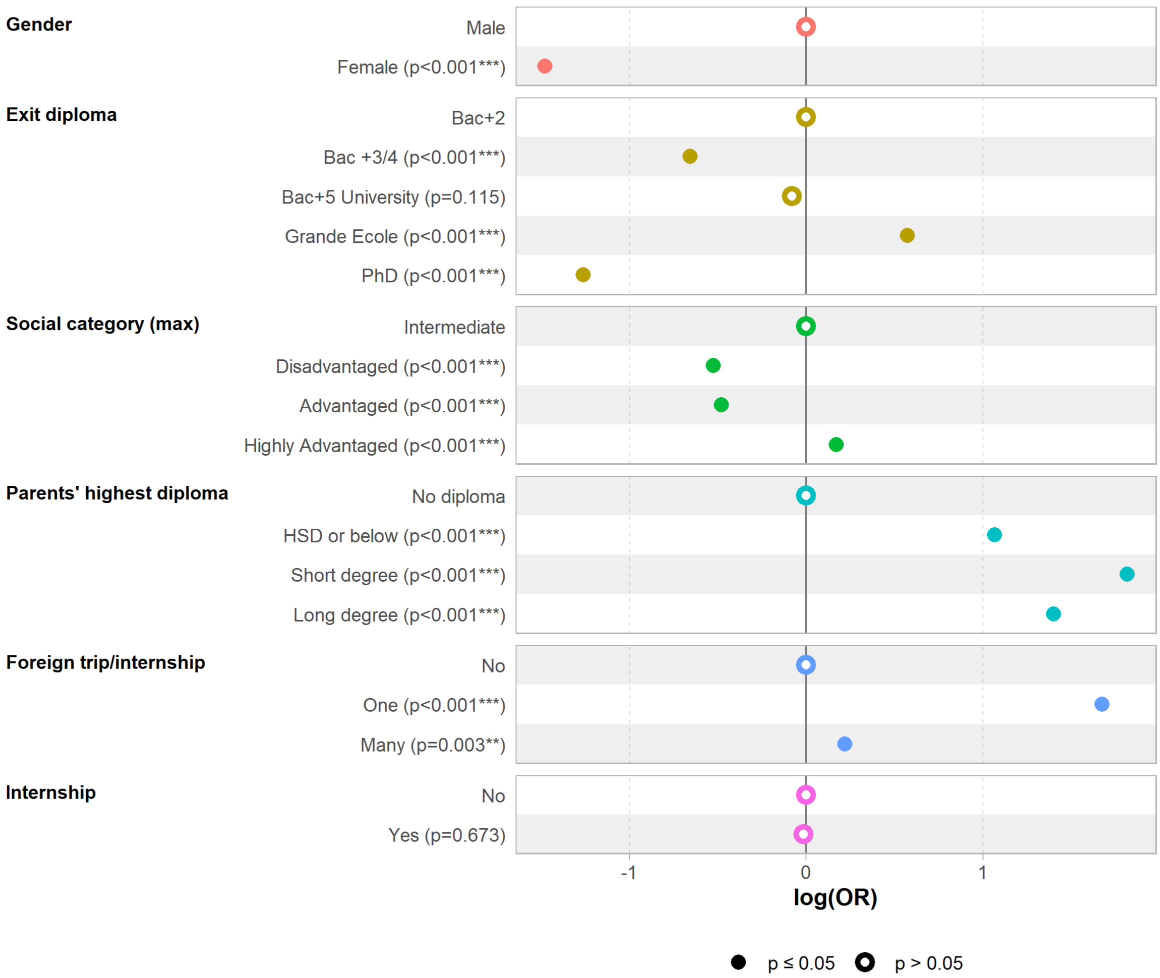


Figure 9: Log(odd ratios) for being in the top 50% of the CATE distribution (First salary)

5 Conclusion

In this paper, I showed that the effect of dropping out on labor market outcomes exhibits statistically significant heterogeneity. I also showed that these heterogeneous effects are conditional to certain education or sociodemographic characteristics.

I used the Generalized Random Forest algorithm in an instrumental variable setting to estimate individual Conditional Average Treatment Effects and then grouped these CATE by subgroups around the median to compute the Local Average Treatment Effect on each subsample. The social composition of these subgroups was studied using a logit model to understand which individual characteristics were more likely to minimize the dropout negative effect.

The main finding is that the effect of dropout on the time in employment and the average wage is heterogeneous and that individual characteristics are actively shaping the distribution of the treatment effects. The heterogeneity in these effects has been tested by using the standard property of the Average (Conditional) Local Average Treatment Effect. We observed two subgroups for each indicator, one with a strong negative effect and the other with a slighter negative to null effect. This result indicates that, while almost every student are penalized for dropout, some are more penalized than others (going from simple to more than the double the magnitude of the effect).

The diploma from which students drop out is highly determinant on the effect on the labor market outcomes. However, family's background such as the social origin or the parents' highest diploma are important to explain the heterogeneity in the dropout effect. The parents' highest diploma is highly determinant, especially for the effect on the average wage. It is also to note that having done a foreign trip or an internship during the study time is helping radically to decrease the effect of dropping out.

I proposed an interpretation of these results in terms of strategic behavior : which type of dropout students tend to take more time to find a job, and do they accept lower wage. The parents' highest diploma is the most determinant variable for both indicators, while the social category indicate that disadvantaged or children of uneducated parents tend to take more time to find a job, and to accept lower wage. Children from highly advantaged background take also more time, but they didn't start with a lower wage.

In terms of effects magnitude, we observe that the family background is at least as important as the student's diploma when determining the heterogeneity in the dropout effect. This result indicate that the social origin or cultural capital of parents are still highly significant for the proper conduct of the integration of the labor market, even in case of a dropout.

This paper raises new results on the need for higher education policymakers to consider multi-dimensional effects of dropout and by extension delay in graduation. By showing that some former students are not as penalized as others by dropping out, this paper brings a less common conclusion on the signal brought by the dropout, and knowing which individuals are the most penalized by dropout can help to broaden even more the access to higher education by designing adapted dropout policy. The importance of family characteristics in shaping the heterogeneity of the dropout effect stresses the fact that policy-makers should focus on making higher education diplomas and achievements valuable experiences for students, even in the case of dropout.

References

Aina, Carmen, et al. "The Economics of University Dropouts and Delayed Graduation: A Survey". *SSRN Electronic Journal* (2018). ISSN: 1556-5068, visited on 01/14/2022. https://doi.org/10.2139/ssrn.3153385. https://www.ssrn.com/abstract=3153385.

Angrist, J. D., and A. B. Krueger. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106, no. 4 (Nov. 1, 1991): 979–1014. ISSN: 0033-5533, 1531-4650, visited on 06/01/2021. https://doi.org/10.2307/2937954. https://academic.oup.com/qje/article-lookup/doi/10.2307/2937954.

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo. "Beyond Signaling and Human Capital: Education and the Revelation of Ability". *American Economic Journal: Applied Economics* 2, no. 4 (Oct. 1, 2010): 76–104. ISSN: 1945-7782, 1945-7790, visited on 11/29/2020. https://doi.org/10.1257/app.2.4.76. https://pubs.aeaweb.org/doi/10.1257/app.2.4.76.

Aronow, Peter M., and Allison Carnegie. "Beyond LATE: Estimation of the Average Treatment Effect with an Instrumental Variable". *Political Analysis* 21, no. 4 (2013): 492–506. https://EconPapers.repec.org/RePEc:cup:polals:v:21:y:2013:i:04:p:492-506_01.

Athey, Susan, and Guido Imbens. "Recursive partitioning for heterogeneous causal effects". *Proceedings of the National Academy of Sciences* 113, no. 27 (July 5, 2016): 7353–7360. ISSN: 0027-8424, 1091-6490, visited on 10/06/2020. https://doi.org/10.1073/pnas.1510489113. http://www.pnas.org/lookup/doi/10.1073/pnas.1510489113.

Athey, Susan, Julie Tibshirani, and Stefan Wager. "Generalized random forests". *The Annals of Statistics* 47, no. 2 (2019): 1148–1178.

Athey, Susan, and Stefan Wager. "Estimating Treatment Effects with Causal Forests: An Application". *arXiv:1902.07409 [stat]* (Feb. 20, 2019). Visited on 04/28/2021. arXiv: 1902.07409. http://arxiv.org/abs/1902.07409.

— . "Policy Learning with Observational Data". *arXiv:1702.02896 [cs, econ, math, stat]* (Sept. 4, 2020). Visited on 04/25/2021. arXiv: 1702.02896. http://arxiv.org/abs/1702.02896.

Becker, Gary S. *Human capital: a theoretical and empirical analysis, with special reference to education.* 3rd ed. Chicago: The University of Chicago Press, 1993. ISBN: 978-0-226-04119-3 978-0-226-04120-9.

Bjerk, David. "Re-examining the impact of dropping out on criminal and labor outcomes in early adulthood". *Economics of Education Review* 31, no. 1 (Feb. 2012): 110–122. ISSN: 02727757, visited on 01/20/2022. https://doi.org/10.1016/j.econedurev.2011.09.003. https://linkinghub.elsevier.com/retrieve/pii/S0272775711001506.

Breiman, L., et al. "Classification and Regression Trees". 1983.

Breiman, Leo. "Random Forests". *Machine Learning* 45, no. 1 (2001): 5–32. ISSN: 08856125, visited on 04/21/2021. https://doi.org/10.1023/A:1010933404324. http://link.springer.com/10.1023/A:1010933404324.

Brodaty, Thomas, Robert Gary-Bobo, and Ana Prieto. "Does Speed Signal Ability? The Impact of Grade Repetitions on Employment and Wages". *C.E.P.R. Discussion Papers, CEPR Discussion Papers* (Jan. 1, 2008).

Card, David. *Using Geographic Variation in College Proximity to Estimate the Return to Schooling*. NBER Working Papers 4483. National Bureau of Economic Research, Inc, 1993. https://EconPapers.repec.org/RePEc:nbr:nberwo:4483.

Chernozhukov, Victor, et al. "Locally robust semiparametric estimation". *arXiv preprint arXiv:1608.00033* (2016).

Fang, Hanming. "DISENTANGLING THE COLLEGE WAGE PREMIUM: ESTIMATING A MODEL WITH ENDOGENOUS EDUCATION CHOICES". *International Economic Review* 47, no. 4 (Nov. 2006): 1151–1185. ISSN: 0020-6598, 1468-2354, visited on 11/25/2020. https://doi.org/10.1111/j.1468-2354.2006.00409.x. http://doi.wiley.com/10.1111/j.1468-2354.2006.00409.x.

Flores-Lagunes, Alfonso, and Audrey Light. "Interpreting sheepskin effects in the returns to education" (Apr. 1, 2007).

Fox, John, and Georges Monette. "Generalized Collinearity Diagnostics". *Journal of the American Statistical Association* 87, no. 417 (Mar. 1992): 178–183. ISSN: 0162-1459, 1537-274X, visited on 01/19/2022. https://doi.org/10.1080/01621459.1992.10475190. http://www.tandfonline.com/doi/abs/10.1080/01621459.1992.10475190.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York, 2009. ISBN: 978-0-387-84857-0 978-0-387-84858-7, visited on 11/28/2020. https://doi.org/10.1007/978-0-387-84858-7. http://link.springer.com/10.1007/978-0-387-84858-7.

Imbens, Guido W., and Joshua D. Angrist. "Identification and Estimation of Local Average Treatment Effects". *Econometrica* 62, no. 2 (Mar. 1994): 467. ISSN: 00129682, visited on 11/26/2020. https://doi.org/10.2307/2951620. https://www.jstor.org/stable/2951620?origin=crossref.

Mahjoub, Mohamed-Badrane. "The treatment effect of grade repetitions". *Education Economics* 25, no. 4 (July 4, 2017): 418–432. ISSN: 0964-5292, 1469-5782, visited on 05/31/2021. https://doi.org/10.1080/09645292.2017.1283006. https://www.tandfonline.com/doi/full/10.1080/09645292.2017.1283006.

Matkovic, T., and I. Kogan. "All or Nothing? The Consequences of Tertiary Education Non-Completion in Croatia and Serbia". *European Sociological Review* 28, no. 6 (Dec. 1, 2012): 755–770. ISSN: 0266-7215, 1468-2672, visited on 01/20/2022. https://doi.org/10.1093/esr/jcr111. https://academic.oup.com/esr/article-lookup/doi/10.1093/esr/jcr111.

Navarro, Salvador, Jane Fruehwirth, and Yuya Takahashi. "How The Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatment Effects". *Journal of Labor Economics* 34 (2016). https://doi.org/10.1086/686262.

Psacharopoulos, George, and Harry Anthony Patrinos. "Returns to investment in education: a decennial review of the global literature". *Education Economics* 26, no. 5 (Sept. 3, 2018): 445–458. ISSN: 0964-5292, 1469-5782, visited on 06/01/2021. https://doi.org/10.1080/09645292.2018.1484426. https://www.tandfonline.com/doi/full/10.1080/09645292.2018.1484426.

Reisel, Liza. "Is more always better? Early career returns to education in the United States and Norway". *Research in Social Stratification and Mobility* 31 (Mar. 2013): 49–68. ISSN: 02765624, visited on 01/31/2022. https://doi.org/10.1016/j.rssm.2012.10.002. https://linkinghub.elsevier.com/retrieve/pii/S0276562412000558.

Rubin, Donald B. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology* 66, no. 5 (1974): 688–701. ISSN: 0022-0663, visited on 11/28/2020. https://doi.org/10.1037/h0037350. http://content.apa.org/journals/edu/66/5/688.

Schnepf, Sylke V. *Do Tertiary Dropout Students Really Not Succeed in European Labour Markets?* IZA Discussion Papers 8015. Bonn: Institute for the Study of Labor (IZA), 2014. http://hdl.handle.net/10419/96694.

Scholten, Mirte, and Nicole Tieben. "Vocational qualification as safety-net? Education-to-work transitions of higher education dropouts in Germany". *Empirical Research in Vocational Education and Training* 9 (Dec. 2017). https://doi.org/10.1186/s40461-017-0050-7.

Spence, Michael. "Job Market Signaling". *The Quarterly Journal of Economics* 87, no. 3 (Aug. 1973): 355. issn: 00335533, visited on 11/29/2020. https://doi.org/10.2307/1882010. https://academic.oup.com/qje/article-lookup/doi/10.2307/1882010.

Vignoles, Anna F., and Nattavudh Powdthavee. "The Socioeconomic Gap in University Dropouts". *The B.E. Journal of Economic Analysis & Policy* 9, no. 1 (2009). https://doi.org/doi:10.2202/1935-1682.2051. https://doi.org/10.2202/1935-1682.2051.

Wager, Stefan, and Susan Athey. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests". *Journal of the American Statistical Association* 113, no. 523 (July 3, 2018): 1228–1242. issn: 0162-1459, 1537-274X, visited on 04/23/2021. https://doi.org/10.1080/01621459.2017.1319839. https://www.tandfonline.com/doi/full/10.1080/01621459.2017.1319839.

Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data.* 2nd ed. Cambridge, Mass: MIT Press, 2010. isbn: 978-0-262-23258-6.

6 Appendix

6.1 IV conditional on individual characteristics

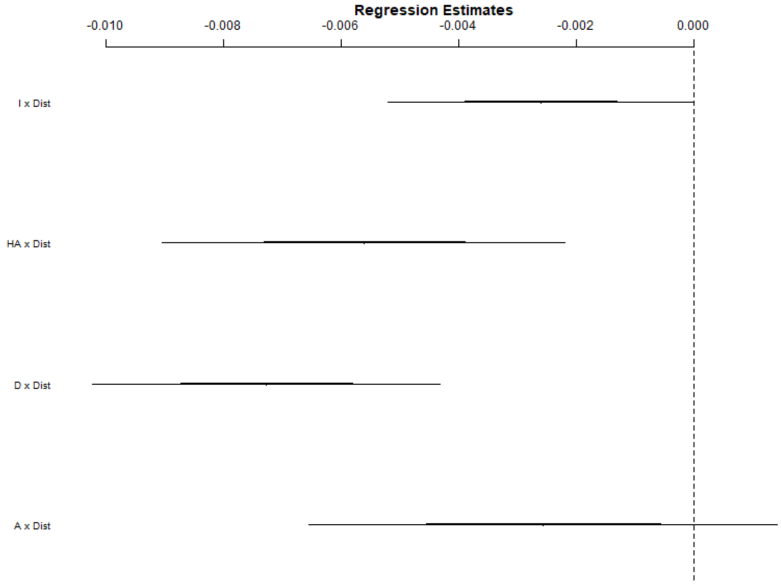


Figure 10: First stage IV - conditional on the parents' social category

D = disadvantaged, I = Intermediate, A = advantaged, HA = highly advantaged

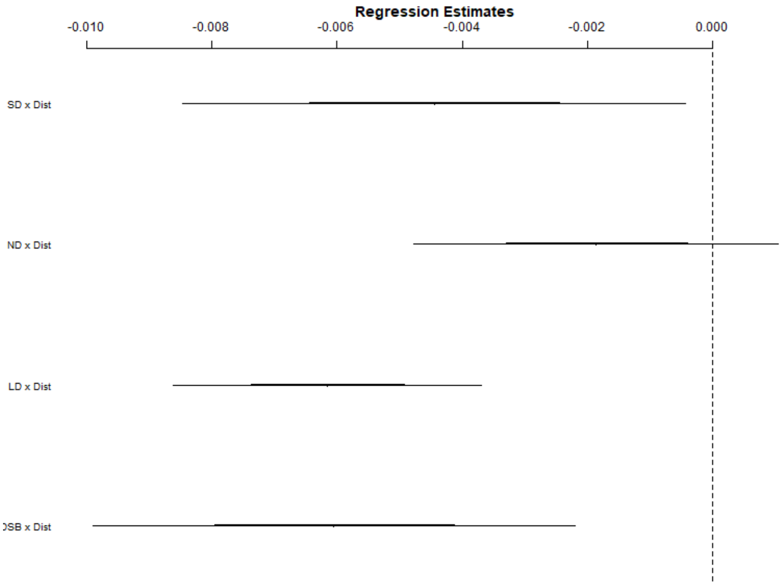


Figure 11: First stage IV - conditional on the parents' highest diploma

SD = Short degree, ND = No diploma, LD = Long Diploma, HSDB = High school diploma or below

6.2 GVIF : Assessing potential multicolinearity

	GVIF	Df	GVIF^(1/(2*Df))	V4
Gender	1.16	1.00	1.08	1.16
Highest dip (12 levels)	3.41	10.00	1.06	1.13
Social origin	1.51	3.00	1.07	1.15
Parents highest dip	1.82	3.00	1.10	1.22
HSD (field)	1.92	3.00	1.11	1.24
HSD (grade)	1.24	3.00	1.04	1.07
HSD Region	40679.90	22.00	1.27	1.62
Higher education Region	42640.38	22.00	1.27	1.62
Foreign trip	1.42	2.00	1.09	1.19
Internship	1.13	1.00	1.06	1.13

As suggested in Fox and Monette, using GVIF^(1/(2*Df) allows to compare the value of GVIF across different number of parameters. I elevate this measure to the square to use the standard rule of thumb of GVIF. Here, no GVIF goes above 2, so I can safely include and interpret all the parameters in the Logit model.

6.3 Logit table : top 50% for CATE on time in employment

	Dependent variable:
	median_2
Female	-0.255^{***}
	(0.033)
Highest diploma tried	
Bac +3/4	-0.649^{***}
	(0.044)
Bac+5 University	-0.257^{***}
	(0.050)
Grande Ecole	0.501^{***}
	(0.073)
PhD	-0.797^{***}
	(0.060)
Other academic characteristics	
Foreign trip : One	1.358^{***}
	(0.051)
Foreign trip : Many	-0.047
	(0.071)
Internship : Yes	0.528^{***}
	(0.034)
Social category	
Disadvantaged	-0.105^{**}
	(0.053)
Advantaged	-0.310^{***}
	(0.048)
Highly Advantaged	0.165^{***}
	(0.045)
Highest diploma (parents)	
HSD or below	0.649^{***}
	(0.048)
Long degree	0.517^{***}
	(0.061)
Short degree	1.816^{***}
	(0.054)
Observations	24,201
Log Likelihood	-12,301.630
Akaike Inf. Crit.	24,733.250
Note:	*p<0.1; **p<0.05; ***p<0.01

Reference levels : Bac +2 (Highest diploma tried), No diploma (Parents' diploma), Intermediate (Parents' social category).

6.4 Logit table : top 50% for CATE on log(average wage)

	Dependent variable:
	$\hat{\tau} > Median(\hat{\tau})$
Female	-0.255^{***}
	(0.033)
Highest diploma tried	
Bac +3/4	-0.649^{***}
	(0.044)
Bac+5 University	-0.257^{***}
	(0.050)
Grande Ecole	0.501^{***}
	(0.073)
PhD	-0.797^{***}
	(0.060)
Other academic characteristics	
Foreign trip : One	1.358^{***}
	(0.051)
Foreign Trip : Many	-0.047
	(0.071)
Internship : Yes	0.528^{***}
	(0.034)

Social category	
Disadvantaged	-0.105^{**}
	(0.053)
Advantaged	-0.310^{***}
	(0.048)
Highly Advantaged	0.165^{***}
	(0.045)
Highest diploma (parents)	
HSD or below	0.649^{***}
	(0.048)
Long degree	0.517^{***}
	(0.061)
Short degree	1.816^{***}
	(0.054)
Observations	24,201
Log Likelihood	-12,301.630
Akaike Inf. Crit.	24,733.250
Note:	*p<0.1; **p<0.05; ***p<0.01

Reference levels : Bac +2 (Highest diploma tried), No diploma (Parents' diploma), Intermediate (Parents' social category).

6.5 Time in employment distribution conditional on dropout and individual characteristics

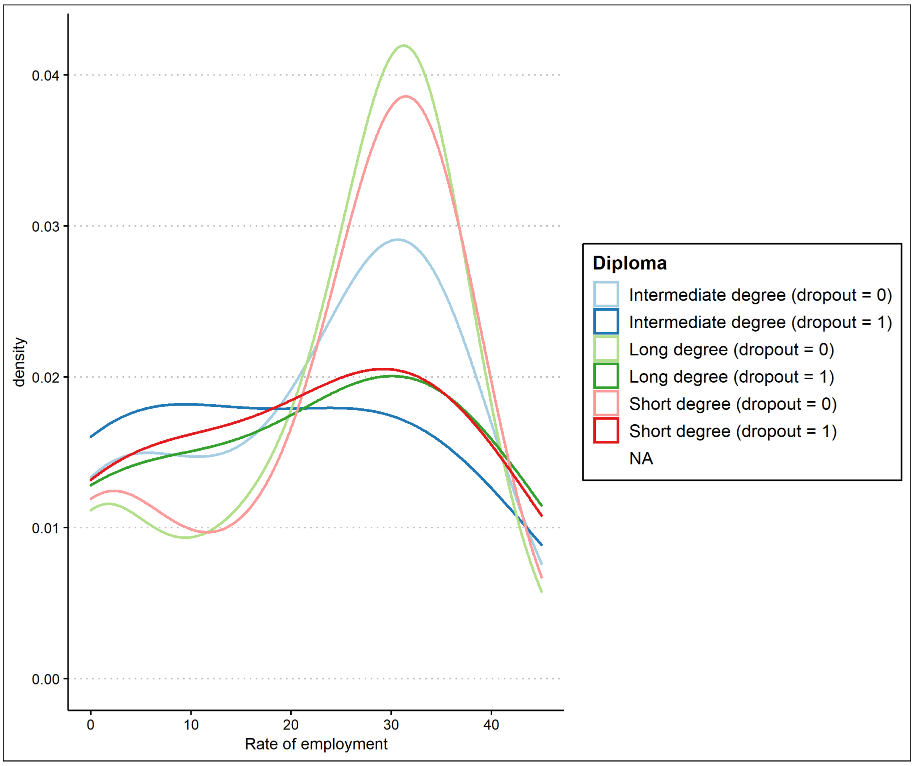


Figure 12: Time in employment distribution conditional on highest diploma tried

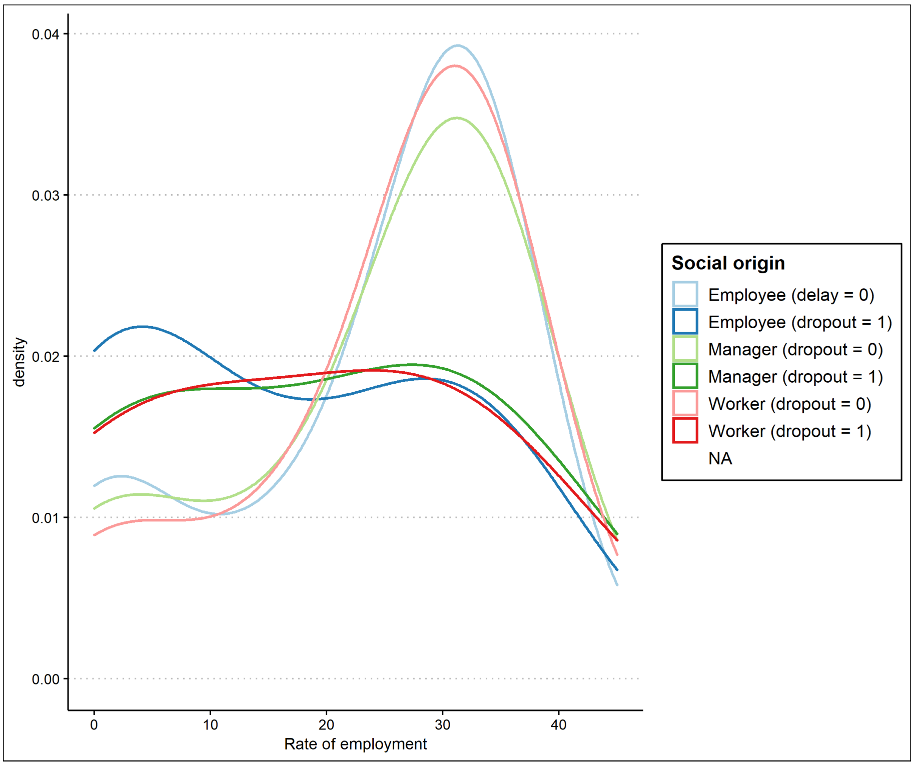


Figure 13: Time in employment distribution conditional on the mother's occupation

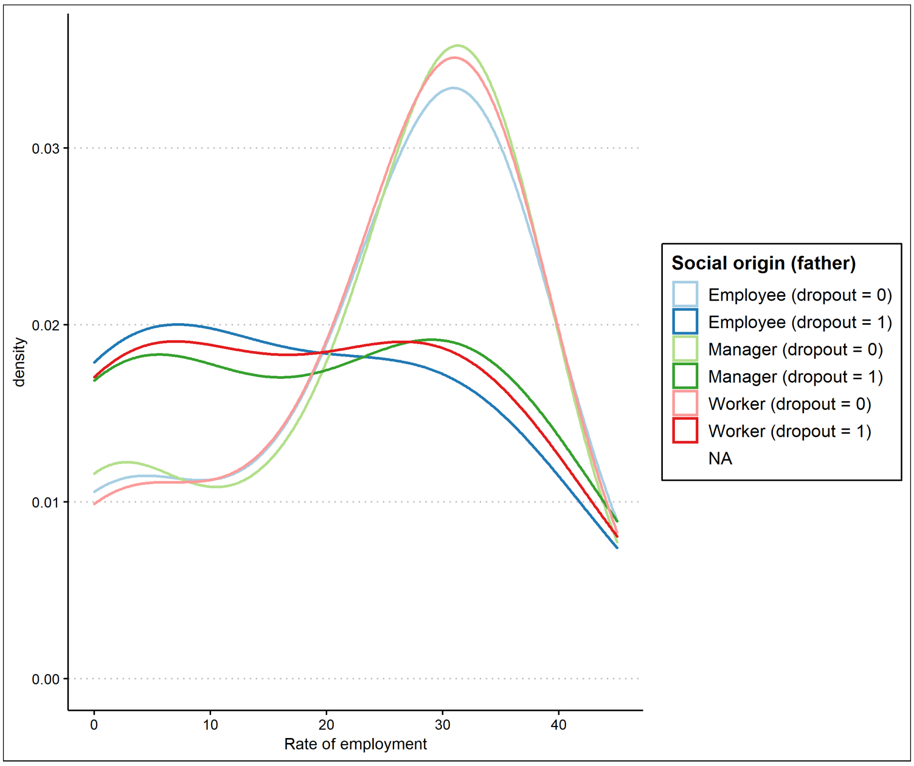


Figure 14: Time in employment distribution conditional on the father's occupation

6.6 Time in employment distribution conditional on dropout and individual characteristics

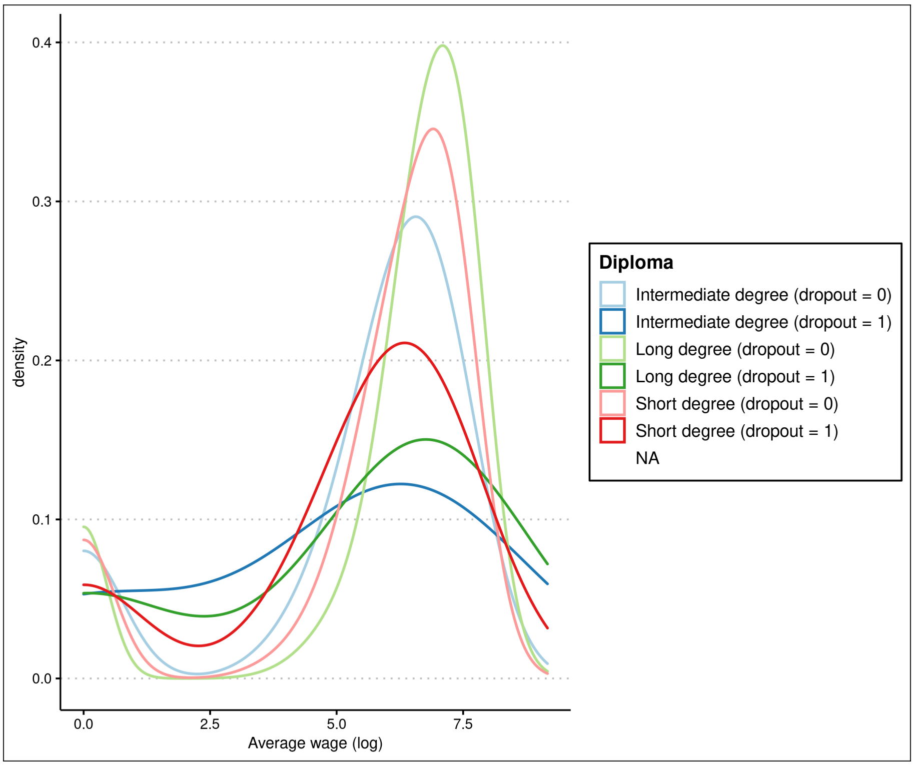


Figure 15: Average wage distribution conditional on highest diploma tried

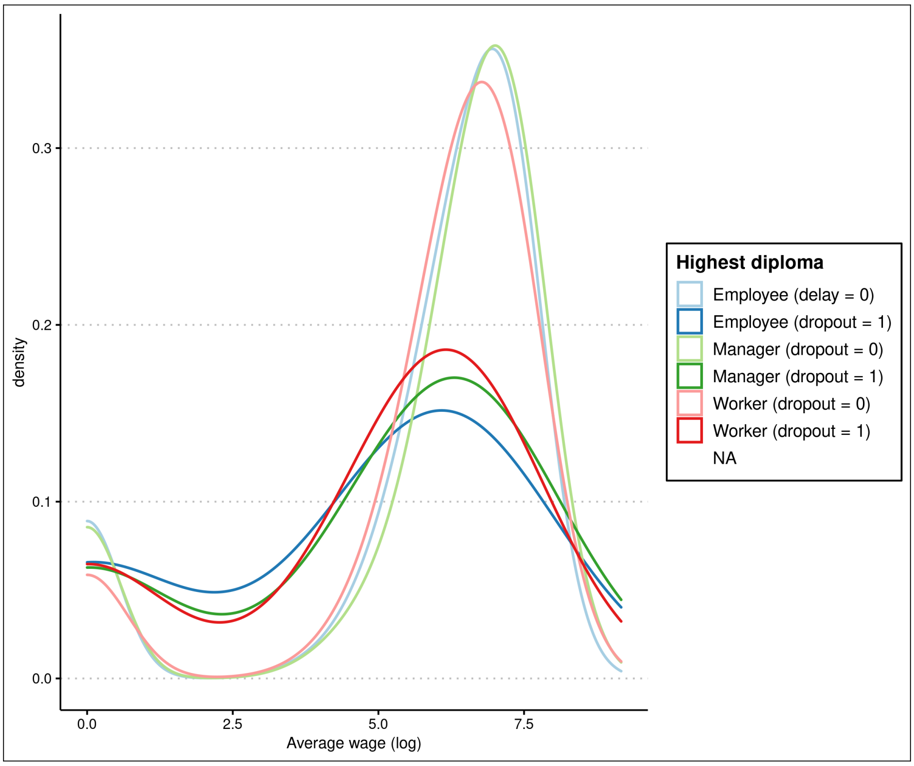


Figure 16: Average wage distribution conditional on the mother's occupation

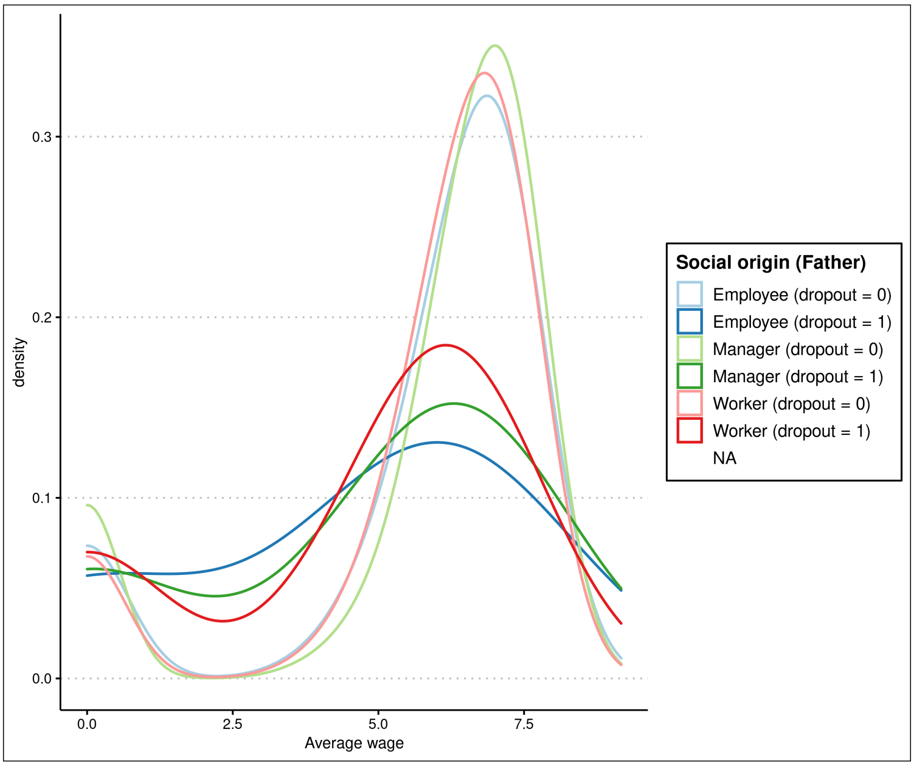


Figure 17: Average wage distribution conditional on the father's occupation