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

Gaspard Tissandier - Université Paris 1 Panthéon Sorbonne

20/01/2022

Abstract:

In the sense of Spence (1973), 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 rate of 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 individual Conditional Average Treatment Effects. In line with the literature, I find a negative effect of delayed graduation on the rate of employment. However, this effect is heterogeneous across individuals, ranging from -71 pp to -42 pp. I find two subgroups with a significant and negative effect of dropping out on the average wage, with effects ranging from -1200€ per month to a null effect. The gender, study field, study duration, and the social origin of parents, especially the mother's one, 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. 2018). Dropping out has a strong effect on labor market outcomes and can penalize new entrants in the long run ((Schnepf 2014)). 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 (1973), 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 2007, Matkovic and Kogan 2012 ,Reisel 2013, Schnepf 2014) . It has been observed that the negative effect of dropout can be heterogeneous conditional on the motivation of the dropout (Bjerk 2012) or even on the diploma or the age of the individual (Brodaty, Gary-Bobo, and Prieto 2008, Navarro, Fruehwirth, and Takahashi 2016, Scholten and Tieben 2017).

It has been shown that the social origin and individual characteristics of students are highly determinant in the dropout process, and that specific variables such has the academic path or the gender have a 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 with respect to a dense set of covariables regarding individual characteristics. Thus, is it crucial to understand the effect of dropout and for the orientation of higher education policy to understand which students are the most penalized by 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 charac-

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

teristics and more specifically if exogenous variables such as social origin or gender play a more important role than endogenous variables such as diplomas.

According to both fundamental models of education (Becker 1993 or Spence 1973), acquiring more year of education bring higher earnings, as detailed in the analysis of Fang (2006) or the review of Psacharopoulos and Patrinos (2018). The role of higher education as a signaling method is widely explored in the literature. The study of Arcidiacono, Bayer, and Hizmo (2010) 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 labor market to better fitting individuals more quickly.

However, dropout raises less consensus about its effects on labor market performance. As highlighted by Schnepf (2014), the literature about labor market performance of dropout is scarce. Bjerk (2012) 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 finding 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 (2014), 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 with Reisel (2013). The author finds that in the United States, it is actually beneficial to integrate higher education even without graduating. Similarly, Matkovic and Kogan (2012) compare the effect of dropping out on labor market performances in Croatia and Serbia and corroborate 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 (2007) 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 2008 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 (2013) finds heterogeneity in the return to education due to the distribution of women and minorities across the income distribution, while Scholten and Tieben 2017 finds that in Germany, for individual 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 (2014). On the other hand, recent papers like Mahjoub 2017 use the period of birth as an instrument, inspired by Angrist and Krueger (1991). An alternative instrument is the distance to the closest higher education institution, as proposed by Card (1993). In Brodaty, Gary-Bobo, and Prieto (2008), 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 (2018), on a French database of 18000 young workers who finished their education in 2010. Their work records are surveyed from 2010 to 2013, which helps us to construct two variables indicating the average of wages and rate of employment for every individual. The GRF algorithm, based on the Random Forest structure (Breiman 2001), allows us to estimate individual Conditional Average Treatment Effect (CATE). Individuals are then gathered in subgroups following their CATE magnitude to compute Local Average Treatment Effect (LATE) on these subgroups (Imbens and Angrist 1994). The asymptotic property of the 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 the parameters of a logistic model for being less penalized by dropout. This method takes into account the highly dimensional interactions between variables by using the belonging in the top 50% of the CATE distribution as a dichotomous variable, and not the individual CATE directly.

The endogeneity of delay in graduation 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 delayed graduation, I obtain an efficient instrumental variable setting allowing me to identify heterogeneous causal effects of delayed graduation 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 institution and the high school diploma city (Brodaty, Gary-Bobo, and Prieto 2008).

I find a statistically significant heterogeneous structure of the CATE on the rate of employment ranging from -58 to -42 percentage points. I succeed to identify heterogeneous subgroups for CATE on monthly average wages, with individuals showing a negative effect from -1250 to -25€ per month as average wage. These individuals are grouped following their CATE: above and below the CATE median for the distribution of both indicators. The LATE is then estimated on each subgroup. The difference between the LATE of these subgroups is tested with a Student test. To understand which individual characteristics are the most important in shaping the individual's CATE distributions, I estimate a logit model for the likelihood of ending above the median.

For the CATE of dropping out on the rate of employment, the most important variables are the mother's diploma and occupation. Having a mother who acquired more years of education or work in a higher position (and blue-collar position) will increase a student's likelihood of being in the less penalized part of the CATE distribution. The student's diploma is also significant with a clear advantage for individuals who tried to acquire short and professional degrees. Compared to fathers having a High School diploma or an employee occupation, all others education and occupation are decreasing the likelihood of being above the median. For this indicator, the social origin is the most important element for reducing the penalty of dropping out.

When studying the dropout CATE on average wages, the parents' occupation and education play a far less active role in shaping the CATE distribution. The father's diploma does not have an effect anymore and the mother's diploma exhibits smaller coefficients. The effect of the mother's occupation is still similar while more heterogeneous depending on the occupation. The highest diploma tried is however more important in determining the magnitude of the dropout effect. Short and professional degrees are still reducing the dropout penalty while longer degrees now reduce the likelihood of being in the top 50% of the distribution.

Generally, social origin and especially the mother's one are more important than the tried diploma to determine the magnitude of the dropout effect. However, the prevalence for exogenous variables is less pronounced for the average wage than for the rate of employment.

This paper sheds new light on the effect of higher education dropout on the rate of employment and average wages of former students in their first three years on the labor market. An important contribution is the application of a machine learning technique that helps to account for individuals' heterogeneity and propose better design policy when discussing higher education paths. The main result departing from the literature is the vast heterogeneity of treatment effect raised by the accounting of complex interactions between covariates. I find that the treatment effect is conditional on individual characteristics such as diploma, social origin, or gender.

2 Data

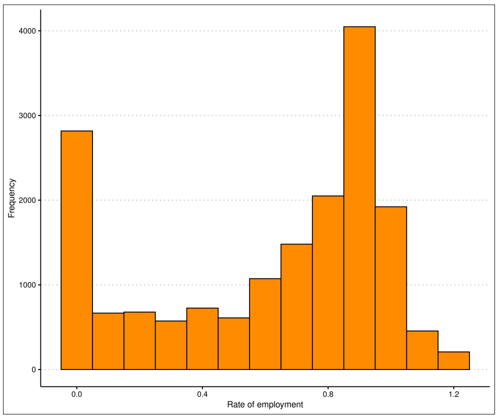
To identify the effect of dropping out on former students' labor market outcomes, I use "Génération 2010", a longitudinal survey provided by the CEREQ (Centre d'Etudes et de Recherches sur les Qualifications) [2]. This survey is conducted on individuals who have finished their education in 2010 (between 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 education to Ph.D. graduate. This represents a data set of 17305 individuals.

I create indicators variables for dropping out, the individual rate of employment and average wage. 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, 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 drop out. The database consists of 4335 individuals who dropped out, and 12970 who didn't.

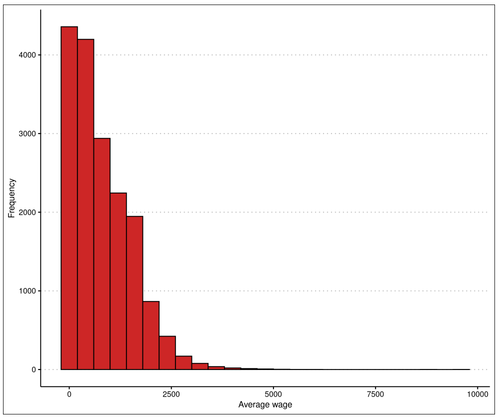
[2] Génération 2010 – Interrogation à 3 ans – 2013 (2013, CEREQ)

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

The first variable, *Rate of Employment*, consists in the average of months worked over 36 months (3 years). For example, if a former student work 12 months, she will have a $RoE = 33\%$. 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€.



(a) Rate of Employment



(b) Average wages

Figure 1: Distribution of dependent variables

I keep in the dataset the following variables : diploma on 32 levels and 12 levels, geographical situation 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 and diplomas of both parents and information about past education such as an indicator of having delay in 6th grade and the discretize grade of the high school diploma. 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 rate of employment and average wage conditional on dropout and individual characteristics are presented in Appendix.

In the notation, HSD correspond to the High School Diploma (or Baccalaureate).

	No dropout	Dropout	No dropout (in %)	Dropout (in %)	Total	Percentage
			Gender			
Male	5472	2239	70.96 %	29.04 %	7711	45.11%
Female	7313	2070	77.94%	22.06%	9383	54.89%
			Highest diploma tried			
HSD +2 Health	3006	48	98.43%	1.57%	3054	17.87%
HSD +2 Industrial	827	416	66.53%	33.47%	1243	7.27%
HSD +2 Tertiary	1290	2339	35.55%	64.45%	3629	21.23%
Vocational Degree	952	60	94.07%	5.93%	1012	5.92%
HSD +3/4 Soc sci., Econ Law	955	690	58.05%	41.95%	1645	9.62%
HSD +3/4 STEM	396	287	57.98%	42.02%	683	4.00%
HSD +5 Soc sci., Econ Law	1408	112	92.63%	7.37%	1520	8.89%
HSD +5 STEM	865	43	95.26%	4.74%	908	5.31%
Business School	343	12	96.62%	3.38%	355	2.08%
Ingeniering School	935	29	96.99%	3.01%	964	5.64%
PhD	1808	273	86.88%	13.12%	2081	12.17%
			Mother's highest diploma			
N.A	1328	532	71.40%	28.60%	1860	10.88%
No diploma	2428	1112	68.59%	31.41%	3540	20.71%
Below HSD	2013	685	74.61%	25.39%	2698	15.78%
HSD	2435	852	74.08%	25.92%	3287	19.23%
HSD +2 years	1618	407	79.90%	20.10%	2025	11.85%
HSD +3/4 years	1775	433	80.39%	19.61%	2208	12.92%
HSD 5+ years	1188	288	80.49%	19.51%	1476	8.63%
			Mother's occupation			
N.A	1314	596	68.80%	31.20%	1910	11.17%
Blue collar	1017	443	69.66%	30.34%	1460	8.54%
Employee	5735	1992	74.22%	25.78%	7727	45.20%
Intermediary	1031	299	77.52%	22.48%	1330	7.78%
White collar	3016	778	79.49%	20.51%	3794	22.19%
Independent	463	155	74.92%	25.08%	618	3.62%
Farmer	209	46	81.96%	18.04%	255	1.49%

Table 1: Summary statistics by dropout status

The time needed to acquire the diploma are counted as "+y" : HSD +2 corresponds to two years or study after the HSD. "Licence Pro" correspond to vocational degree, needing 3 years to complete. The short degrees (two and three years) leads to precise field, and are considered as "professional degree". The rest of the notation is common.

Concerning parents' occupation, the levels are defined using the type of occupation. It is actually entirely possible to have a Blue collar earning more than an Employee. The independent level groups craftsman and company owners. The farmer level is broad and include agronomic workers.

The unequal distribution of male and female is common and well documented in France, as female are more represented in higher education. The distribution of the professional situation of the mother shows that they are mostly employed in employee position, while it is more balanced for the fathers (see Appendix).

On table 1, we can also observe the distribution of dropout by highest diploma tried. The percentage presented for the dropout columns are the distribution of delay for the considered diploma, while it is the percentage of the considered diploma among the whole population in the total column. For short degrees holders, Health major are very less likely to dropout than tertiary or industrial graduates. For three or four years degrees, the percentage or dropout is the same with around 40% of the students who are dropping out. For all others degrees, the dropout rate is noticeably lower, with a maximum of 13% for Ph.D.

Regarding social origin, the dropout rates conditional on mother's education or professional occupation do not exhibits very different values.

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 Harvesine distance between both points. The geographical unit is the *zone d'emploi*, dividing France in around 310 area. 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 2. I use the square of the distance as an instrument.

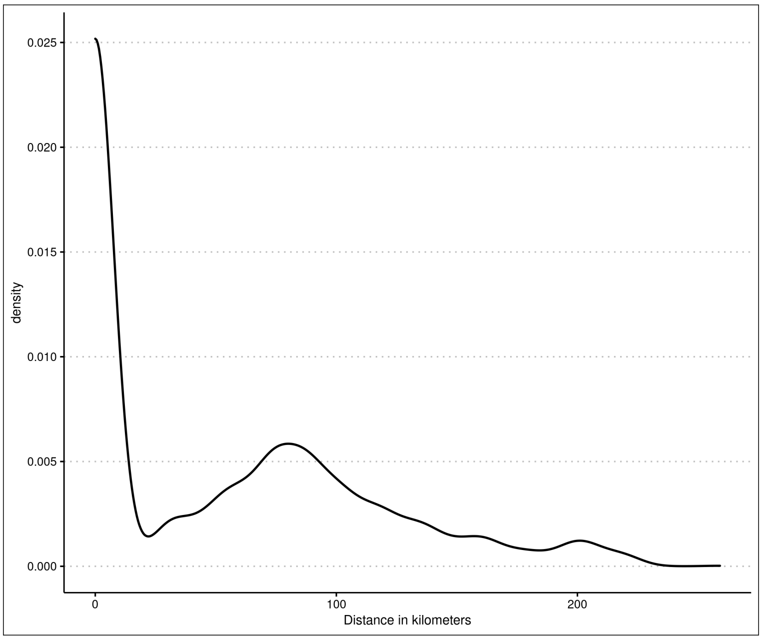


Figure 2: Distribution of the distance from 6th grade home to the closest university

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 an hypothesis of rational individual seeing education as a sequential process, the distance to the closest university will has a negative effect of educational attainment : for two similar students but one living farther than the other from the local university, the first one will have an higher cost of education, thus have higher probability of not engage in the next year.

The distance to the closest university also affects negatively the dropout probability by three mains vectors. First, the instrument affect 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 proportion, some benefits of having acquire 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 higher sunk cost, and then will be less likely to dropout. Finally, for students living close to an university, the cost of enrolling in a degree is lower than those living far. Then, it will be more likely for students without any real academic of professional project to enroll in higher education. These students being less motivated, they will be less likely to persist in programs, thus increasing the probability of dropout (Bernardo et al. 2016).

3 Methodology

The objective is to identify subgroups with different treatment effects of delayed graduation, conditional on a vector of covariates X. If we want to test for every interactions that this vector allow, 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 (2018), and use the data structure to identify heterogeneous treatment effect . This method allow us to compute Conditional Average Treatment Effect (CATE), the individual treatment effect, with standard error , and then to average these effects on selected partitions of the population as Local Average Treatment Effects (LATE). This method rely on regression trees to estimate the CATE, and average the estimated CATE across all trees. This methodology is called the Random Forest (Breiman 2001) and help us to dodge the dependency of estimated CATE to the initial random splitting needed with only few trees model. To avoid spurious correlations due to using similar data to construct the tree and estimating the treatment effect, the authors rely on the "honest methodology". As explained below, the endogeneity of the delay 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 Local Average Treatment Effect estimation.

The objective of our paper is to evaluate the causal effect of delayed in graduation. As defined by Rubin (1974), 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 the appendix 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 (2009), Athey and Imbens (2016), Wager and Athey (2018) or Athey, Tibshirani, and Wager (2018)).

The GRF is based on the regression tree algorithm developed by Breiman et al. (1983) (called the CART for Classification and Regression Trees) and adapted as causal honest tree by Athey and Imbens (2016). 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 wz can evaluate the Conditional Average Treatment Effects (CATE). 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 have delay in graduation she shows $W_i = 1$ and $W_i = 0$ if she has a study duration less than the mean of her diploma.

In the initial paper by Breiman et al. (1983), 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 (number of individuals in subsamples, 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 (2016), 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 base MSE, but add a negative effect of within subsample variance of the CATE. This allow the algorithm to 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 (2001) and applied to causal inference by Wager and Athey (2018). 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 individuals CATE given by all trees to compute the individual CATE. This method provide us 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 delay in graduation, I have to include a instrumental variable setting in our framework.

The Generalized Random Forest developed by Athey, Tibshirani, and Wager (2018) 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 us 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, the higher education institution region, the type of high school diploma (general, technical or professional), a categorical variable for the grade at the high school diploma, the professional occupation and diploma of both parents and the gender.

3.2 The instrumental variable setting

In this setting, I want to identify the effect of delayed in graduation on wages. 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 (showing delay in graduation) 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, found 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 use different type of IV. Knowing that the treatment is binary, we can implement an IV setting with a binary or continuous instrumental variable. However, since the Local Average Treatment Effect is used to estimate treatment effect on group of former students, we need a dichotomous instrumental variable. We need to compute the LATE as a doubly robust score as proposed by Athey and Wager (2020). This setting help to get unbiased results for every LATE. As precised in Athey and Wager (2020), 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 (2010) 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 having delay. 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 $(\alpha) = 0.80$. The LATE estimation and the compliance scores computation are developed in Appendix.

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

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 delayed in graduation, conditional on individuals' covariates.

3.3 The Local Average Treatment Effect

Our methodology produce individual treatment effect for $i = 1, ..., N$. Since I can't analyze every individual CATE, I need a way to group the individual treatment effects. For the Individual Treatment Effect on wages average, I divide individuals following if they present a above or below median effect for the rate of employment CATE distribution, or a positive or negative effects for the average wages distribution, and then estimate Local Average Treatment Effect (LATE) on each parts. Since the LATE follow a standard distribution, I perform a t-test to prove that our distribution of ITE present heterogeneity, and not only noise.

Then, the CATE distribution is segmented in clusters using the k-means algorithm, and the LATE is estimated on each clusters.

Following the "leave one out" method described by Athey and Wager (2020), I am able to compute LATE on each clusters, with the associated variance. Since the LATE present asymptotic property and follow the normal distribution, I can use the t-statistic to test for the significance of clusters' LATE.

4 Results

4.1 First stage regression

The first stage regression is run in the first model without controls, then with controls i.e having delay before 6th grade, diplomas and professional position of both parents, gender and the size of 6th grade city.

Dropout	(1)	(2)
(min^2 /100)	-0.000132	-0.00011
	(0.000033)	(0.000037)
Constant	0.26	0.49
	(0.0039)	(0.023)
Control variables	No	Yes
R Squared	0.001	0.034
F-Stat for the instrument	16.74	8.44

Table 2: Linear first stage

The initial correlation of the instrument with delay is strong, however it fall after controlling. Since the vector of controls is quite dense, I keep the logarithm of the minimum distance as an instrumental variable and consider it efficient. Using the predicted probability build with a vector of controls and the distance as an instrumental variable helps to get efficient and unbiased estimation of individual Conditional Average Treatment Effect.

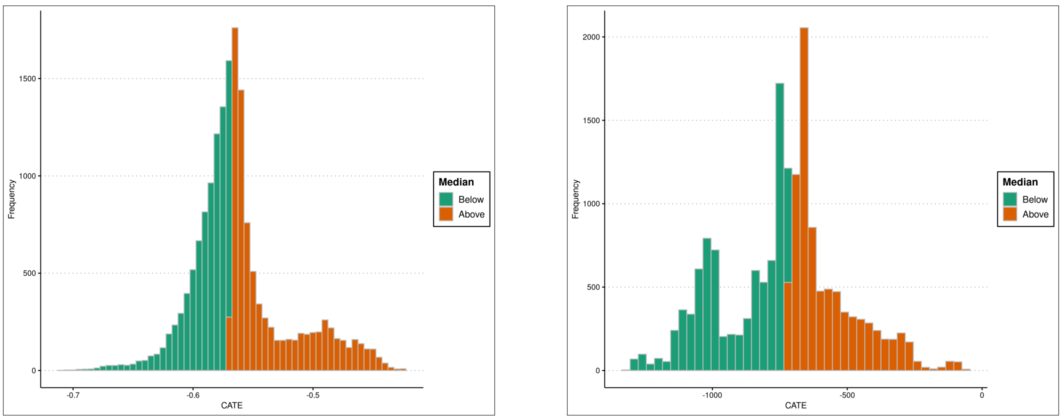
4.2 Assessing treatment heterogeneity

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 property to test for the presence of heterogeneity in the distribution.

Since we dispose of a a ordering of potential effect of dropout, the best estimator for an average effect is the Local Average Treatment Effect (Imbens and Angrist 1994). The LATE presents asymptotic property and tends to the standard distribution. It is possible to test for the difference between two LATE estimators using a Student test.

In this section, I split the dropout effect distributions on the rate of employment and the average wage around the median, compute the LATE on both subsamples and test for the statistical significance of their differences.

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



(a) Rate of Employment

(b) Average wages

Figure 3: Conditional Average Treatment Effect of delayed graduation

	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Rate of Employment	-0.71	-0.58	-0.57	-0.56	-0.55	-0.42
Average Wage	-1264.32	-809.34	-671.00	-678.79	-555.83	-25.17

Table 3: Distribution of CATE of dropout on the Rate of Employment and the Average Wage

The CATE on rate of employment range from -0.70 to -0.47, with the median at -0.30 and the mean at -0.55. This range indicate that individuals show CATE from a slight effect (less 10 more months of unemployment out of 36 months), to very important individuals effects with more than two years of unemployment. This wide gap on individuals effect needs to be validated trough a LATE groups difference test. Half of the individuals show a negative effect of -0.55 of less, which mean that around 50% of the sample could undergo a negative effect of more than a year and a half of unemployment, out of three years.

The CATE on average wages range from -1264€ to -25€, with median at -671€ and mean at -678€in monthly wage. This first step results seriously denote with precedent findings, as the distribution show individuals effects going from deeply negative (first quartile at -809€ per month) to almost null effect for the less disadvantaged individual.

For the both CATE distribution, I split the sample 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 $\hat{\tau}_1$ correspond to a group showing a strong negative effect of dropout, while the $\hat{\tau}_2$ correspond to a group showing slighter negative effect.

The LATE on each subgroups are presented in Table 4 for LATE on rate of employment, and in Table 5 for LATE on average wages.

	LATE	Std.Error	Mean($\hat{\tau}$)
$\hat{\tau}_1^{roe} \leq Median(\hat{\tau}^{roe})$	-0.78	0.09	-0.59
$\hat{\tau}_2^{roe} > Median(\hat{\tau}^{roe})$	-0.24	0.09	-0.53
T-Statistic	-4.39 (p.val $<$ 1%)		

Table 4: Local Average Treatment Effect of dropping out on the Rate of Employment

For the effect of dropping out on the rate of employment, we observe one subsamples with a strong negative effect on the compliers of -80 percentage points, and the other a negative effect of -24 percentage points. The student test of the difference of LATE between these two groups indicate that this difference is statistically significant at less than 1% level. As explained in the methodology part, the LATE always indicate more extreme results than the mean of CATE as the average of causal effect is divided by a weighting function based on the compliance scores, which is always less than 1. Consequently, while the average CATE is useful to understand the dynamic of our model, the LATE is necessary to effectively assess the causal effect of dropping out.

	LATE	Std.Error	Mean($\hat{\tau}$)
$\hat{\tau}_1^{aw} \leq Median(\hat{\tau}^{aw})$	-1703.65.97	260.05	-903.10
$\hat{\tau}_2^{aw} > Median(\hat{\tau}^{aw})$	-393.79	248.06	-569.43
T-Statistic		-3.64 (p.val $<$ 1%)	

Table 5: Local Average Treatment Effect of droping out on the Average Wages

The LATE of dropping out on the average wage indicate that 50% of the population is highly penalized by dropout with an effect of -1700€ per month on average wage. The top 50% of the distribution show a LATE of -393.79, but is not significantly different from 0. The difference between both LATE is significant at less than 1% confidence level.

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.

4.3 Social composition of subsamples

In this section, we study the social 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, we estimate a Logit model of being in the top 50% of the CATE distribution.

I include all the variables used to build the Generalized Random Forest. The main independent variables of interest are the gender, the diplomas, the education and the occupation of both parents. The objective of this analysis is, first, to understand how these variables are influencing the CATE distribution i.e if having parents with a certain occupation can increase the chance to be in the null effect group. Second, to understand which variables are the most important between exogenous and endogenous variables to shape the distribution of the CATE.

4.3.1 Rate of employment

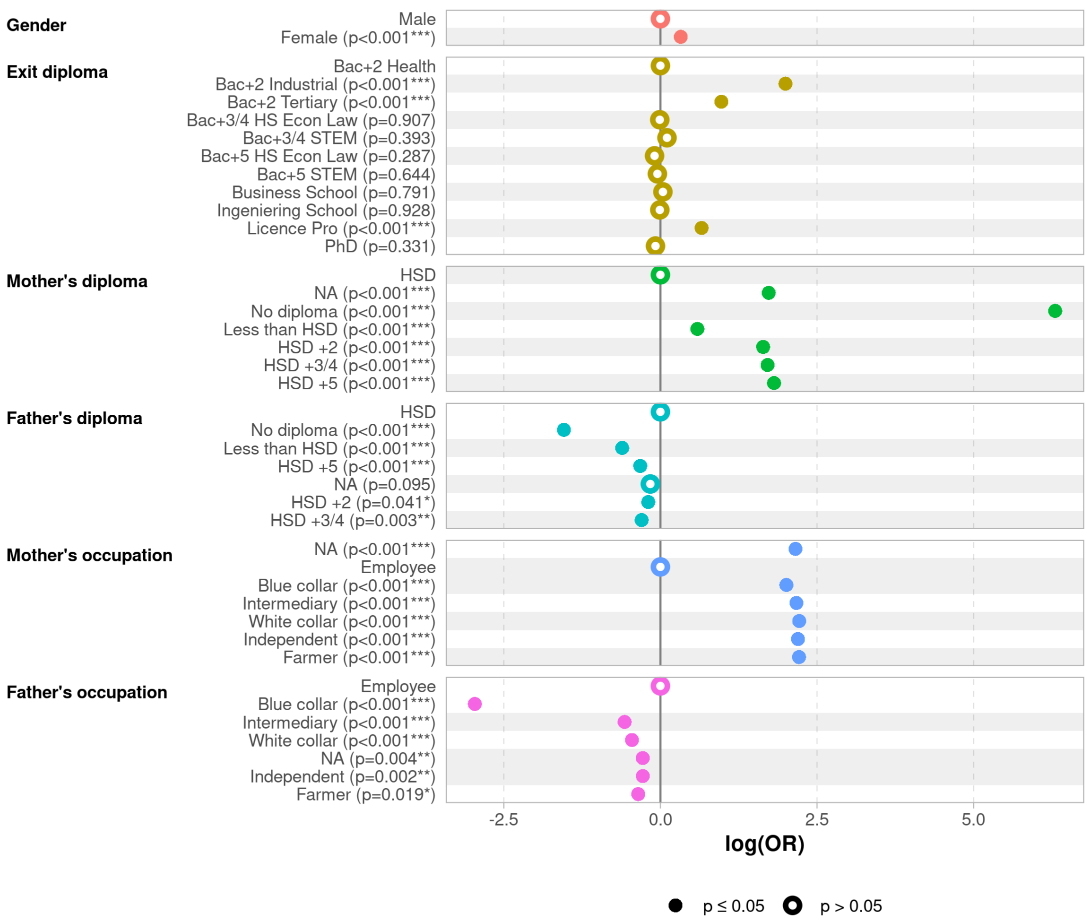
The logarithm of the odds ratios for the gender and highest diploma tried are reported in table **??**. The other odds ratio are reported in the Appendix. I perform the GVIF analysis in order to detect potential multicolinearity in the variables, however no measure goes above the recommended level. This analysis can be found table **??**. The results for every variables of interest are presented in the Appendix, section 6.2.

	Dependent variable:
	$\hat{\tau} > Median(\hat{\tau})$
Female	0.325***
	(0.046)
Highest diploma tried	
Bac+2 Industrial	1.996***
	(0.106)
Bac+2 Tertiary	0.973***
	(0.074)
Bac+3/4 HS Econ Law	−0.010
	(0.087)
Bac+3/4 STEM	0.105
	(0.123)
Bac+5 HS Econ Law	−0.094
	(0.088)
Bac+5 STEM	−0.048
	(0.104)
Business School	0.040
	(0.151)
Ingeniering School	−0.009
	(0.104)
Licence Pro	0.658***
	(0.104)
PhD	−0.078
	(0.081)
Observations	17,094
Log Likelihood	-6,939.218
Akaike Inf. Crit.	14,038.440
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6: Logit model for being in the top 50% (ROE)

Following the estimation of the Logit model, we observe that being a female represents an advantage to mitigate the negative effect of dropping out.

The effect of the highest diploma tried is mixed but with an overall trend : short degrees are less penalized by dropout. Dropping out of a short degree in health doesn't influence the magnitude of the effect on the rate of employment. However, very few students drop out of these diplomas (only 1.57%, as seen in table 1). For the rest of these professional degrees, the effect is positive and significant at the 1% level for the industrial field. This result can be explained by a high demand for these degrees, inciting the labor market to not penalized too much dropout. The important share of dropout can also lead employers to not penalized these students.



htbp

Figure 4: Log(odd ratios) for being in the top 50% of the CATE distribution (Rate of employment)

In this figure, the reference levels are also reported. The reference level for mother and father highest diploma is the high school diploma. All the other mother's diploma have a positive effect compared to having a mother with an HSD. Having a mother with no diploma increase a lot the likelihood of being above the median of the CATE. All the other levels show a positive effect compared to the HSD with a smaller effect for having a mother with less than a diploma. On the overall, the mother's education has a positive impact on the likelihood of being highly penalized by dropout. On the contrary, having a father with an HSD is the more advantaging social origin. All the others levels exhibit negative effects for being in the top 50%, with a decrease of the magnitude of this effect when the level of the diploma increase. The positive effect of the mother's diploma is higher than the father's diploma, indicating that it is the mother which can helps the most to mitigate the negative effect of dropout.

Regarding the parent's occupation, the reference level is the Employees, which is also the biggest group in terms of social original of the parents. Compared to having a employee mother, every other social origin have a positive impact on the likelihood of being the less penalized students, with a slightly smaller effect for the Blue collars. On the contrary, every other occupation than employee have a negative effect for the father. Having a Blue collar father is deeply penalizing, and all the other levels present almost the same negative effect.

By analyzing the conditional effect of dropout on the rate of employment, it appears that having an educated and well employed mother affects positively a student's probability of being less penalized by dropout, while the effect is completely reversed regarding the father's occupation and diploma. On the overall, the social origin of both is playing a far more active role in shaping the CATE distribution than the highest diploma tried.

4.3.2 Average wage

The logarithm of the odd ratios are reported in table 7. The complete table is reported in Appendix.

	Dependent variable:
	$\hat{\tau} > Median(\hat{\tau})$
Female	3.037***
	(0.059)
Highest diploma tried	
Bac+2 Industrial	1.492***
	(0.103)
Bac+2 Tertiary	1.663***
	(0.075)
Licence Pro	1.058***
	(0.107)
Bac+3/4 HS Econ Law	−0.067
	(0.087)
Bac+3/4 STEM	0.184
	(0.129)
Bac+5 HS Econ Law	−0.296***
	(0.089)
Bac+5 STEM	−0.139
	(0.111)
Business School	−0.365**
	(0.167)
Ingeniering School	−0.113
	(0.116)
PhD	−0.355***
	(0.084)
Observations	17,094
Log Likelihood	-6,444.328
Akaike Inf. Crit.	13,006.660
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: Logit model for being in the top 50% (Average wage)

With the short degree in Health as a reference, the case remains similar for short degree, with a strong probability of being in the top 50% of the CATE distribution. We can also observe that long degrees (5 years) non related to STEM field have lower probability of being in the top 50% than the reference group. The case is similar for Ph.D dropout.

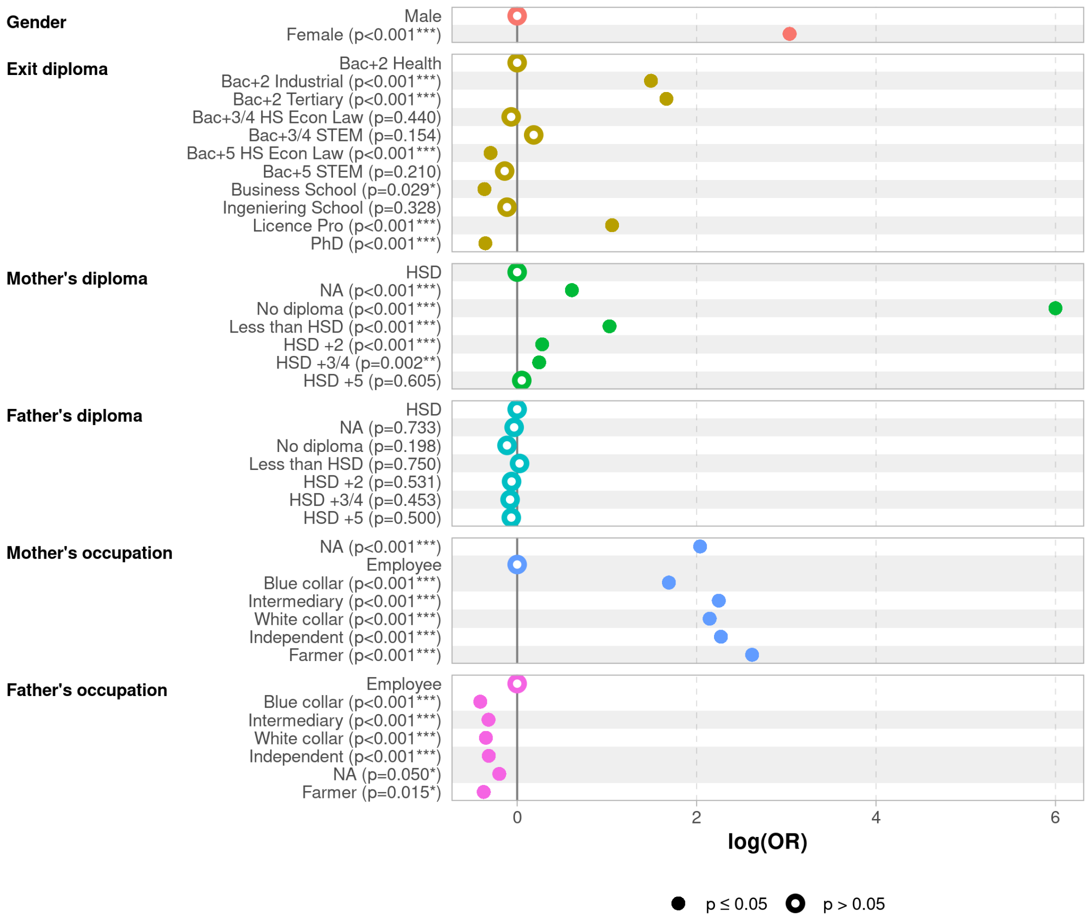


Figure 5: Log(odd ratios) for being in the top 50% of the CATE distribution (Average wage)

The effect of the mother's diploma is less straightforward in this situation. While having a mother with no diploma still affects strongly the likelihood of having a CATE above the median, the other levels do not have the same effects. A mother having less than an HSD is the second most positive effect compared to the HSD level, while longer degrees have only a slight positive effect. However, having a mother with 5 years of study or more is not significant anymore. Compared to the conditional effect measured in the first logit regression, having a mother with a longer study duration does not protect as much as on the CATE on rate of employment. It is also the case for the father's education, which does not impact the structure of the CATE as every levels don't have an effect significantly different from the HSD.

The mother's education, however, continue to plays a role in shaping the CATE distribution. All levels have a significantly different from 0 and positive effect compared to the Employee occupation. The Blue collar occupation have a slightly weaker effect than the others, while the farmer's children are less disadvantaged. The results for the father's occupation are quite similar. All the levels have negative effects compared to having an employee father.

While still being significant in the structure of the distribution of the CATE, the social origin of both parents have a less important role for the average wage than for the rate of employment. The parent's diploma and mostly the parent's occupation are the most important variables compared to the diploma, but are less affecting the distribution. The gender continue to have the same effect.

We have seen that, while being important in shaping the distribution of the dropout CATE on both indicators, diplomas are still less important than social origin of the parents, especially the mother's one. For the CATE on the rate of employment, the most important variables are the mother's education and occupation, while it is mostly the mother's occupation which drives the effect for the average wage. Concerning the tried diploma, it is less penalized to dropout from a short, professional degree than from longer degrees. On the overall, endogenous variables such as social origin and gender have a more important role in shaping the CATE distribution for the rate of employment, while this effect is less visible for the average wage.

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 Effect 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 rate of 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 Local Average Treatment Effect.

The diploma of workers plays a decisive role in the heterogeneity of the treatment effect. Short degree holders are less penalized in terms of rate of employment and the average wage. Dropping out has a negative effect on the average wage for longer degree dropouts.

The social origin is also detrimental in the amplitude of the effect on the rate of employment. Students coming from lower classes are mostly penalized, while children of executives or intermediate parents are far less penalized. However, this effect is mitigated for the average wage. The mitigation of the effects of exogenous variables once the monthly salary is accounted for shows encouraging results concerning the labor market capacity to perceive a uniform signal from delayed graduation.

This paper raises new results on the need for higher education policymakers to consider multi-dimensional effects of grade retention or re-orientation. 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 higher education and dropout.

The results on the influence of personal characteristics can help policymakers to consider the heterogeneity of situation for former students as an essential dimension in the delayed graduation debate.

References

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

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

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

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

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2018. "Generalized Random Forests". *arXiv:1610.01271 [econ, stat]* (). Visited on 04/15/2021. arXiv: 1610.01271. http://arxiv.org/abs/1610.01271.

Athey, Susan, and Stefan Wager. 2020. "Policy Learning with Observational Data". *arXiv:1702.02896 [cs, econ, math, stat]* (). Visited on 04/25/2021. arXiv: 1702.02896. http://arxiv.org/abs/1702.02896.

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

Bernardo, Ana, et al. 2016. "Comparison of Personal, Social and Academic Variables Related to University Drop-out and Persistence". *Frontiers in Psychology* 7 (). ISSN: 1664-1078, visited on 01/20/2022. doi:10.3389/fpsyg.2016.01610. http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01610/full.

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

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

Breiman, Leo. 2001. "Random Forests". *Machine Learning* 45 (1): 5–32. ISSN: 08856125, visited on 04/21/2021. doi:10.1023/A:1010933404324. http://link.springer.com/10.1023/A:1010933404324.

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

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

Fang, Hanming. 2006. "DISENTANGLING THE COLLEGE WAGE PREMIUM: ESTIMATING A MODEL WITH ENDOGENOUS EDUCATION CHOICES". *International Economic Review* 47, no. 4 (): 1151–1185. ISSN: 0020-6598, 1468-2354, visited on 11/25/2020. doi: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. 2007. "Interpreting sheepskin effects in the returns to education" ().

Fox, John, and Georges Monette. 1992. "Generalized Collinearity Diagnostics". *Journal of the American Statistical Association* 87, no. 417 (): 178–183. ISSN: 0162-1459, 1537-274X, visited on 01/19/2022. doi:10.1080/01621459.1992.

10475190. http://www.tandfonline.com/doi/abs/10.1080/01621459.1992.10475190.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning.* Springer Series in Statistics. New York, NY: Springer New York. ISBN: 978-0-387-84857-0 978-0-387-84858-7, visited on 11/28/2020. doi: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. 1994. "Identification and Estimation of Local Average Treatment Effects". *Econometrica* 62, no. 2 (): 467. ISSN: 00129682, visited on 11/26/2020. doi:10.2307/2951620. https://www.jstor.org/stable/2951620?origin=crossref.

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

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

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

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

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

Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology* 66 (5): 688–701. ISSN: 0022-0663, visited on 11/28/2020. doi:10.1037/h0037350. http://content.apa.org/journals/edu/66/5/688.

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

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

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

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

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data.* 2nd ed. OCLC: ocn627701062. Cambridge, Mass: MIT Press. ISBN: 978-0-262-23258-6.

6 Appendix

6.1 GVIF : Assessing potential multicolinearity

	GVIF	Df	GVIF^(1/(2*Df))	GVIF^(1/(2*Df))^2
gender	1.18	1.00	1.08	1.18
Highest diploma tried	1.79	10.00	1.03	1.06
Diploma (mother)	8.30	6.00	1.19	1.42
Diploma (father)	6.87	6.00	1.17	1.38
Occupation (mother)	3.54	6.00	1.11	1.23
Occupation (father)	6.83	6.00	1.17	1.38
Region of highest diploma	22682.19	22.00	1.26	1.58
Region of High School diploma	21370.35	22.00	1.25	1.57

Table 8: Various measure of VIF

As suggested in Fox and Monette 1992, 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.

6.2 Logit table : top 50% for CATE on rate of employment

	Dependent variable:
	median_2
Female	0.325***
	(0.046)
Highest diploma tried	
Bac+2 Industrial	1.996***
	(0.106)
Bac+2 Tertiary	0.973***
	(0.074)
Bac+3/4 HS Econ Law	−0.010
	(0.087)
Bac+3/4 STEM	0.105
	(0.123)
Bac+5 HS Econ Law	−0.094
	(0.088)
Bac+5 STEM	−0.048
	(0.104)
Business School	0.040
	(0.151)
Ingeniering School	−0.009
	(0.104)
Licence Pro	0.658***
	(0.104)
PhD	−0.078
	(0.081)
Mother's diploma	
NA	1.729***
	(0.093)
No diploma	6.302***
	(0.120)
Less than HSD	0.590***
	(0.081)
HSD +2	1.640***
	(0.078)
HSD +3/4	1.711***
	(0.080)
HSD +5	1.814***
	(0.094)
Father diploma	
NA	−0.161*
	(0.096)
No diploma	−1.540***
	(0.095)
Less than HSD	−0.610***
	(0.085)
HSD +2	−0.194**
	(0.095)
HSD +3/4	−0.299***
	(0.100)
HSD +5	−0.322***
	(0.091)
Mother's occupation	
NA	2.156***
	(0.077)
Blue collar	2.011***
	(0.096)
Intermediary	2.171***
	(0.081)
White collar	2.216***
	(0.063)
Independent	2.196***
	(0.114)
Farmer	2.213***
	(0.203)
Father's occupation	
NA	−0.281***
	(0.099)
Blue collar	−2.962***
	(0.104)
Intermediary	−0.572***
	(0.089)
White collar	−0.454***
	(0.073)
Independent	−0.281***
	(0.090)
Farmer	−0.354**
	(0.152)
Observations	17,305
Log Likelihood	-6,331.953
Akaike Inf. Crit.	12,825.910
Note:	*p<0.1; **p<0.05; ***p<0.01

Reference levels : Bac +2 Health (Highest diploma tried), High School Diploma/Baccalaureate (father/mother diploma), Employee (father/mother occupation).

6.3 Logit table : top 50% for CATE on average wage

	Dependent variable:
	$\hat{\tau} > Median(\hat{\tau})$
Female	3.037***
	(0.059)
Highest diploma tried	
Bac+2 Industrial	1.492***
	(0.103)
Bac+2 Tertiary	1.663***
	(0.075)
Licence Pro	1.058***
	(0.107)
Bac+3/4 HS Econ Law	−0.067
	(0.087)
Bac+3/4 STEM	0.184
	(0.129)
Bac+5 HS Econ Law	−0.296***
	(0.089)
Bac+5 STEM	−0.139
	(0.111)
Business School	−0.365**
	(0.167)
Ingeniering School	−0.113
	(0.116)
PhD	−0.355***
	(0.084)
Mother's diploma	
NA	0.610***
	(0.095)
No diploma	5.999***
	(0.125)
Less than HSD	1.029***
	(0.072)
HSD +2	0.279***
	(0.078)
HSD +3/4	0.245***
	(0.081)
HSD +5	0.051
	(0.099)

Father diploma	
NA	−0.034
	(0.099)
No diploma	−0.113
	(0.088)
Less than HSD	0.027
	(0.084)
HSD +2	−0.062
	(0.099)
HSD +3/4	−0.079
	(0.105)
HSD +5	−0.065
	(0.097)
Mother's occupation	
NA	2.039***
	(0.085)
Blue collar	1.691***
	(0.102)
Intermediary	2.247***
	(0.087)
White collar	2.146***
	(0.069)
Independent	2.271***
	(0.126)
Farmer	2.619***
	(0.215)
Father's occupation	
NA	−0.199**
	(0.102)
Blue collar	−0.411***
	(0.084)
Intermediary	−0.320***
	(0.090)
White collar	−0.348***
	(0.075)
Independent	−0.316***
	(0.093)
Farmer	−0.373**
	(0.153)
Observations	17,094
Log Likelihood	-6,444.328
Akaike Inf. Crit.	13,006.660
Note:	*p<0.1; **p<0.05; ***p<0.01

Reference levels : Bac +2 Health (Highest diploma tried), High School Diploma/Baccalaureate (father/mother diploma), Employee (father/mother occupation).

6.4 Rate of employment distribution conditional on dropout and individual characteristics

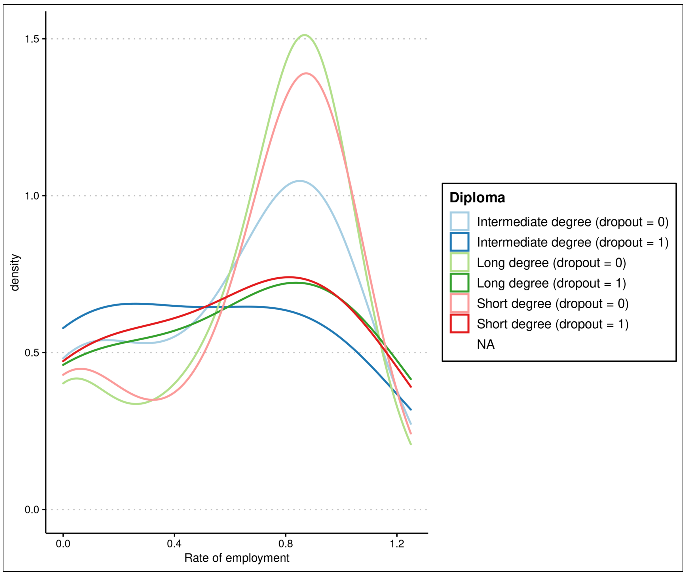


Figure 6: Rate of employment distribution conditional on highest diploma tried

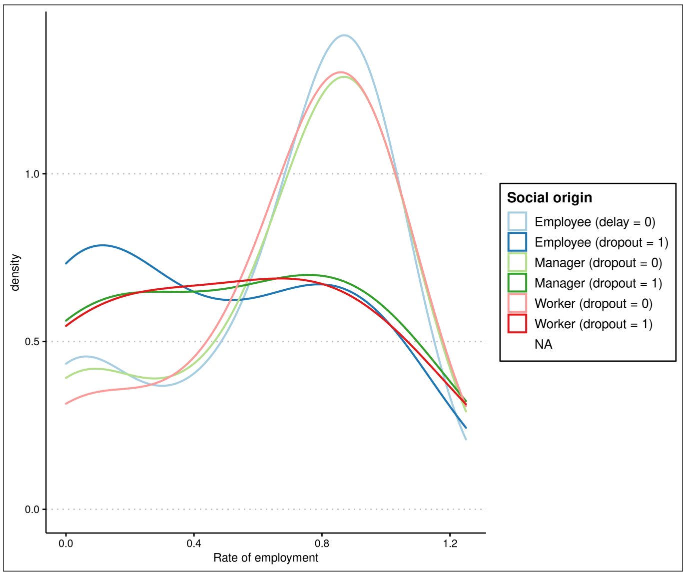


Figure 7: Rate of employment distribution conditional on the mother's occupation

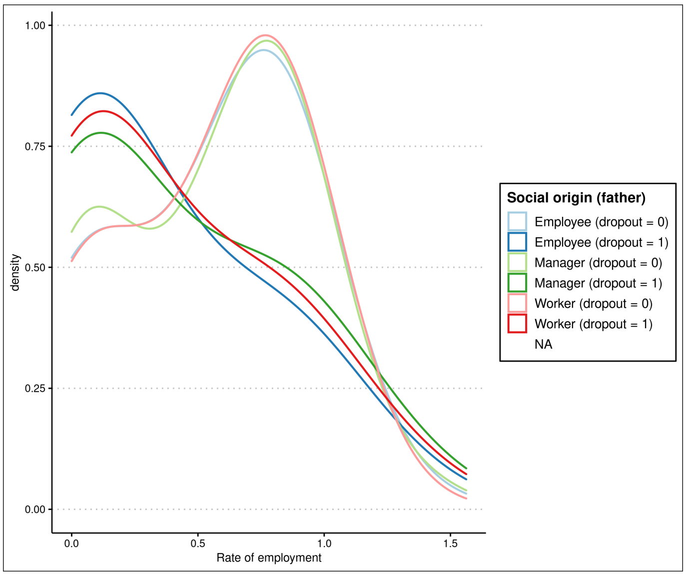


Figure 8: Rate of employment distribution conditional on the father's occupation

6.5 Rate of employment distribution conditional on dropout and individual characteristics

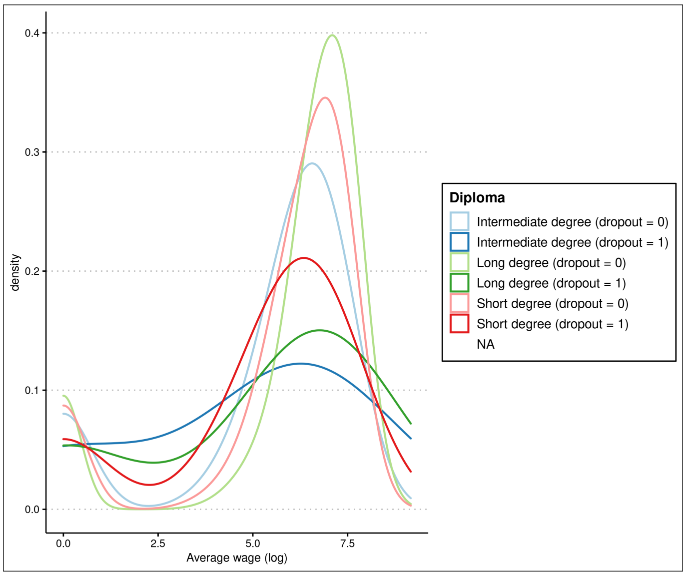


Figure 9: Average wage distribution conditional on highest diploma tried

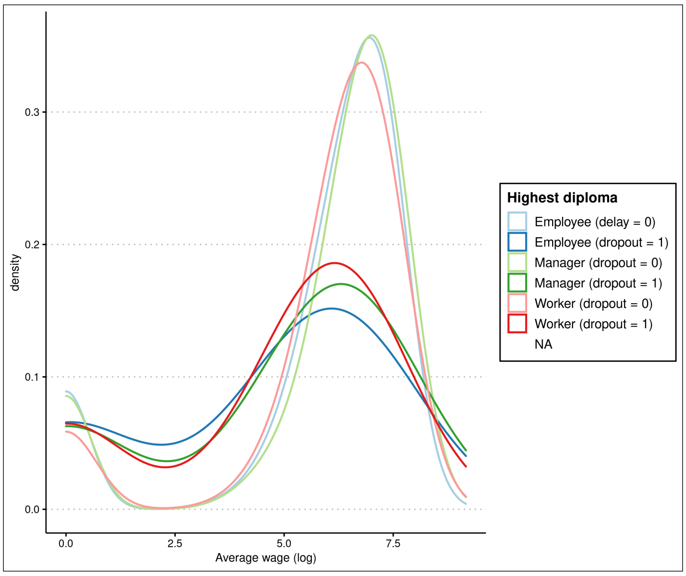


Figure 10: Average wage distribution conditional on the mother's occupation

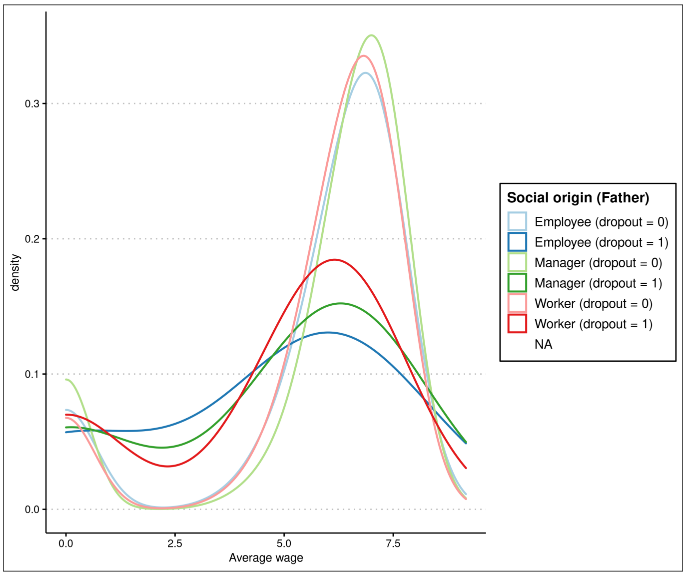


Figure 11: Average wage distribution conditional on the father's occupation