Does training pay?

Estimating the wage returns to vocational training in France

MASTER'S THESIS

Oliver CASSAGNEAU-FRANCIS*

22nd May 2017

Abstract

This paper exploits a novel dataset to estimate the wage returns to training in France. Using standard estimation techniques (OLS and diff-in-diff) the estimated returns are around 2-3%—similar to previous estimates in France and elsewhere. However, novel IV estimates—using data on whether firms provide information on training to their employees as the instrument—differ greatly suggesting issues with the assumptions required for OLS and differences-in-difference methods. Estimates obtained using Angrist and Imbens [1995]'s *saturate-and-weight* procedure produce an estimated weighted average treatment effect of 25%, suggesting high returns for those most affected by the instrument.

SciencesPo

*I would like to thank Professor Robin and Professor Gary-Bobo for their invaluable advice and guidance. Professor Gary-Bobo also kindly provided the data used in this paper. I am also grateful to Julie Pernaudet for her helpful advice, and in particular her suggestion of the IV I use in this paper. Finally, and most importantly, I thank my family and friends for their support: Fabio, Hugo and Théo for sharing this experience with me; Laélia and my parents for their constant support.

List of Figures

List of Tables

We are not heading towards Keynes' vision of a 15-hour work week; rather an increasingly cheaper supply of automated and offshore labour is changing the nature of the skills demanded of the labour force. Many displaced workers are not willing to suffer the loss in income that a return to formal education would require, making the ability to retrain while working a necessity. Given the current rate of technological progress, learning while earning is only likely to increase in importance. Governments are starting to catch on. Singapore recently introduced "lifelong learning accounts", providing its citizens with credits to spend on approved courses [Economist, 2017]. France introduced a similar system in 2015, *le compte personnel de formation* (CPF), which entitle all employees to a certain number of hours training per year, funded by their employer and/or the state.[1] Skills also play an increasing role in inequality. The increases in wages paid to those at the top of the skill distribution while others' wages stagnate and fall in real terms suggest that recent increases in inequality are linked to skills [Autor et al., 2003]. This is on top of the traditional view that career wage-paths are driven by the accumulation of human capital—through training as well as learning by doing. Ensuring the supply of skills keeps up with demand will require the ability to train while working for those already in the labour force, and this ability will likely play a vital role in the success of those yet to enter.

The increasing importance of vocational training, and for an increasingly flexible workforce raises important policy issues. It will be up to governments not only to intervene to address market failures in providing an efficient level of training (where economic theory has and will continue to play an important role), but also to ensure firms and workers have the evidence they need to make the right choices. This paper addresses the second concern, joining a growing literature which makes use of high-quality microdata to measure the returns to vocational training undertaken during employment[2]—as opposed to a related

[1] In fact, the French state has intervened in private sector training since the 1970s.

[2] This distinction is worth elaborating upon. Initially "on-the-job training" was used throughout this paper. However, that has the potential for confusion as some authors use it to mean "training in the course of work", rather than "training undertaken during employment" meant in this paper. To avoid such confusion I have attempted to use "vocational training" throughout; which in the context of this paper was almost exclusively undertaken while the individual was employed.

literature which focuses on training programmes for the unemployed. I make use of a novel French dataset collected by the *Centre d'études et des recherches sur les qualifications*[3] (Céreq), to estimate the wage returns to vocational training in France. The estimated returns using OLS and differences-in-differences specifications are similar to those found in previous studies employing these specifications (both in France and in other countries); however, the difference of these estimates to those obtained using a novel, policy-relevant IV suggest that the assumptions required for OLS and DiD specifications are too restrictive.

The paper is organised as follows. Section 1 discusses the economic theory and evidence concerning on-the-job training and the (unique) state of training in France. Section 2 describes the dataset. Section 3 introduces the model underlying the empirical analysis. The following sections presents different approaches and compare their results. Section 7 concludes with a discussion of further work.

1 Background

1.1 Training in an economic context

1.1.1 A brief history on the theory of vocational training

Pigou [1912]'s analysis of on-the-job training suggests firms will underinvest in training—workers' ability to leave a firm destroys firms' incentives to invest. Becker [1962] suggested there was no such problem; given perfect markets, workers pay for general training that improves their wage, while firms will pay for any specific training for which they capture the benefits. Training provision will be optimal. This belief was widely held up until the 1990s and led to the deregulation and dismantling of apprenticeship systems in the US and the UK [Greenhalgh, 1999]. As microdata became sufficiently detailed for economists to investigate empirically the returns to training Lynch [1992]. Under Becker's model, workers

[3] Centre for studies and research into qualifications.

will face a wage cut, or even negative wages while receiving general training, and will then receive all of the increased productivity in the form of higher wages—testable predictions. Becker's model seems to fit a number of industries, such as law where workers initially accept very low wages, which they suffer in anticipation of high future compensation. In other industries, however, firm's appear to fund general training. Autor [2001] found this to be the case among temporary help firms.

Firm funded general training suggests that Becker's analysis did not capture the whole story. Numerous authors have explored market frictions that give firm sufficient incentives to invest in general training. Acemoglu and Pischke [1999c] present a model where wage compression in imperfect labour markets can provide sufficient incentives for firms to invest in general training. When workers' outside option grows more slowly than their productivity, resulting in wage compression, firms can capture some of the benefits of training and will have sufficient incentives to invest. Acemoglu and Pischke [1999a] suggest sources of wage compression that could lead to firms investing in general training. The same authors show how under certain circumstances, an increase in minimum wages can actually increase training—a policy which under Becker's model reduces a worker's ability to accept lower wages and leads to lower levels of training [Acemoglu and Pischke, 1999b]. Using CPS training supplement data from periods in the US when minimum wage laws changed, they find evidence that levels of training increased when the minimum wage increased. Evidence inconsistent with Becker's theory of on-the-job training.

Another possible source of wage compression is adverse selection [Acemoglu, 1998]. Firms train their workers to learn about their ability. They make good offers to high ability workers, causing low ability workers to quit. The mean ability among non-workers is lower than the whole labour force, lowering the outside option of high-ability workers and allowing firms to pay them less than their marginal product. This leads to wage compression and incentives for firms to pay for general training. An interesting feature of the Acemoglu [1998] model is the presence of multiple equilibria, with different levels of training and worker mobility. The authors suggest that the low-mobility, high-training equilibrium represents the situation in European countries, while the high-mobility, low-training equilibrium reflects what is seen in the US. Autor [2001] models the temporary help industry in the US, and finds evidence that adverse selection effects can lead firms to pay for general training—even for temporary workers.

Acemoglu [1997] discusses how search and match frictions such as costs to workers changing jobs, or to firms hiring new workers can lead to match-specific rents to be bargained over, another source of wage compression. Collective bargaining, such as that seen in unionised industries, can also lead to wage compression. Dustmann and Schönberg [2012] extend the model of Acemoglu [1998] to include the ability of firms to commit to providing training. They suggest that the complexity of firm-provided training makes it difficult to verify and build a model which suggests that more training will be provided when a firm can reliably commit to training its workers. They provide evidence of differences in the level of training provision between countries such as Germany and the United Kingdom, and attribute the difference to the greater regulation of training in Germany which increases a firm's ability to commit.

The theory and evidence suggest that given sufficient labour market frictions, both firms and workers have incentives to invest in general training. The question is then: who will ultimately invest, and will the level of investment be socially optimal? Acemoglu and Pischke [1999a] argue that even when workers are not credit constrained, in a non-cooperative regime training will be under provided. Investment in training is never shared in their model; whichever party has the higher demand for training will invest up to that level, and the other party does not. There are two externalities, however, which the investor does not take into account and which ultimately lead to under investment. Neither the firm or the worker considers the benefit of training to the other party—or the effect on marginal cost when the other party invests. Second, other firms outside of the worker-firm training relationship are also likely to benefit from training given that frictions will allow them to capture some of the returns. Perhaps there is a role for government subsidies and regulation to achieve an optimal level of training.

1.1.2 Vocational training in France

This is the belief in France, where vocational training has long been valued by the state. While France does not have as impressive an apprenticeship system as Germany, there exists a plethora of non-academic qualifications available to young people as an alternative to the traditional *Baccalauréat-Licence-Master*. A professional baccalauréat was introduced in 1986, and has grown steadily in popularity—from 2.8% of those leaving school in 1990, to 24.2% in 2014 [Direction de l'évaluation, 2016]—growth not at the expense of the technological or general baccalauréat. This growth in post-sixteen formal education among the French labour force is likely to impact the propensity to train of French workers.

The importance placed on vocational training by the French state extends to the workplace. In 1971 French companies with more than 10 employees became obliged by law to spend an amount equivalent to 0.8% of their wage bills on training [Verdier, 1994]. The level is currently set at 0.55% for firms with under 10 employees and 1% for all other firms.[4] [LegiFrance, 2014]. The system behind this requirement is rather complex, and very particular to France. There exist a number of "joint collecting bodies" (OPCAs), and a firm can pay the their contribution directly to the OPCAs, who report this to the state. The OPCAs then provide training directly or the firm who made the initial contribution can request the funds back for a specific training. If they do not receive any requests, the OPCAs can use the contribution to fund training at other firms.

It is not left completely up to firms to decide which training to fund. Employees are entitled to "individual training leave" (CIF)—funded by a ringfenced portion of their firm's training contribution. Employees then apply formally for training to their companies and their requests must be approved—companies do not have the right to postpone requested training for more than 9 months, and must reply to formal requests within 30 days. Employees are entitled to between 80% and 100% of their salary during training under

[4] These levels changed and a "personal training account" was introduced, coming into effect in January 2015. However, the majority of training recorded in the DEFIS data took place before this date.

CIF, and must be reinstated in their role on their return. Therefore, in France there is a direct mechanism allowing employees to choose training solely for their benefit, funded in part by their employer, who is obliged accept the request. Though there are likely to be many issues complicating this system in reality, in principal training under the CIF provides an opportunity to study the decision of individuals to train, abstracting from many of the issues faced by unregulated firm-employee dynamics discussed in section 1.1.1.

1.2 Empirical evidence

Early studies used experience and tenure to proxy for training, ignoring somewhat the distinction between specific and general training, and also between formal and informal training and "learning-by-doing". As more detailed microdata became available, researchers focus switched to estimating the causal effect of training, via OLS and difference-in-difference specifications, and increasingly parametric and nonparametric selection models [Leuven, 2004]. The measured effects are not always comparable: authors use different measures of training, from hours completed [Lynch, 1992] to careers of training [Parent, 2003]; different methodologies estimating different underlying effects; and focus on different populations (e.g. the unemployed, young men). Table 1 summarises estimated returns to training incidence across a number of countries. The OLS estimates vary between 5% and 17%, though are generally less than 10%. Fixed-effects and difference-in-differences estimate are slightly lower, falling between 0% and 5%. Estimates using explicit selection models vary greatly: from −5.7% in France [Goux and Maurin, 2000] to 34.2% in the UK [Arulampalam and Booth, 2001]. This variation might reflect misspecification of either the errors—Goux and Maurin [2000] assume jointly normal errors, for example—or of the underlying selection model. In many countries workers are not free to choose to train; the possibilities are intrinsic to their job and the wishes of their employer. Workers in France enjoy this freedom.

The Goux and Maurin [2000] study uses data similar in structure to the DEFIS data used

Table 1: Summary of estimated returns to on-the-job training

	Period	Method	Estimate	Notes
US				
Loewenstein and Spletzer [1998]	1988-91	FE	0.035	
UK				
Booth [1991]	1987	OLS	0.106	male
		OLS	0.166	female
Blundell et al. [1996]	1981-91	q-DiD	0.036	m, 33, on job
			0.066	m, 33, off job
			0.003	f, 33, on job
			0.046	f, 33, off job
Arulampalam and Booth [2001]	1981	SM	0.342	33
France				
Goux and Maurin [2000]	1988-93	SM	−0.057	participation
		OLS	0.071	
Germany				
Kuckulenz and Zwick [2003]	1998-99	SM	0.15	
Norway				
Schøne [2004]	1989-93	OLS	0.053	
		FE	0.011	

here, and was collected by the same agency, Ceréq, in 1993. Their data covers the period between 1988 and 1993 and details training in that period, any changes in employers, and wages. The authors exploit the data using a system of simultaneous equations to model the selection of workers into training, the decision to move to another firm, and wages in 1992. Their preferred specification finds a negative (see Table 1), though not significant, effect of training on wages. Goux and Marin suggest that this may be an artefact of the incentives in the French system; French firms will lose money if they do not appear to be training their workers. However, firms only need provide "cosmetic training scheme[s] that have little impact on productivity" [Goux and Maurin, 2000, p. 15]. As the authors go on to point out, it is hard to believe that firms would go through the motions of investing in and providing training to no benefit if they could just pay a fee. Similarly, it would be quite a scandal if enough firms to produce such results are reporting training their workers when they are in fact not training them at all. Their results perhaps provide evidence that it is firms who capture most of the benefit from increases in productivity due to training. This hypothesis is consistent with evidence from the UK that workers only receive around 50% of their increase in productivity following training [Dearden et al., 2006]. Currently DEFIS data lacks detailed data on firm performance that would enable deeper analysis of how firms and workers share the returns to training; the analysis in this paper focuses solely on measuring the wage returns to vocational training in France.

2 Data

The dataset used in this paper is the first stage of DEFIS, a novel survey conducted by the Céreq), with the specific aim to better understand the training environment in French firms. The survey comprises two components: one focusing on firms and one on their employees. Céreq plans to survey the employees annually over 5 years to build longitudinal data on their training and career paths, a process which began in Autumn 2015. Firms completed the survey in December 2015. Although currently only the first "wave" of the employee surveys and the firm survey are available, this still contains sufficient information for a first attempt at estimating the wage returns to vocational training in France. The following sections present key descriptive statistics of the DEFIS data; the analysis of firm data is from a note by Céreq [Dubois et al., 2016], while the analysis of employee data is my own.

2.1 Firms

The survey includes a total of 4,529 firms in France. The sample of firms with 10 or more employees covers all commercial sectors (except agriculture) and 66% of sectors (and 72% of employees) in smaller firms (3-9 employees).[5] Summary statistics by firm size and sector are presented in Tables 7 and 8. The percentage of *training firms*[6] is over 94% for all firms with over 20 employees and is increasing in firm size (Table 7). Training firms comprise at least 75% of the sampled firms in each sector (Table 8). The percentage of employees who train per firm[7] is also increasing in firm-size, and consistent across sectors. The focus of training effort is homogeneous across firms by size and by sector, with approximately 50% reporting no preference in which category of employee they train. The exceptions are the construction and transport sectors, which predominantly focus their training effort on workers.

Around half the surveyed firms report keeping their expenditure on training constant in recent years, with only 4% reporting a fall in expenditure. The amount spent on training (as a percentage of total expenditure on salaries) is increasing in firm size, consistent across sectors.Firms were also asked about their main reasons for training: training to meet regulations was most popular among small firms, while training to support changes is the most popular reason among larger firms. The second most popular reason for *not training* among firms with less than 50 employees was that they preferred training their

[5] This discrepancy in coverage appears to be driven by differences in the sector classification for small and large firms.

[6] Firms who have at least one employee who has completed training in 2014.

[7] Employees who received training in 2014.

employees "in the course of work". This suggests that the difference in training between large and small firms is not as stark as it first appears, but perhaps is driven in part by different views on what is meant by "training". This hypothesis is supported by evidence on the main mode of skill acquisition: results across firm sizes are nearly identical with "in the performance of work" most popular at approximately 50% followed by "upon hiring" and "during training organised by the company" at around 20%.

Firms were also asked about the "information and dialogue around training". Both the importance placed on training in supervisor-worker discussions, and the likelihood that a company disseminates information on training are increasing in firm size. These are features to bear in mind when reading the results of the IV estimation later in the paper. There are certainly a significant number of employees in firms who do not disseminate information on training, and it appears important to control for firm size in the IV specification.

2.2 Employees

The second component of DEFIS is the most important for my analysis: the data on employees. The data includes detailed information on 16,129 individuals who were employed in one of the surveyed companies in December 2013. Céreq have "enriched" the anonymised data with information from the DADS employee database—currently for 2013 and 2014—including wage and hours data for all employees.

The plots in Figure 2 in the Appendix present descriptive analyses of the sample of employees. The age of employees in the sample are well distributed. Over 27% of individuals hold a CAP/BEP or equivalent as their highest level of education, followed by BAC+2 at 17%. The share of other levels of education varies between 4 and 10%, without a clear preference for higher or lower levels. The sample is predominantly male, at over 67%; this is unsurprising for a sample of mostly *manual workers* in private industry, typically male-dominated roles. Over 75% of the sample remained at the same firm. Over

Table 2: Summary statistics on the employee component of DEFIS

(a) Gender and training

	Share
Female	0.31
Training	0.75
pre-2014	0.64
post-2014	0.28
last year	0.07

(b) Wages and training episodes

	Mean	Median	Std dev.
Wage (EUR)	15.04	12.20	11.05
Training episodes	1.56	1.00	0.89

Notes: Data on the number of training episodes was recorded for post-2014. These are for individuals with at least one training episode.

84% of those surveyed held a permanent contract (CDI) in 2013. The next most common contract is fixed term (CDD) with a share under 7%. The majority of individuals were working full-time in 2013, and the sample is dominated by "qualified workers", "engineers", and "category C, D or equivalents",[8] each contributing nearly a quarter of the sample. Labourers, supervisors and technicians make up around 7% each. The sampling ensures that the share of employees across sectors matches the share of firms.

Wages are from the DADS database, and was matched by Céreq to the sample. The raw information from DADS is data on total yearly earnings and total yearly hours worked in 2013 and 2014. To calculate an average hourly wage for each year I divided total earnings by total hours. Individuals whose resulting hourly earnings are over 200 euros are excluded, along with any individuals reporting zero earnings. This corresponds to 16,064 employees in 2013, and 13,555 in 2014.[9] Figure 3a shows the distribution of wages in the sample for 2013 and 2014, for all individuals earning less than 50 euros per hour. The distribution does not change much between 2013 and 2014 and offers no surprises, with a median wage of 12.6 euros and a mean of 15.5 euros.

The dataset includes detailed information on the training undertaken by the surveyed employees, initially covering three periods:

[8] Category C or D is equivalent to clerical, service, caregivers and babysitters.

[9] The fall in observations is driven mainly by workers employed in a surveyed company in 2013 whose hours are missing from the data in 2014—though it includes their total net earnings. Unfortunately it is not possible to determine why their hours are missing. Table 9 in the Appendix summarises the differences. The majority appear to have been employed in construction in 2013.

◇ from when the individual left education until December 2013

◇ after December 2013

◇ in the year preceding the survey in 2015.

The second and third of these periods overlap and the questions asked on training are not consistent across periods.[10] To avoid issues equating different types of training recorded differently I focus on the *incidence of training*.

Much of the analysis aggregates these three periods, representing training with an indicator variable taking the value 1 if the individual received training before that date. The disadvantages of such an approach are discussed in more detail in later sections, but it allows a crude split the sample into two groups: those who have received vocational training at some point in their career and those who have not. Just over three-quarters of individuals fall into this first category, suggesting training among the sampled population is widespread. In the Appendix are a number of plots showing how wages vary with age, education and training status (Figures 3c and 3b). I recreate Mincer's age-wage profiles using the DEFIS data, and similar profiles when data is grouped by training rather than schooling (Figure 3b). Mean wages are higher for those who have trained even among individuals with the same level of formal education.

3 Econometric model

There is a long tradition of modelling investment in human capital decisions, both in terms of formal education and vocational training. The majority of this work (particularly the empirical work) focuses on the pecuniary benefits of human capital accumulation—i.e. wages. Why wages are an attractive measure may be obvious—and linked to the desire to attach a price to everything: money is unambiguous. Everybody prefers more money.

[10] This is due to this first round of surveys collecting contextual data as well as the first wave of longitudinal data.

While it is possible to measure outcomes in terms of responsibility, hours or even job status, it is not clear that everyone prefers more responsibility to less, or would want to work more hours given the chance. Focusing on wages permits use of the Mincerian wage equation, which models wages as a log-linear function of the explanatory variables [Mincer, 1958]. An alternative is the nonparametric approach, following the treatment effects literature, which formulates everything in terms of conditional expected outcomes. Focusing on wages makes equal sense under this approach; it is valuable to have an outcome that can be compared unambiguously across individuals. I combine these approaches, motivating a treatment effects methodology with a choice-theoretic underlying model following Heckman and Vytlacil [2001] and Angrist [2001].

A common assumption in the literature on training that attempts to model selection is that the decision to train is one taken solely by the employee, rather than jointly with (or solely by) their firm. This does not seem too strong of an assumption for France in comparison with other countries given the unique features of the French training system—in particular the CIF and DIF. I then take the incidence of training to be a single treatment[11] with heterogeneous costs and effects—analogous to how Carneiro et al. [2011] treat college education. This approach is useful as it allows the use of results from the literature on heterogeneous treatment effects, in particular Imbens and Angrist [1994] on identifying local average treatment effects (LATE) and Heckman and his coauthors work on marginal treatment effects [Heckman and Vytlacil, 2001, Heckman and Navarro-Lozano, 2004, Carneiro et al., 2011]. It also reflects the position of the French state; rather than restricting attention on certain types of training, they allow workers to choose the training they undertake, and intervene only to ensure the firm allows the worker his choice, and, at least nominally, funds this choice.

[11] This is not strictly true, of course. However, the distinction between homogeneous treatment with heterogeneous effects and heterogeneous treatment is not clear cut, especially in the social sciences.

3.1 A latent variable approach

I assume the following underlying selection process governing individual's selection into training. Specifically, I apply Heckman and Vytlacil [2005]'s Roy model to the context of vocational training.

Potential outcomes (assuming a Mincerian log-linear wage equation)[12] can be represented as:[13]

$$\ln w_1 = \beta_1 \boldsymbol{X} + U_1 \quad \text{and} \quad \ln w_0 = \beta_0 \boldsymbol{X} + U_0 \tag{1}$$

where $\mathbb{E}\left[\ln w_1 \mid \boldsymbol{X}\right] = \beta_1 \boldsymbol{X}$ and $\mathbb{E}\left[\ln w_0 \mid \boldsymbol{X}\right] = \beta_0 \boldsymbol{X}$, i.e. $\mathbb{E}[U_1 \mid \boldsymbol{X}] = \mathbb{E}[U_0 \mid \boldsymbol{X}] = 0$. This is sometimes called in the literature the *conditional independence assumption* (CIA), and is key; it states that unobserved outcomes are uncorrelated with treatment once we condition on $\boldsymbol{X}$. The validity of the different experimental designs in this paper rests on variations of the CIA.

An employee's choice to train is determined by the latent variable, T^* which is linear in parameters

$$T^* = \beta_T \boldsymbol{Z} + U_T \tag{2}$$

and

$$T = \begin{cases} 1 & \text{if } T^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

In equations (1) and (2), $\boldsymbol{X} \subset \boldsymbol{Z}$[14] are observable characteristics and $U_.$ terms are

[12] This is the standard in the literature on training. It also greatly simplifies the analysis. I discuss the realism of this assumption in section 6.

[13] I denote random variables by upper case letter, and their (potential) realisations by the corresponding lower case.

[14] It will be necessary at times to assume that there are variables in $\boldsymbol{Z}$ but not in $\boldsymbol{X}$; instrument(s) which explain T^* but not wages directly.

unobserved determinants. For example, an alternative formulation of equation (2) is

$$T^* = w_1 - w_0 + \mu(\boldsymbol{Z}) + U_1 - U_0 - U_C \tag{4}$$

In (4), $\mu(\boldsymbol{Z})$ represents the observed non-wage costs/benefits associated with training, and U_C the unobserved.

An observed outcome w_i satisfies

$$\ln w_i = T_i \ln w_{1i} + (1 - T_i) \ln w_{0i} \tag{5}$$

The effect of training is $\Delta \equiv \ln w_1 - \ln w_0$.

Other useful—and hopefully estimable—*treatment effects* can be defined using this notation.

◇ the *average effect of training conditional on* $\boldsymbol{X} = \boldsymbol{x}$

$$\Delta^{ATE}(\boldsymbol{x}) \equiv \mathbb{E}[\Delta \mid \boldsymbol{X} = \boldsymbol{x}] = (\beta_1 - \beta_0)\boldsymbol{x}$$

◇ the *average effect of training on the trained* given $\boldsymbol{X} = \boldsymbol{x}$

$$\Delta^{ATT}(\boldsymbol{x}) \equiv \mathbb{E}[\Delta \mid \boldsymbol{X} = \boldsymbol{x}, T = 1] = (\beta_1 - \beta_0)\boldsymbol{x} + \mathbb{E}[U_1 - U_0 \mid T = 1]$$

◇ the *local average treatment effect* given $\boldsymbol{X} = \boldsymbol{x}$ for a change in $\boldsymbol{Z}$ from $\boldsymbol{z}$ to $\boldsymbol{z'}$

$$\Delta^{LATE}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{z'}) \equiv \mathbb{E}[\Delta \mid \boldsymbol{X} = \boldsymbol{x}, T(\boldsymbol{z}) = 1, T(\boldsymbol{z'}) = 0]$$

◇ the *marginal effect of training* given $\boldsymbol{X} = \boldsymbol{x}$ and $U_T = u_T$

$$\Delta^{MTE}(\boldsymbol{x}, u_T) \equiv \mathbb{E}[\Delta \mid \boldsymbol{X} = \boldsymbol{x}, U_T = u_T]$$

Heckman and Vytlacil [2005] demonstrate how Δ^{ATE}, Δ^{ATT}, Δ^{LATE} along with other traditionally estimated *treatment effects* can be expressed as weighted sums of Δ^{MTE} over the distribution of unobserved factors affecting selection, U_T, shown by Heckman and Vytlacil [1999]:

$$\Delta^{TE}(x) = \int_0^1 \Delta^{MTE}(\boldsymbol{x}, u_T) h_{TE}(\boldsymbol{x}, u_T) \mathrm{d}u_T \tag{6}$$

The weights for a given treatment effect are:

$$h_{ATE}(\boldsymbol{x}, u_T) = 1 \tag{7}$$

$$h_{OLS}(\boldsymbol{x}, u_T) = \begin{cases} 1 + \Omega & \text{if } \Delta^{\text{MTE}}(\boldsymbol{x}, u_T) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where

$$\Omega = \frac{\mathbb{E}\left[U_1 \mid \boldsymbol{X} = \boldsymbol{x}, U_T = u_T\right] h_1(\boldsymbol{x}, u_T) - \mathbb{E}\left[U_0 \mid \boldsymbol{X} = \boldsymbol{x}, U_T = u_T\right] h_0(\boldsymbol{x}, u_T)}{\Delta^{\text{MTE}}(\boldsymbol{x}, u_T)}$$

and

$$\begin{aligned} h_1(\boldsymbol{x}, u_T) &= \frac{1}{\mathbb{E}[P \mid \boldsymbol{X} = \boldsymbol{x}]} \int_{u_T}^1 f(p \mid \boldsymbol{X} = \boldsymbol{x}) \mathrm{d}p \\ h_0(\boldsymbol{x}, u_T) &= \frac{1}{\mathbb{E}[1 - P \mid \boldsymbol{X} = \boldsymbol{x}]} \int_0^{u_T} f(p \mid \boldsymbol{X} = \boldsymbol{x}) \mathrm{d}p \end{aligned}$$

where $f(\cdot \mid \boldsymbol{X} = \boldsymbol{x})$ is the pdf of u_T conditional on $\boldsymbol{x}$, and $P = \Pr(T = 1)$. The local average treatment effect (LATE), discussed by Imbens and Angrist [1994], is defined for a range of u_T values, corresponding to *marginal individuals* at either value of Z (when Z is discrete), which describe the location of the LATE:

$$\Delta^{LATE}(\boldsymbol{x}, u_T, u_T') = \frac{1}{u_T - u_T'} \int_{u_T'}^{u_T} \Delta^{MTE}(\boldsymbol{x}, u) \mathrm{d}u \tag{9}$$

Heckman and Vytlacil [2005] show that $\Delta^{MTE}(\boldsymbol{x}, u_T)$ is identified for certain values of

u_T corresponding to suitable instruments. Their work provides a useful framework in which to compare the results of the three estimation methodologies employed here—OLS, differences-in-differences, and IV—and to discuss the assumptions under which each is informative. Note that these effects are all currently defined *conditional on* $\boldsymbol{X} = \boldsymbol{x}$. An assumption that greatly simplifies interpretation of the estimates in the following sections is that these treatment effects are *independent of* $\boldsymbol{X}$. Under this assumption equation (1) becomes

$$\ln w_1 = \beta \boldsymbol{X} + \Delta + U_1 \text{ and } \ln w_0 = \beta \boldsymbol{X} + U_0 \tag{10}$$

where $\mathbb{E}[\Delta \mid \boldsymbol{X}] = \mathbb{E}\Delta = \Delta^{ATE}$. I proceed under this stronger assumption and discuss the implications if it does not hold. To make the notation clear; under the stronger assumption I write Δ^{TE}, while $\Delta^{TE}(\boldsymbol{x})$ refers to cases when the assumptions permit correlation between $\boldsymbol{X}$ and the treatment effect.

4 Pooled OLS

A natural starting point when estimating the returns to vocational training for workers is the classic Mincerian wage equation. This approach has been widely used to estimate the returns to schooling, and following Mincer [1958], models wages as log-linear in the explanatory variables.

$$\ln w_i = \beta \boldsymbol{x}_i + \delta_p T_i + u_i \tag{11}$$

$\boldsymbol{x}_i$ is a vector of observable characteristics that might explain wages, T_i is an indicator which takes the value 1 if an individual has received training in their career, and u_i contains unobserved factors affecting wages. Under equation (10), the OLS estimate of δ_p is a natural estimator of Δ^{ATE} under certain additional assumptions

$$\delta_p = \mathbb{E}[\ln w \mid \boldsymbol{X} = \boldsymbol{x}, T = 1] - \mathbb{E}[\ln w \mid \boldsymbol{X} = \boldsymbol{x}, T = 0] \tag{12}$$

$$= \Delta^{ATE} + \{\mathbb{E}[U_1 \mid \boldsymbol{x}, T = 1] - \mathbb{E}[U_0 \mid \boldsymbol{x}, T = 0]\} \tag{13}$$

Table 3: OLS results on pooled data

	Dependent variable: log(wage)					
Model:	NC	FFE	IC	FC	ICFC	ICFFE
Training	0.25^{***}	0.15^{***}	0.04^{***}	0.21^{***}	0.03^{***}	0.02^{***}
	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
R^2	0.06	0.49	0.58	0.15	0.60	0.75
Adj. R^2	0.06	0.42	0.58	0.15	0.60	0.71
Num. obs.	29,619	29,619	29,619	29,619	29,619	29,619
RMSE	0.45		0.30	0.43	0.30	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Notes: Model NC regresses ln(wage) on training. FFE adds firm fixed effects. IC adds controls for education, contract type, socio-professional category, role, FT/PT, gender, age and graduation year. ICFC adds controls for sector and firm-size. ICFFE drops adds firm fixed effects to IC. All controls (including firm fixed effects) refer to 2013. Regressions without fixed effects were implemented in `R` using command `lm()`. Regressions including fixed effects were implemented in `R` using command `felm()`.

Recall the CIA, $\mathbb{E}[U_1 \mid \boldsymbol{x}] = \mathbb{E}[U_0 \mid \boldsymbol{x}] = 0$. Then δ_p is a consistent estimator of Δ^{ATE} if $(U_1, U_0) \perp T$ conditional on $\boldsymbol{X}$; if the term in braces in (13) is zero. I discuss the plausibility of this assumption in the context of French training in section 4.2.

4.1 Pooled-OLS results

Equation (11) is straightforward to estimate by OLS, and the key results of this estimation are in Table 3.[15] The OLS estimates report a positive correlation between training and wages. The decreasing magnitude of the estimate across the table demonstrates the importance of the added controls. Model NC simply regresses the logarithm of the wage against a dummy indicating the person received training *before or in the year to which the wage refers.* The estimate of 0.25 suggests that the mean of trained worker wages in the sample is 25% higher than the mean wage of their untrained colleagues. This does not represent a causal effect; trained workers may be more likely to hold senior positions, or work in industries with higher wages, leading to omitted variable bias. Model FFE includes firm fixed effects, to control for the effects of industry and other firm-specific factors, though not (explicitly) individual characteristics, on wages. The inclusion of firm

[15] The full results are in the Appendix, Table 10. The controls were chosen to match the choices of researchers in similar studies (for example Goux and Maurin [2000]) restricted by the availability of variables in the DEFIS data.

fixed effects reduces the estimate of δ_p, suggesting firm-specific factors are important in determining wages—firm-fixed effects might also proxy for individual characteristics if certain types of worker select into certain types of firms. Model IC controls for individual characteristics and FC controls only for sector and firm-size. The estimated δ_p in IC is much smaller in magnitude, suggesting there are numerous important variables missing from FC included in IC. ICFC combines the controls from IC and FC with the slightly smaller estimate suggesting that firm size and sector are important beyond acting as proxies for individual characteristics. Finally ICFFE includes individual characteristics as controls and firm fixed effects. At first glance, this would appear to be the preferred specification. However, descriptive analysis of the DEFIS data shows that the vast majority of firms have fewer than 10 sampled workers, and more two-thirds have less than five. Therefore, a significant number of cells will be empty when including both individual controls and firm fixed effects.

I am hesitant to mention p-values as many authors have expressed doubt over their validity. For example Gelman [2016] argues that the standard errors used in statistical packages to calculate p-values are incorrect as they do not account for the "researcher degrees of freedom" implicit in the choice of variables, and the data collection. However, given the size of the DEFIS sample, and the uncontroversial variables included as controls, I feel they are useful at least to facilitate comparison with previous work. All estimates have p-values less than 0.001, despite their small magnitude. Once individual controls are included, the OLS estimates are slightly below the range of OLS estimates from other authors. This is likely to be caused by the inclusion of variables that might themselves be affected by past training as controls. For example, Goux and Maurin [2000] were careful to only include post-training controls that they argued would be unaffected by training. I discuss this issue in greater depth in the next section.

Measurement error is also likely to be an issue. Employees in the survey are asked if they have undertaken any training since they left school. This may introduce errors due to what different employees class as training and also, given the retrospective nature of the question, due to difficulties in remembering such details over a long career. The diff-in-diff specification solves the memory issue to some degree, and the questions surrounding training in this period leave less up to the interpretation of the employee.

4.2 Discussion of pooled-OLS assumptions

There are a number of reasons to be cautious of claiming to have identified a causal effect based on these results. The first relates to the timing of events. Ideally, only controls that were determined *before* the training took place would be included in each regression. To illustrate consider role in the firm. Suppose an individual was given a role with more responsibility as a direct result of training, with an associated higher salary. At least part of this increase in salary is directly attributable to training—as without training the person would not have been promoted and so would not be on the same salary. Controlling for the after-training role risks excluding this effect. Unfortunately the timing of the observations in the DEFIS data makes controlling for variables affected by training before training impossible in this specification. Leaving them out is also not an option: as many of the controls are certainly correlated with wages, their exclusion would invalidate the CIA (as demonstrated by the results in table 3). The difference-in-differences design in the next section solves this issue.

A separate issue is that individuals are likely to select into training based on *unobserved* characteristics, such as ability, which are captured in (2) by U_T. If $w_1, w_0 \perp U_T$, i.e. these unobserved factors affecting training do not otherwise affect wages, then the last term (in braces) of equation (13) is zero and the pooled specification in equation (11) consistently estimates Δ^{ATE} [Heckman and Vytlacil, 2005]. However, when $w_1, w_0 \not\perp U_T$, the resulting troublesome selection on unobserved characteristics falls into two categories: (i) individuals selecting to be trained based on unobserved characteristics that *separately* impact their wages; (ii) individuals select into training based on their *heterogeneous* returns to training. This is an important distinction.

Take ability as an example. Assuming (i), one might argue that ability affects the costs associated with training, so that the disutility of training is lower for higher ability individuals. Ability is also likely to affect their wages directly, regardless of training. Crucially, however, under (i) ability does not affect the impact of training on wages.[16] Also, assuming (i), the treatment effects defined in 3.1 coincide:

$$\Delta^{ATE} = \Delta^{ATT} = \Delta^{LATE} = \Delta^{MTE}.$$

Therefore a *uniquely defined treatment effect* and can be consistently estimated via IV and difference-in-differences designs. The key assumption is that although the effects of training may be heterogeneous, the *variation in effects is not correlated with* U_T.

Conversely, under condition (ii) individuals select into training based on their ability, precisely because the *wage returns to training are correlated with ability.* This is the situation described in Heckman and Vytlacil [2005]. Using their framework it is straightforward to see the treatment effects all coincide if Δ^{MTE} is constant in u_T, as in (i), and also why OLS is a consistent estimator of Δ_{ATE} when $(U_1, U_0) \perp T$, as the OLS weights become unity. Under condition (ii), however, this is not the case. The second term in equation (8) is not necessarily zero, and the OLS weights may be negative. The effect estimated by OLS does not have a natural interpretation. If an instrument is available that affects the propensity to train, but not wages directly, it is possible to estimate a local average treatment effect as demonstrated in section 6.

The final issue concerns the assumption that the treatment effect is independent of $\boldsymbol{X}$. Without this assumption, Δ^{ATE} is only defined conditional on $\boldsymbol{X}$. The OLS estimate of δ_p is then a weighted average of the $\Delta^{ATE}(\boldsymbol{x})$, with the weights in the current specification not clearly defined. An alternative specification that has a more useful interpretation is suggested by Angrist and Pischke [2008]. Their specification is "saturated-in-$\boldsymbol{X}_i$" as it

[16] Assumption (i) also allows the effects of training on wages to vary with ability, as long as individuals do not take this into account when deciding whether to train.

includes a dummy for *every possible combination of controls*

$$\ln w_i = \sum_{\boldsymbol{x}} d_{i\boldsymbol{x}} \beta_x + \delta_R D_i + \varepsilon_i \tag{14}$$

where $d_{i\boldsymbol{x}}$ takes the value 1 when $\boldsymbol{X}_i = \boldsymbol{x}$ and 0 otherwise. δ_R is then defined as the following weighted average of $\Delta^{ATE}(\boldsymbol{x})$

$$\delta_R = \frac{\sum_{\boldsymbol{x}} \Delta^{ATE}(\boldsymbol{x}) \left[\Pr(T_i = 1 \mid \boldsymbol{X}_i = \boldsymbol{x})(1 - \Pr(T_i = 1 \mid \boldsymbol{X}_i = \boldsymbol{x}))\right] \Pr(\boldsymbol{X}_i = \boldsymbol{x})}{\sum_{\boldsymbol{x}} \left[\Pr(T_i = 1 \mid \boldsymbol{X}_i = \boldsymbol{x})(1 - \Pr(T_i = 1 \mid \boldsymbol{X}_i = \boldsymbol{x}))\right] \Pr(\boldsymbol{X}_i = \boldsymbol{x})} \tag{15}$$

In words, δ_R puts most the most weight on $\Delta^{ATE}(\boldsymbol{x})$ for values of $\boldsymbol{x}$ with the highest variance in T_i—individuals with values of $\boldsymbol{x}$ that do not make them particularly likely or unlikely to train. Due to the computational difficulties associated with including a separate dummy for every possible value of $\boldsymbol{X}_i$, the results of such a regression are not presented here. However, this is certainly worth exploring.

5 Difference-in-differences

As discussed in section 4.2, in the case where (i) individuals select into training based on their ability, (ii) ability impacts directly on wages, but (iii) ability does not impact the effects of training, differences-in-differences (DiD) is a valid approach. As the DEFIS data is a true panel, with wages for each individual in 2013 and in 2014, it also permits inclusion of individual fixed effects.

The DiD estimator can be obtained by estimating a variation on equation (11):

$$\ln w_{i,t} = \alpha_t + \beta \boldsymbol{x}_{i,t} + \delta_{DiD} T_{i,2014} + u_{i,t} \tag{16}$$
$$i = 1, 2, \ldots, N, \quad t = 2013, 2014$$

where $T_{i,2014}$ is a dummy variable indicating whether i received training in 2014. Note that $\boldsymbol{x}_{i,t}$ is a vector of time-varying *and* time-invariant individual (and firm) characteristics.

There are a number of possibilities contained in this design. My preferred specification is to include all individuals, with pre-2014 training in $\boldsymbol{x}$ as a control.

Returning to the notation introduced in section 3, δ_{DiD} is

$$\begin{aligned}\delta_{DiD} = \mathbb{E}[\ln w \mid \boldsymbol{x}, T_{2014} = 1, t = 2014] - \mathbb{E}[\ln w \mid \boldsymbol{x}, T_{2014} = 1, t = 2013] \\ - \{\mathbb{E}[\ln w \mid \boldsymbol{x}, T_{2014} = 0, t = 2014] - \mathbb{E}[\ln w \mid \boldsymbol{x}, T_{2014} = 0, t = 2013]\}. \end{aligned} \tag{17}$$

Using the definitions for $\ln w_1$ and $\ln w_0$ in (10) δ_{DiD} rewrites as

$$\begin{aligned}\delta_{DiD} = \Delta^{ATE} + \mathbb{E}[U_1 \mid \boldsymbol{x}, T_{2014} = 1, t = 2014] - \mathbb{E}[U_0 \mid \boldsymbol{x}, T_{2014} = 1, t = 2013] \\ - \{\mathbb{E}[U_0 \mid \boldsymbol{x}, T_{2014} = 0, t = 2014] - \mathbb{E}[U_0 \mid \boldsymbol{x}, T_{2014} = 0, t = 2013]\} \end{aligned} \tag{18}$$

Recall assumptions (i)-(iii) made at the start of this section; individuals select into training based on unobserved factors that affect their wages but that do not affect the impact of training on wages. Under these assumptions one can decompose U as follows:

$$U_1 = \varepsilon_i + \varepsilon_t + V_1 \text{ and } U_0 = \varepsilon_i + \varepsilon_t + V_0 \tag{19}$$

where ε_i represents time-invariant individual factors that affect wages (and possibly selection into training), while ε_t represents time-varying factors common to all individuals. If $\mathbb{E}[V_0 \mid \boldsymbol{x}, T, t] = \mathbb{E}[V_0 \mid \boldsymbol{x}, t]$, the estimator $\hat{\delta}_{DiD}$ is a consistent estimator of Δ^{ATT}. Under the (stronger) assumption that $\mathbb{E}[V_1 \mid \boldsymbol{x}, T, t] = \mathbb{E}[V_1 \mid \boldsymbol{x}, t]$, $\hat{\delta}_{DiD}$ is a consistent estimator of Δ^{ATE}.

5.1 DiD results

Table 4 includes the estimated coefficient on T_{2014} and pre-2014 training for variations on (16) with different controls. Model NC includes only dummy variables for training in 2014 and pre-2014, and the year in which the wage was observed; equivalent to comparing the

Table 4: DiD wage regressions

	Dependent variable: log(wage)					
Model:	NC	IC	FC	IFC	IFE	IFE(SF)
Training	0.12***	0.04***	0.09***	0.02***	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Prev. trained	0.24***	0.03***	0.20***	0.02***		
	(0.01)	(0.00)	(0.01)	(0.00)		
R^2	0.07	0.58	0.16	0.60	0.03	0.02
Adj. R^2	0.07	0.58	0.16	0.60	−1.12	−1.08
Num. obs.	29,619	29,619	29,619	29,619	29,619	23,671
RMSE	0.45	0.30	0.43	0.29		

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Model NC regresses $\ln(w)$ on past and 2014 training with year dummies. IC includes controls for contract type, education, socio-professional category, role, FT, gender, age and graduation year. FC includes controls for firm size and sector. ICFC combines all controls from IC and FC. IFE includes individual and year fixed effects and controls for education, sector, FT/PT, age, and graduation year. IFE(SF) uses the same specification as IFE but only including individuals who remained at the same firm in 2014. Regressions without fixed effects were implemented in `R` using command `lm()`. Regressions including fixed effects were implemented in `R` using command `felm()`.

change in wages from 2013-2014 for those who trained in 2014 and those who did not, without controlling for different trends due to selection.

The validity of model NC rests on the strong assumption that trends in wages do not depend on individual or firm characteristics. From Mincer [1958] age-earnings profiles, we know that wages increase faster at the beginning of careers—and should therefore expect wage trajectories to vary with age and experience. These wage-earnings profiles hold for cross-sections of individuals in the data with and without prior training (see Appendix A, Figure 3c). This makes conditioning on age and experience necessary—otherwise the measured effect could be capturing age and experience differences between those who train in 2014 and those who do not, rather than the effects of training. The decrease in magnitude of the measured effect from model NC to IC/IFC supports the inclusion of controls. The results for models IC and FC are broadly similar to the pooled-OLS estimations, suggesting one of three possibilities: (i) the pooled-OLS CIA assumption holds; (ii) the pooled-CIA assumption does not hold, and the DiD-CIA does, but the incorrect assumption does not impact the results to a great degree; or (iii) both DiD and pooled-OLS assumptions are incorrectly specified. I discuss case (iii) in the next section.

The final two columns in table 4 report coefficients of regressions including individual fixed effects. As the data contains only two periods these are equivalent to a first-differences design. Model IFE includes all individuals, though in an individual effects model this implies exclusion of 2013 wages for individuals whose 2014 wage is unavailable—predominantly the approximately $2,000$ individuals mentioned in section 2.2 (table 9 in the appendix compares these workers with the rest of the sample). These individuals could be driving the difference between the DiD and individual fixed effects estimators, as they generally receive lower wages than the rest of the population. However, model IFC including only employees with valid wages in both years gives an estimate of 2%, suggesting this is not the issue. Note also that when only non-movers are included with individual fixed effects, a small effect is found (model IFE(SF)).

The likelihood of measurement error is greatly reduced in the DiD design compared to pooled-OLS; not only is the period in question shorter and more recent (alleviating cognitive difficulties in recalling past training), but the data collected by Céreq for this period is more specific, which should help to alleviate possible confusion over the definition of training among those surveyed.

5.2 Validity of the DiD assumptions

Recall the discussion in section 4.2 concerning the problems with the pooled-OLS specification and the assumptions upon which it rests. The key assumption that invalidated the consistency of pooled-OLS options *but which is solvable through a DiD design* relates to an omitted variable. In the omission of variables that affect the decision of people to train, and also affect wages directly, but crucially do not affect the *impact of training on wages*, pooled-OLS estimates are inconsistent but DiD estimates, by controlling for unobserved, time-invariant factors correlated with selection into training, are consistent. The similarity of pooled-OLS and DiD estimates suggest this type of omitted variable bias is not a serious issue. Therefore under the CIA that $(U_1, U_0) \perp T$ conditional on $\boldsymbol{x}$, both the OLS and DiD estimates are consistent estimators of a *unique treatment effect* as

Δ^{ATE}, Δ^{ATT}, and Δ^{MTE} are all equal.

Alternatively, condition (i) from section 4.2 applies and both the DiD and pooled-OLS estimates do not consistently estimate Δ^{ATE}. People not only select into training based on unobserved factor(s), but they select into training based on the *impact of unobserved factor(s) on the wage returns to training.* $\Delta^{MTE}(u_T)$ is not constant. As Δ^{OLS} and Δ^{DiD} are weighted sums of Δ^{MTE} (recall the OLS weights in section 3.1) they are difficult to interpret. The DiD weights are identical to the OLS weights; the only difference being the period in which selection into training occurred, which is likely to have affected which individuals select into training, causing the values of the weights in the sample to differ.The change in weights is cause by a change in $P = \Pr(T = 1)$ as the training variable is *training in 2014* rather than training in their career. The set of controls, $\boldsymbol{X}$ also differs as it now includes past training. The similarity of the DiD and pooled-OLS estimates suggest under condition (i) in section 3.1 the OLS and DiD weights are quite close.

Finally I return to the assumption imposed at the end of section 3.1. As for the pooled-OLS estimates, the interpretation of δ_{DiD} as an estimator of Δ^{ATE} (or Δ^{ATT}) rests on the assumption of no correlation between the returns to training and the covariates. Relaxing this assumption would be interesting. It would require a saturated-in-$\boldsymbol{X}$ model similar to that described at the end of section 4.2 and the estimated coefficient would be a weighted sum of estimated $\Delta^{ATE}(\boldsymbol{x})$, with the weights proportional to the variance of $T \mid \boldsymbol{X}$—i.e. the maximum weight would be on $\boldsymbol{x_i}$ such that $\Pr(T = 1 \mid \boldsymbol{X} = \boldsymbol{x_i}) = \frac{1}{2}$.

6 Instrumenting training

The previous estimation techniques rely on individuals not selecting into training based on their returns; in the notation of Heckman and Vytlacil [2005] $\Delta^{MTE}(u_T)$, was assumed to be constant. As I have noted, there are reasons to believe this may not be the case. In this section I exploit a *policy-relevant instrumental variable* to achieve dual aims: (i) in estimating Imbens and Angrist [1994]'s LATE and comparing the estimate with the ATE estimates of DiD and OLS designs, I conduct a crude test of whether Δ^{MTE} is constant in u_T; (ii) in the case the test suggests selection on unobservable training returns, I estimate a policy-relevant LATE.

6.1 The instrument

The proposed instrument is whether the employee's firm provides information on training.[17] Under the model presented in section 3.1, an individual selects into training based on the (expected) benefits of training exceeding the (expected) costs. These benefits and costs are likely to be multidimensional. Although the focus in this paper is on the pecuniary benefits, other benefits and costs affecting the selection decision present opportunities for exogenous variation. This is the rationale behind using information on training as an instrument. A worker choosing their own training under the DIF, will require information on not only the available training, but also the associated costs and benefits.[18] For individuals in firms that provide information on training to their employees, these informational costs are reduced. This reduction in costs is likely to cause some individuals to train when they would otherwise not, suggesting $Cov(T, Z) \neq 0$. The first stage coefficients in table 6 support this assumption. Whether a firm provides information on training is also unlikely to be correlated with the wage they pay in the absence of training—especially once factors such as firm size and sector are taken into account. In the case where $\Delta^{MTE}(u_T)$ is constant in u_T the LATE estimate associated with this IV will coincide with Δ^{ATE}. Recall that Δ^{LATE} is the sum of $\Delta^{MTE}(u_T)$ who switch into treatment due to a change in the instrument: $\int_{u'_T}^{u_T} \Delta^{MTE}(u)\mathrm{d}u$, where u'_T and u_T correspond to unobserved factors influencing selection for employees indifferent over training at $Z = 0$ and $Z = 1$ respectively. Therefore a crude test of this condition is whether the estimated IV-LATE

[17] Julie Pernaudet suggested the use of information on training as an instrument.

[18] The literature is framed in terms of homogeneous treatments with heterogeneous effects; in the context of training, however, a more realistic assumption is heterogeneous treatments, which may have heterogeneous effects.

Table 5: Summary statistics by training and info. on training

	Entire sample	*By training participation*		*By info. on training*	
		Trained	Not trained	Informed	Not informed
Treatments					
Trained	0.77			0.77	0.67
	(0.42)			(0.42)	(0.39)
Trained (2014)	0.46	0.36		0.28	0.19
	(0.50)	(0.48)		(0.45)	(0.47)
Instrument					
Informed	0.82	0.83	0.74		
	(0.39)	(0.38)	(0.44)		
Outcome					
Wage (2014)	16.01	16.96	12.79	16.38	14.37
	(10.50)	(10.78)	(8.75)	(10.77)	(8.98)
Covariates (mode)					
Age	45-49	40-44	20-24	45-49	45-49
Socio-prof. cat.	C/D	Engineer	C/D	C/D	Qual. worker
Firm size	$> 1{,}000$	$> 1{,}000$	20-49	$> 1{,}000$	10-19

Notes: Mean (or modal) values are displayed in the first row for each variable, with standard deviations in parentheses underneath. For covariates not shown in the table, the modal group did not vary across subsamples. Plots summarising the full sample are available in the Appendix, Figure 2.

effect is close to the pooled-OLS and DiD estimates.

This analysis is complicated if covariates are required to achieve exogeneity of Z. In Imbens and Angrist [1994]'s introduction of LATE, covariates are not mentioned—the instruments they discuss are assumed to be randomly assigned and do not require covariates for exogeneity. Table 5 suggests that this might not be the case here. Larger firms appear more likely to disseminate information on training. Information also appears more prevalent in some sectors than others. As both firm size and sector are likely to be correlated with wages, achieving exogeneity without covariates seems unlikely. Nevertheless, under the assumption that the effects of training are uncorrelated with covariates, Δ^{LATE} is identified when covariates are included. Section 6.3 discusses the consequences of relaxing this assumption. The LATE associated with information on training is policy relevant. For the LATE this is easy to see. If the LATE is significant, a simple policy to harness this benefit would be to oblige firms to provide information on training.

Table 6: IV first stage and wage regressions

	Dependent variable: log(wage)				
Model:	NC (s1)	NC	FC (s1)	FC	SAT
Train. info.	0.09^{***}		0.05^{***}		
	(0.01)		(0.01)		
Training		1.24^{***}		0.72^{***}	$0.24^{\dagger}$
		(0.10)		(0.14)	(0.06)
R^2	0.01	-0.92	0.03	-0.10	0.97
Adj. R^2	0.01	-0.92	0.03	-0.10	0.97
Num. obs.	29,619	29,619	29,619	29,619	29,619
RMSE	0.47	0.65	0.46	0.49	0.43

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. †The manual two-stage estimation produces incorrect standard errors.
NC uses the IV information on training with no other covariates. FC includes controls for sector and firm size. (s1) models are the coefficients of a (linear) first stage. The IV estimates for NC and FC were obtained in `R` using command `ivreg()` from the `AER` package. SAT was run manually in two stages. The full first stage is in the Appendix and the coefficient on training information is not presented here as it does not have the same interpretation. Full average marginal effects are available in the Appendix.

6.2 IV results

Table 6 displays IV estimates of the LATE associated with information on training. Model NC uses training information to instrument for training with no other controls; assuming exogeneity of information on training without the need to condition on covariates. An effect of training on wages of 124% as suggested by the coefficient for NC seems implausibly large—though as noted in the previous section the exogeneity of the instrument likely requires conditioning on firm characteristics. Model FC includes controls for firm sector and firm size, and the estimated effect of training decreases dramatically. An effect of training on wages of 72%, however, still seems very high. These results call into question the constancy of $\Delta^{MTE}\boldsymbol{x}, u_T$ in u_T and hence also in $\boldsymbol{x}$. It is hard to claim constant marginal effects across observed variables faced with evidence of variation across unobserved factors.

6.3 An alternative approach

Although Imbens and Angrist [1994]'s introduction of the LATE does not mention covariates, the same authors discuss the issue in a later paper [Angrist and Imbens, 1995]. The procedure can also be viewed as the 2SLS analogy to that described for dealing with heterogeneous treatment effects that are correlated with covariates in the OLS case (see section 4.2). As in the OLS case, a *fully-saturated* model is required, hence the need for discrete covariates. The fully-saturated model required here is much less computationally intensive than it would be for OLS as exogeneity of the instrument requires fewer covariates.

The procedure involves the following two stages [Angrist and Imbens, 1995]. The first stage is a fully-saturated linear model of training incidence on *all* interactions between the instrument (training information), Z, and the covariates, $\boldsymbol{X}$

$$T_i = \pi_{\boldsymbol{X}} + \pi_{1\boldsymbol{X}} Z_i + \xi_{1i} \tag{20}$$

where $\pi_{\boldsymbol{X}}$ represents the parameters on a full set of dummies for *all possible covariate combinations* and $\pi_{1\boldsymbol{X}}$ the parameters for all combinations interacted with Z. The second stage equation is

$$\ln w_i = \alpha_{\boldsymbol{X}} + \delta_S T_i + \eta_i \tag{21}$$

where $\alpha_{\boldsymbol{X}}$ is the saturated-in-$\boldsymbol{X}$ term for this equation and δ_S is a weighted average of $\Delta^{LATE}(\boldsymbol{x})$'s. Angrist and Imbens [1995] show the weights are equal to

$$\frac{\mathbb{E}\left[\Pr\left(T_i = 1 \mid \boldsymbol{X}_i, Z_i\right)\left(1 - \Pr\left(T_i = 1 \mid \boldsymbol{X}_i, Z_i\right)\right)\right]}{\mathbb{E}\left[\mathbb{E}\left[T_i \mid \boldsymbol{X}_i, Z_i\right]\left(1 - \Pr\left(T_i = 1 \mid \boldsymbol{X}_i, Z_i\right)\right)\right]}. \tag{22}$$

Note that the numerator is the variance of $\mathbb{E}\left[T_i \mid \boldsymbol{X}_i, Z_i\right]$ at each $\boldsymbol{X}_i$. δ_S puts the highest weight on $\Delta^{LATE}(\boldsymbol{x})$'s for which $\Pr(T_i = 1 \mid \boldsymbol{X_i} = \boldsymbol{x_i}, Z_i) = \frac{1}{2}$. These are employees with values of covariates that make them most likely to be affected by a change in Z_i: the *compliers*.

An estimate of δ_S using the DEFIS data is in column SAT of table 6. The estimate of 24% is considerably smaller than the IV estimates for the no-controls or non-saturated firm-controls models. This seems much more realistic, and suggests the NC and FC specifications are capturing effects unrelated to training. Why this might be the case is unclear and warrants further investigation.

7 Discussion and further work

In general the results found in this study are interesting, though perhaps not unexpected. OLS and differences-in-difference specifications found similar estimates of the effect of training on wages to previous studies—both in France and in other countries. The OLS and DiD estimates were around 3% and 1% respectively, with standard IV estimates significantly higher at 74%. The stark difference of the initial IV estimates suggest that the assumptions underlying these specification are too strong—considering that under the required specifications these estimands should coincide. An alternative IV specification which accounts for correlation between effect strength and covariates produces an estimated weighted average treatment effect of 24%. That OLS and DiD appear inadequate is not surprising. The effects of training are certainly heterogeneous. This heterogeneity is in part driven by differences in the type of training chosen by different employees (i.e. heterogeneous treatments), and these choices are likely to be correlated with both observed and unobserved employee characteristics. Therefore I would expect the marginal effect of training to vary with both observed and unobserved characteristics. Given that the OLS and DiD estimates are weighted averages of $\Delta^{MTE}(\boldsymbol{x}, u_T)$ with possibly negative weights [Heckman and Vytlacil, 2005], the IV estimates are evidence that at least for part of the population, training has a positive effect on wages.

The results also pose a number of questions. The large difference between the standard 2SLS estimates and the other estimates, including Angrist and Imbens [1995]'s *saturate-and-weight* estimator certainly warrants further investigation. Abadie [2003] proposes a similar methodology which enables estimation of the training effect on the specific population of compliers, rather than the weighted estimate which places more weight on these individuals. Abadie's contribution also allows identification of the compliers, which would be an interesting exercise from both academic and policy positions. I would also like to relax the assumption maintained throughout most of the paper that Δ^{ATE} is independent of the covariates. I discuss how this might be possible, and comparing these less restrictive estimations with the IV estimates presented here, and estimates using Abadie [2003]'s procedure would permit a deeper understanding of who chooses to train and the relationship between this choice and the wage returns to training.

Finally, something that has not been considered in this paper is the effect of training on productivity. The analysis presented has ignored this process, focusing solely on wage returns to training. Currently the unavailability of detailed data on firm profits and the limited data on wages restrict the possibilities to identify direct productivity effects, and therefore which party captures a larger slice of the pie. The DEFIS data, however, has the possibility to be linked to both firm data held by the French state, and also to longer periods of employee wages. Exploiting this additional data, were it made available, along the lines of Dearden et al. [2000]'s study of British industry would be an interesting extension. This could be combined with estimates of how workers' returns to training vary with their characteristics to further our understanding of the relationship between firm and worker returns to training; a key issue in the theoretical literature on training.

A Data

A.1 Firms

Table 7: Firms and training by firm size

Firm size (employees)	Number of firms	Percentage training	
		Firms	Employees
3-9	1,150	n/a	n/a
10-19	931	81	34
20-49	866	94	36
50-249	787	97	47
250-499	188	99	50
500-1999	284	99	53
> 2000	323	100	57
Total	4,525	88	49

Source: Dubois et al. [2016].
Note: For example, 81% of firms with 10-19 employees have trained at least one employee.

Table 8: Firms and training by sector

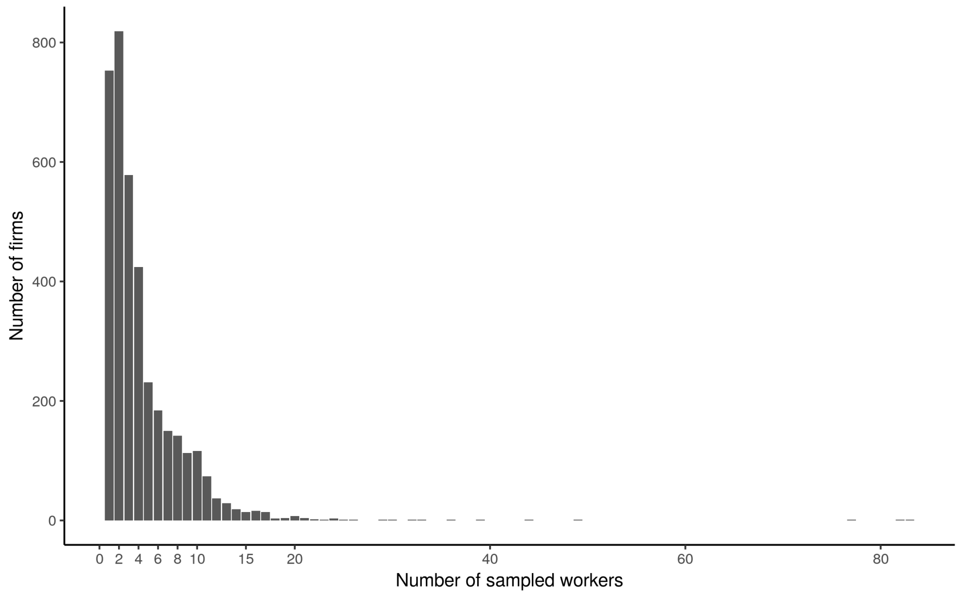
Sector	Number of firms	Percentage training: Firms	Percentage training: Employees
C3 / C4 - Manufacture of electrical, electronic and computer equipment; Manufacturing machinery; Manufacture of transport equipment	161	89	53
C5 - Manufacture of other industrial products	421	89	50
CR - Manufacture of food, drink and beverages Tobacco—coking and refining	226	84	52
DE - Mining, energy, water, waste management and decontamination	47	91	74
FZ - Construction	563	83	42
GZ - Commerce, repair of automobiles and motorcycles	1,153	91	45
HZ - Transport and storage	211	90	55
IZ - Accommodation and catering	457	75	38
JZ - Information and Communication	147	96	54
KZ / LZ - Real estate, financial and insurance activities	195	98	73
MN - Scientific and technical activities; Administrative and support	665	89	39
OQ / UK - Other service activities	283	88	46
Total	4,529	88	49

Source: Dubois et al. [2016].
Note: For example, 83% of construction firms have trained at least one employee, and construction firms train 42% of their employees on average.

Table 9: Comparison statistics for workers with zero hours in 2014

	Zero hours (2014)	Rest of sample
Training	0.64	0.72
(2014)	0.19	0.26
episodes	0.44	0.64
Wage (2013) EUR	12.95	15.48
Trained	14.05	16.88
Female	0.24	0.32
Same firm (2015)	0.58	0.82
Socio-prof. cat.	Qual. worker	Engineer
Sector	FZ	GZ
Contract (2013)	CDI	CDI
Contract (2015)	CDI	CDI
Age	45-49	40-44

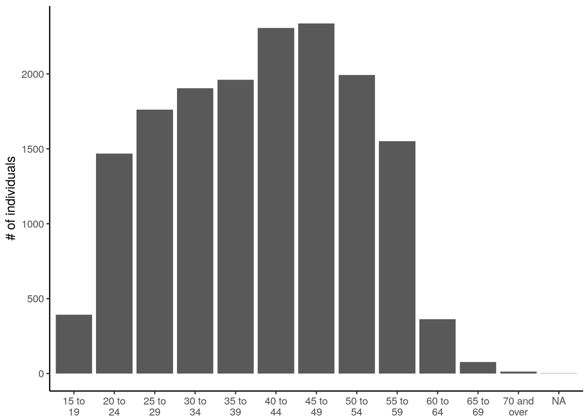
Figure 1: Count of firms by number of sampled workers



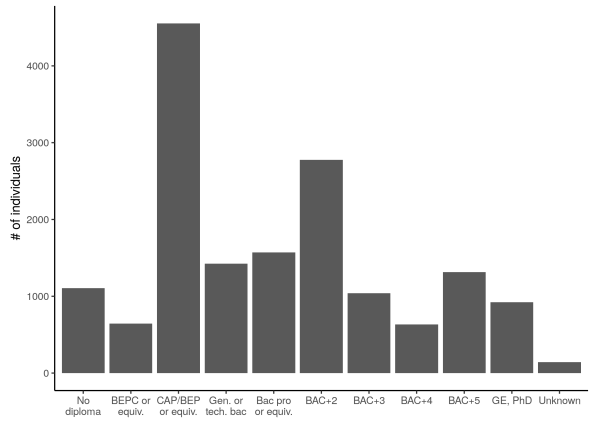
A.2 Employees

Figure 2: Summary plots of the DEFIS data

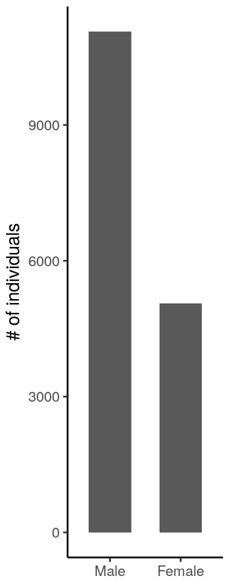
(a) Age distribution



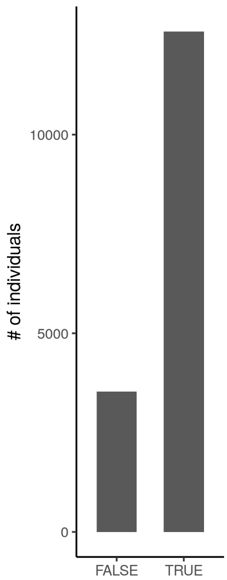
(b) Highest education level



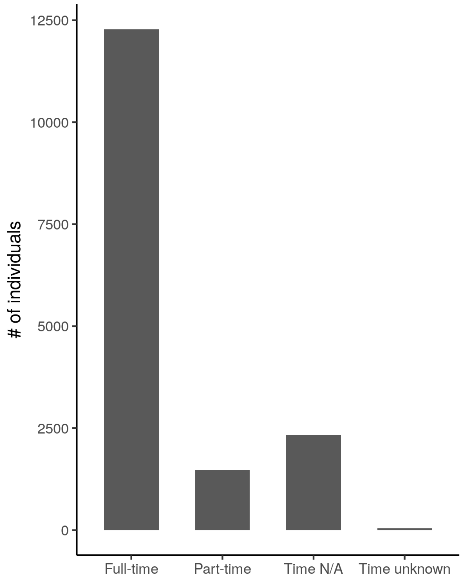
(c) Gender



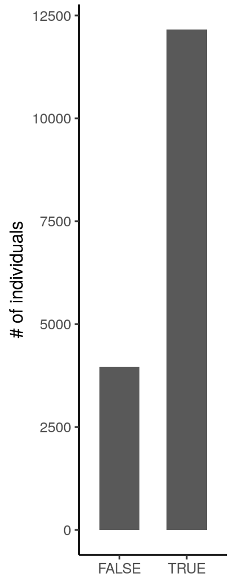
(d) Mobility



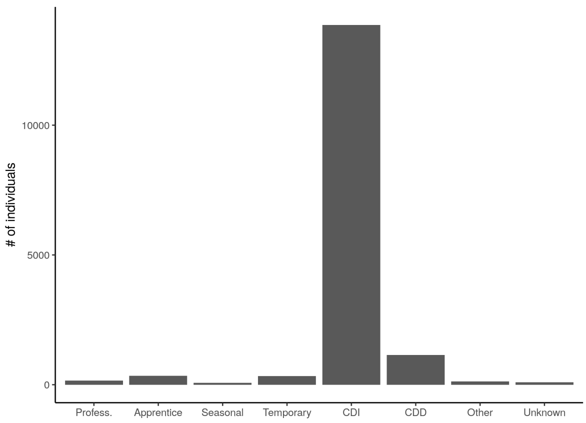
(e) FT/PT



(f) Training



(g) Contract type in 2013



(h) Socio-professional category in 2013

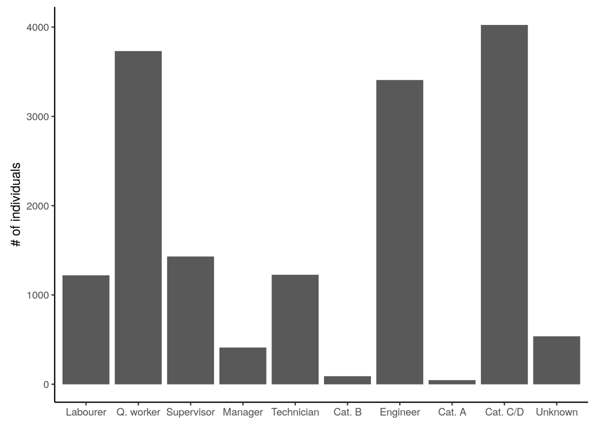
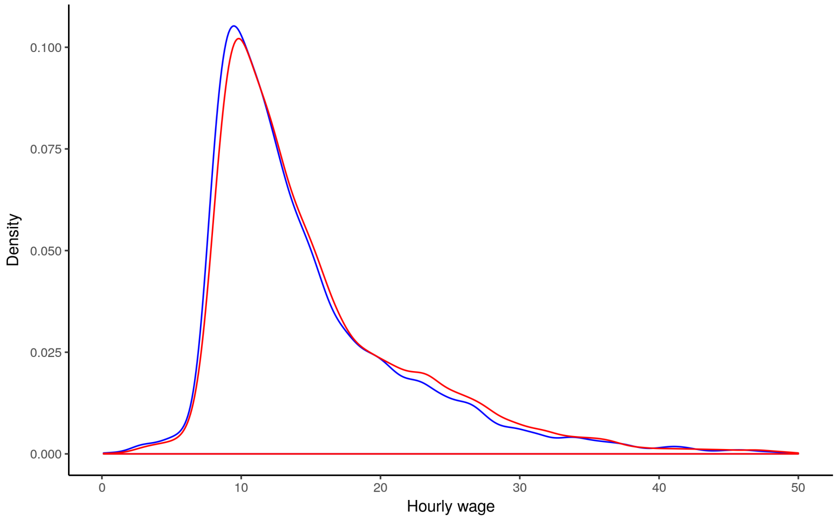
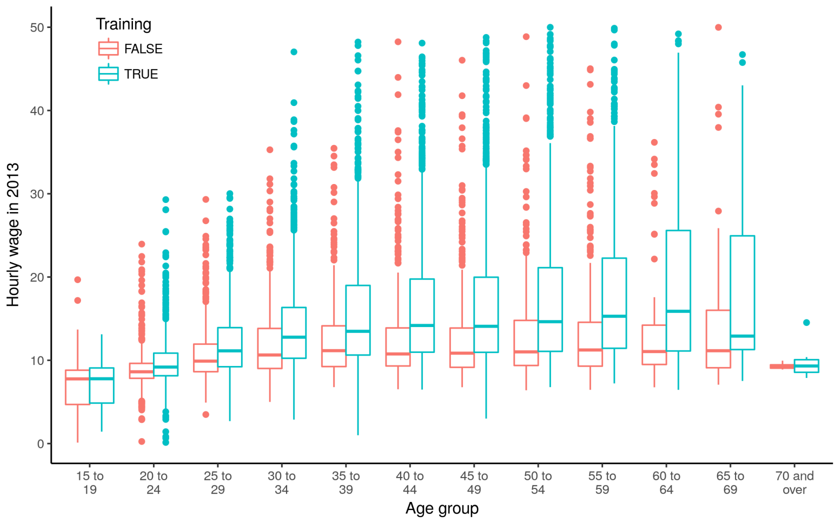


Figure 3: Descriptive analysis of wages and training

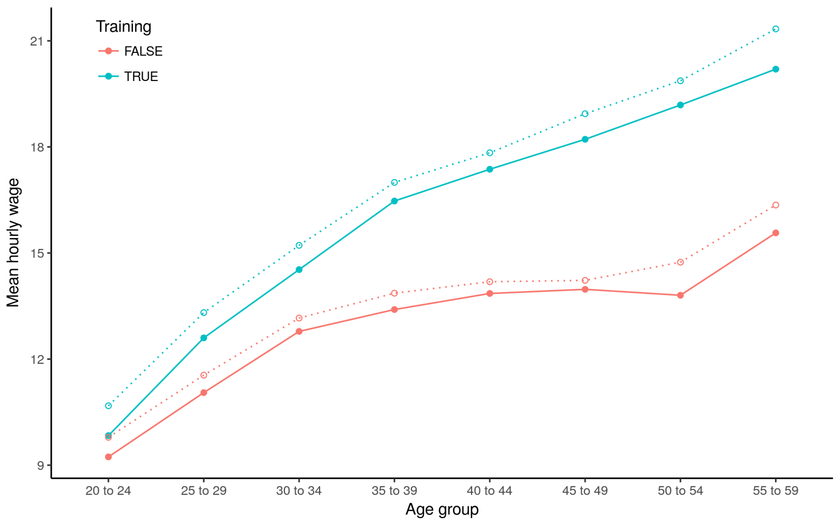
(a) Wage distributions in 2013 (blue) and 2014 (red)



(b) Age-wage profiles by training status



(c) Age-mean wage profiles in 2013 and 2014 by training status



B Regression outputs

Table 10: Full OLS results on pooled data

	IC	FC	ICFC	ICFFE
(Intercept)	1.98***	2.46***	1.95***	2.42***
	(0.03)	(0.02)	(0.03)	(0.18)
Training	0.04***	0.21***	0.03***	0.02***
	(0.00)	(0.01)	(0.00)	(0.00)
Contract type				
Apprentice (2013)	−0.32***		−0.29***	−0.34***
	(0.06)		(0.06)	(0.06)
Seasonal work (2013)	0.15***		0.18***	0.15***
	(0.03)		(0.03)	(0.03)
Temporary contract (2013)	0.10		0.14*	0.14*
	(0.06)		(0.06)	(0.06)
CDI inc. civil service (2013)	0.12***		0.14***	0.12***
	(0.02)		(0.02)	(0.02)
CDD ou vacataire (2013)	0.09***		0.10***	0.08***
	(0.02)		(0.02)	(0.02)
Other	−0.00		0.03	−0.03
	(0.03)		(0.03)	(0.03)
Unknown	0.05		0.09**	0.05
	(0.03)		(0.03)	(0.03)
Education				
BEPC or similar	0.01		0.02	0.02*
	(0.01)		(0.01)	(0.01)
CAP, BEP or similar	0.01		0.02*	0.02*
	(0.01)		(0.01)	(0.01)
Bac gen or tech	0.10***		0.09***	0.07***
	(0.01)		(0.01)	(0.01)
Bac pro or similar	0.09***		0.09***	0.07***
	(0.01)		(0.01)	(0.01)
Bac+2 (DEUG, BTS, DUT)	0.16***		0.15***	0.11***
	(0.01)		(0.01)	(0.01)
Bac+3 (licence gen or pro)	0.18***		0.17***	0.12***
	(0.01)		(0.01)	(0.01)
Bac+4 (Maitrise, M1)	0.20***		0.19***	0.13***
	(0.01)		(0.01)	(0.01)
Bac+5 (DESS, DEA, M2)	0.28***		0.26***	0.19***
	(0.01)		(0.01)	(0.01)
Grande ecole, PhD	0.39***		0.37***	0.28***
	(0.01)		(0.01)	(0.01)
Education unknown	0.08***		0.09***	0.07***
	(0.02)		(0.02)	(0.02)

	IC	FC	ICFC	ICFFE
Socio-prof. cat.				
Qualified worker	0.03***		0.04***	0.02**
	(0.01)		(0.01)	(0.01)
Supervisor	0.18***		0.19***	0.16***
	(0.01)		(0.01)	(0.01)
Manager or deputy manager	0.73***		0.76***	0.81***
	(0.01)		(0.01)	(0.01)
Technician, draftsman	0.15***		0.13***	0.09***
	(0.01)		(0.01)	(0.01)
Category B	0.19***		0.20***	0.16***
	(0.02)		(0.02)	(0.02)
Engineer	0.46***		0.46***	0.43***
	(0.01)		(0.01)	(0.01)
Category A	0.18***		0.22***	0.15***
	(0.03)		(0.03)	(0.04)
Category C or D	0.04***		0.06***	0.05***
	(0.01)		(0.01)	(0.01)
Soc-prof unknown	0.06***		0.08***	0.07***
	(0.01)		(0.01)	(0.01)
Role (2013)				
Repair, maintenance	−0.01		−0.00	−0.01*
	(0.01)		(0.01)	(0.01)
Security, cleaning	−0.15***		−0.13***	−0.08***
	(0.01)		(0.01)	(0.01)
Logisitics	−0.08***		−0.07***	−0.05***
	(0.01)		(0.01)	(0.01)
Admin	−0.03**		−0.00	−0.00
	(0.01)		(0.01)	(0.01)
Management, accounting	0.02**		0.04***	0.03***
	(0.01)		(0.01)	(0.01)
Commercial	0.02***		0.04***	0.06***
	(0.01)		(0.01)	(0.01)
R&D	0.00		0.01*	−0.01
	(0.01)		(0.01)	(0.01)
Role unknown	−0.01		0.01	0.03***
	(0.01)		(0.01)	(0.01)
FT/PT				
Part-time	−0.06***		−0.04***	−0.02**
	(0.01)		(0.01)	(0.01)
Time N/A	−0.00		−0.02	−0.04
	(0.06)		(0.05)	(0.05)
Time unknown	0.04		0.06*	0.06*
	(0.03)		(0.03)	(0.03)
Female	−0.11***		−0.11***	−0.10***
	(0.00)		(0.00)	(0.00)
Age				
20 to 24	0.09***		0.06***	0.02

	IC	FC	ICFC	ICFFE
	(0.01)		(0.01)	(0.01)
25 to 29	0.07***		0.05**	0.00
	(0.02)		(0.02)	(0.02)
30 to 34	0.14***		0.11***	0.06***
	(0.02)		(0.02)	(0.02)
35 to 39	0.20***		0.16***	0.11***
	(0.02)		(0.02)	(0.02)
40 to 44	0.24***		0.21***	0.14***
	(0.02)		(0.02)	(0.02)
45 to 49	0.29***		0.25***	0.18***
	(0.02)		(0.02)	(0.02)
50 to 54	0.34***		0.30***	0.22***
	(0.02)		(0.02)	(0.02)
55 to 59	0.41***		0.36***	0.28***
	(0.02)		(0.02)	(0.02)
60 to 64	0.44***		0.40***	0.32***
	(0.02)		(0.02)	(0.02)
65 to 69	0.36***		0.33***	0.26***
	(0.03)		(0.03)	(0.03)
70 and over	0.14*		0.11	0.22***
	(0.07)		(0.06)	(0.07)
Graduation year				
1976 to 1982	0.04***		0.04***	0.03**
	(0.01)		(0.01)	(0.01)
1983 to 1993	0.08***		0.08***	0.06***
	(0.01)		(0.01)	(0.01)
1994 to 2010	0.06***		0.06***	0.05***
	(0.01)		(0.01)	(0.01)
2011 to 2014	−0.05**		−0.04*	−0.05**
	(0.02)		(0.02)	(0.02)
No schooling	0.02		0.01	−0.01
	(0.02)		(0.02)	(0.02)
Still in school	−0.10***		−0.11***	−0.13***
	(0.02)		(0.02)	(0.02)
Grad. year unknown	−0.01		0.00	−0.01
	(0.01)		(0.01)	(0.01)
Sector				
C4		−0.03	0.03*	
		(0.02)	(0.01)	
C5		−0.06***	0.02*	
		(0.01)	(0.01)	
CR		−0.22***	−0.04***	
		(0.02)	(0.01)	
DE		−0.07**	0.03*	
		(0.02)	(0.02)	
FZ		−0.19***	−0.06***	
		(0.02)	(0.01)	

	IC	FC	ICFC	ICFFE
GZ		−0.21***	−0.08***	
		(0.01)	(0.01)	
HZ		−0.22***	−0.05***	
		(0.02)	(0.01)	
IZ		−0.34***	−0.12***	
		(0.02)	(0.01)	
JZ		0.10***	−0.06***	
		(0.02)	(0.01)	
KZ		0.08***	0.02	
		(0.02)	(0.01)	
LZ		−0.09***	−0.05**	
		(0.02)	(0.02)	
MN		−0.16***	−0.08***	
		(0.01)	(0.01)	
OQ		−0.22***	−0.07***	
		(0.02)	(0.01)	
RU		−0.20***	−0.04*	
		(0.02)	(0.02)	
Firm size				
1 to 9		0.09***	0.04***	
		(0.01)	(0.01)	
10 to 19		0.11***	0.06***	
		(0.01)	(0.01)	
20 to 49		0.14***	0.07***	
		(0.01)	(0.01)	
50 to 249		0.21***	0.12***	
		(0.01)	(0.01)	
250 to 499		0.17***	0.12***	
		(0.01)	(0.01)	
500 to 999		0.20***	0.14***	
		(0.01)	(0.01)	
1000 plus		0.24***	0.15***	
		(0.01)	(0.01)	
R^2	0.58	0.15	0.60	
Adj. R^2	0.58	0.15	0.60	
Num. obs.	29619	29619	29619	29619
RMSE	0.30	0.43	0.30	
R^2 (full model)				0.75
R^2 (proj model)				0.75
Adj. R^2 (full model)				0.71
Adj. R^2 (proj model)				0.71

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Notes: Model NC regresses ln(wage) on training. FFE adds firm fixed effects. IC adds controls for education, contract type, socio-professional category, role, FT/PT, gender, age and graduation year. ICFC adds controls for sector and firm-size. ICFFE drops adds firm fixed effects to IC. All controls (including firm fixed effects) refer to 2013. Regressions without fixed effects were implemented in `R` using command `lm()`. Regressions including fixed effects were implemented in `R` using command `felm()`.

Table 11: DiD wage regressions

	IC	FC	IFC	IFE	IFE(SF)
(Intercept)	1.97***	2.44***	1.93***		
	(0.03)	(0.02)	(0.03)		
Training (2014)	0.04***	0.09***	0.02***	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Training (pre-2014)	0.03***	0.20***	0.02***		
	(0.00)	(0.01)	(0.00)		
Contract type					
Apprentice (2013)	−0.32***		−0.29***		
	(0.06)		(0.06)		
Seasonal work (2013)	0.15***		0.18***		
	(0.03)		(0.03)		
Temporary contract (2013)	0.10		0.14*		
	(0.06)		(0.06)		
CDI inc. civil service (2013)	0.12***		0.13***		
	(0.02)		(0.02)		
CDD ou vacataire (2013)	0.09***		0.10***		
	(0.02)		(0.02)		
Other contract (2013)	−0.00		0.03		
	(0.03)		(0.03)		
Contract (2013) unknown	0.05		0.09**		
	(0.03)		(0.03)		
Education					
BEPC or similar	0.01		0.02		
	(0.01)		(0.01)		
CAP, BEP or similar	0.01		0.02*		
	(0.01)		(0.01)		
Bac gen or tech	0.10***		0.09***		
	(0.01)		(0.01)		
Bac pro or similar	0.09***		0.09***		
	(0.01)		(0.01)		
Bac+2 (DEUG, BTS, DUT)	0.16***		0.15***		
	(0.01)		(0.01)		
Bac+3 (licence gen or pro)	0.18***		0.17***		
	(0.01)		(0.01)		
Bac+4 (Maitrise, M1)	0.20***		0.19***		
	(0.01)		(0.01)		
Bac+5 (DESS, DEA, M2)	0.28***		0.26***		
	(0.01)		(0.01)		
Grande ecole, PhD	0.39***		0.36***		
	(0.01)		(0.01)		
Education unknown	0.08***		0.09***		
	(0.02)		(0.02)		
Socio-prof. cat.					

	IC	FC	IFC	IFE	IFE(SF)
Qualified worker	0.03***		0.04***		
	(0.01)		(0.01)		
Supervisor	0.18***		0.19***		
	(0.01)		(0.01)		
Manager or deputy manager	0.73***		0.76***		
	(0.01)		(0.01)		
Technician, draftsman	0.15***		0.13***		
	(0.01)		(0.01)		
Category B	0.19***		0.20***		
	(0.02)		(0.02)		
Engineer	0.46***		0.46***		
	(0.01)		(0.01)		
Category A	0.18***		0.21***		
	(0.03)		(0.03)		
Category C or D	0.04***		0.06***		
	(0.01)		(0.01)		
Soc-prof unknown	0.06***		0.08***		
	(0.01)		(0.01)		
Role (2013)					
Repair, maintenance	−0.01*		−0.00		
	(0.01)		(0.01)		
Security, cleaning	−0.15***		−0.13***		
	(0.01)		(0.01)		
Logisitics	−0.08***		−0.07***		
	(0.01)		(0.01)		
Admin	−0.03**		−0.00		
	(0.01)		(0.01)		
Management, accounting	0.02**		0.04***		
	(0.01)		(0.01)		
Commercial	0.02***		0.04***		
	(0.01)		(0.01)		
R&D	−0.00		0.01*		
	(0.01)		(0.01)		
Role unknown	−0.01		0.01		
	(0.01)		(0.01)		
FT/PT					
Part-time	−0.05***		−0.04***		
	(0.01)		(0.01)		
Time N/A	−0.00		−0.02		
	(0.06)		(0.05)		
Time unknown	0.04		0.06*		
	(0.03)		(0.03)		
Female	−0.11***		−0.11***		
	(0.00)		(0.00)		
Age					
20 to 24	0.09***		0.06***		
	(0.01)		(0.01)		

	IC	FC	IFC	IFE	IFE(SF)
25 to 29	0.07***		0.05**		
	(0.02)		(0.02)		
30 to 34	0.15***		0.12***		
	(0.02)		(0.02)		
35 to 39	0.20***		0.16***		
	(0.02)		(0.02)		
40 to 44	0.25***		0.21***		
	(0.02)		(0.02)		
45 to 49	0.29***		0.25***		
	(0.02)		(0.02)		
50 to 54	0.34***		0.30***		
	(0.02)		(0.02)		
55 to 59	0.41***		0.36***		
	(0.02)		(0.02)		
60 to 64	0.44***		0.40***		
	(0.02)		(0.02)		
65 to 69	0.36***		0.33***		
	(0.03)		(0.03)		
70 and over	0.14*		0.11		
	(0.07)		(0.06)		
Graduation year					
1976 to 1982	0.04***		0.04***		
	(0.01)		(0.01)		
1983 to 1993	0.08***		0.08***		
	(0.01)		(0.01)		
1994 to 2010	0.06***		0.06***		
	(0.01)		(0.01)		
2011 to 2014	−0.05**		−0.04*		
	(0.02)		(0.02)		
No schooling	0.01		0.01		
	(0.02)		(0.02)		
Still in school	−0.11***		−0.11***		
	(0.02)		(0.02)		
Grad. year unknown	−0.01		0.00		
	(0.01)		(0.01)		
2014	0.04***	0.03***	0.04***	0.04***	0.03***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Sector					
C4		−0.03	0.03*		
		(0.02)	(0.01)		
C5		−0.06***	0.02*		
		(0.01)	(0.01)		
CR		−0.22***	−0.04***		
		(0.02)	(0.01)		
DE		−0.07**	0.03*		
		(0.02)	(0.02)		
FZ		−0.17***	−0.05***		

	IC	FC	IFC	IFE	IFE(SF)
		(0.02)	(0.01)		
GZ		-0.21^{***}	-0.08^{***}		
		(0.01)	(0.01)		
HZ		-0.21^{***}	-0.05^{***}		
		(0.02)	(0.01)		
IZ		-0.34^{***}	-0.12^{***}		
		(0.02)	(0.01)		
JZ		0.10^{***}	-0.06^{***}		
		(0.02)	(0.01)		
KZ		0.08^{***}	0.02		
		(0.02)	(0.01)		
LZ		-0.09^{***}	-0.05^{**}		
		(0.02)	(0.02)		
MN		-0.15^{***}	-0.08^{***}		
		(0.01)	(0.01)		
OQ		-0.22^{***}	-0.07^{***}		
		(0.02)	(0.01)		
RU		-0.20^{***}	-0.04^{*}		
		(0.02)	(0.02)		
Firm size					
1 to 9		0.09^{***}	0.04^{***}		
		(0.01)	(0.01)		
10 to 19		0.11^{***}	0.06^{***}		
		(0.01)	(0.01)		
20 to 49		0.14^{***}	0.07^{***}		
		(0.01)	(0.01)		
50 to 249		0.21^{***}	0.12^{***}		
		(0.01)	(0.01)		
250 to 499		0.17^{***}	0.12^{***}		
		(0.01)	(0.01)		
500 to 999		0.19^{***}	0.14^{***}		
		(0.01)	(0.01)		
1000 plus		0.23^{***}	0.15^{***}		
		(0.01)	(0.01)		
R^2	0.58	0.16	0.60	0.03	0.02
Adj. R^2	0.58	0.16	0.60	-1.12	-1.08
Num. obs.	29619	29619	29619	29619	23671
RMSE	0.30	0.43	0.29		

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Model NC regresses $\ln(w)$ on past and 2014 training with year dummies. IC includes controls for contract type, education, socio-professional category, role, FT, gender, age and graduation year. FC includes controls for firm size and sector. ICFC combines all controls from IC and FC. IFE includes individual and year fixed effects and controls for education, sector, FT/PT, age, and graduation year. IFE(SF) uses the same specification as IFE but only including individuals who remained at the same firm in 2014. Regressions without fixed effects were implemented in `R` using command `lm()`. Regressions including fixed effects were implemented in `R` using command `felm()`.

Table 12: IV first stage and wage regressions

	FC (s1)	FC	SAT
(Intercept)	0.60 (0.02)***	2.13 (0.09)***	
Training info.	0.05 (0.01)***		
SECTORC4	−0.01 (0.02)	−0.02 (0.02)	2.24 (0.16)***
SECTORC5	−0.05 (0.01)***	−0.03 (0.02)	2.25 (0.09)***
SECTORCR	−0.15 (0.02)***	−0.14 (0.03)***	1.99 (0.04)***
SECTORDE	−0.06 (0.03)*	−0.04 (0.03)	2.38 (0.16)***
SECTORFZ	−0.10 (0.02)***	−0.13 (0.02)***	2.17 (0.04)***
SECTORGZ	−0.09 (0.01)***	−0.17 (0.02)***	2.24 (0.04)***
SECTORHZ	−0.10 (0.02)***	−0.16 (0.02)***	2.29 (0.08)***
SECTORIZ	−0.22 (0.02)***	−0.23 (0.04)***	2.12 (0.04)***
SECTORJZ	0.00 (0.02)	0.09 (0.02)***	2.83 (0.09)***
SECTORKZ	−0.01 (0.02)	0.08 (0.02)***	2.67 (0.19)***
SECTORLZ	0.02 (0.03)	−0.10 (0.03)***	2.41 (0.08)***
SECTORMN	−0.11 (0.01)***	−0.10 (0.02)***	2.40 (0.05)***
SECTOROQ	−0.08 (0.02)***	−0.18 (0.02)***	2.29 (0.09)***
SECTORRU	−0.11 (0.02)***	−0.14 (0.03)***	2.35 (0.06)***
NB_EMPLYEES_X1 to 9	0.08 (0.01)***	0.05 (0.02)**	0.18 (0.16)
NB_EMPLYEES_X10 to 19	0.08 (0.01)***	0.06 (0.02)***	0.11 (0.16)
NB_EMPLYEES_X20 to 49	0.10 (0.01)***	0.08 (0.02)***	0.21 (0.16)
NB_EMPLYEES_X50 to 249	0.12 (0.01)***	0.14 (0.02)***	0.27 (0.16)
NB_EMPLYEES_X250 to 499	0.13 (0.01)***	0.10 (0.03)***	0.29 (0.16)
NB_EMPLYEES_X500 to 999	0.13 (0.02)***	0.12 (0.03)***	0.29 (0.16)
NB_EMPLYEES_X1000 plus	0.15 (0.01)***	0.15 (0.03)***	0.48 (0.16)**
TRAINING1TRUE		0.72 (0.14)***	
pred_train			0.24 (0.06)***
SECTORC3			2.31 (0.16)***
SECTORC4:NB_EMPLYEES_X1 to 9			−0.13 (0.17)
SECTORC5:NB_EMPLYEES_X1 to 9			−0.00 (0.18)
SECTORCR:NB_EMPLYEES_X1 to 9			−0.00 (0.17)
SECTORDE:NB_EMPLYEES_X1 to 9			−0.16 (0.25)
SECTORFZ:NB_EMPLYEES_X1 to 9			−0.05 (0.16)
SECTORGZ:NB_EMPLYEES_X1 to 9			−0.07 (0.16)
SECTORHZ:NB_EMPLYEES_X1 to 9			−0.25 (0.18)
SECTORIZ:NB_EMPLYEES_X1 to 9			−0.10 (0.16)
SECTORJZ:NB_EMPLYEES_X1 to 9			−0.48 (0.18)**
SECTORKZ:NB_EMPLYEES_X1 to 9			−0.29 (0.25)
SECTORLZ:NB_EMPLYEES_X1 to 9			−0.09 (0.19)
SECTORMN:NB_EMPLYEES_X1 to 9			−0.12 (0.16)
SECTOROQ:NB_EMPLYEES_X1 to 9			−0.14 (0.19)
SECTORRU:NB_EMPLYEES_X1 to 9			−0.29 (0.17)
SECTORC4:NB_EMPLYEES_X10 to 19			0.05 (0.08)
SECTORC5:NB_EMPLYEES_X10 to 19			0.05 (0.18)
SECTORCR:NB_EMPLYEES_X10 to 19			0.17 (0.17)
SECTORDE:NB_EMPLYEES_X10 to 19			−0.15 (0.23)
SECTORFZ:NB_EMPLYEES_X10 to 19			0.04 (0.16)

	FC (s1)	FC	SAT
SECTORGZ:NB_EMPLYEES_X10 to 19			0.07 (0.16)
SECTORHZ:NB_EMPLYEES_X10 to 19			−0.16 (0.17)
SECTORIZ:NB_EMPLYEES_X10 to 19			−0.03 (0.16)
SECTORJZ:NB_EMPLYEES_X10 to 19			−0.33 (0.18)
SECTORKZ:NB_EMPLYEES_X10 to 19			−0.15 (0.25)
SECTORLZ:NB_EMPLYEES_X10 to 19			0.03 (0.19)
SECTORMN:NB_EMPLYEES_X10 to 19			−0.05 (0.16)
SECTOROQ:NB_EMPLYEES_X10 to 19			−0.04 (0.18)
SECTORRU:NB_EMPLYEES_X10 to 19			−0.21 (0.17)
SECTORC4:NB_EMPLYEES_X20 to 49			0.01 (0.05)
SECTORC5:NB_EMPLYEES_X20 to 49			0.02 (0.17)
SECTORCR:NB_EMPLYEES_X20 to 49			0.23 (0.16)
SECTORDE:NB_EMPLYEES_X20 to 49			−0.15 (0.22)
SECTORFZ:NB_EMPLYEES_X20 to 49			0.11 (0.16)
SECTORGZ:NB_EMPLYEES_X20 to 49			−0.10 (0.16)
SECTORHZ:NB_EMPLYEES_X20 to 49			−0.21 (0.17)
SECTORIZ:NB_EMPLYEES_X20 to 49			−0.14 (0.16)
SECTORJZ:NB_EMPLYEES_X20 to 49			−0.32 (0.17)
SECTORKZ:NB_EMPLYEES_X20 to 49			−0.18 (0.24)
SECTORLZ:NB_EMPLYEES_X20 to 49			−0.09 (0.17)
SECTORMN:NB_EMPLYEES_X20 to 49			−0.20 (0.16)
SECTOROQ:NB_EMPLYEES_X20 to 49			−0.07 (0.18)
SECTORRU:NB_EMPLYEES_X20 to 49			−0.21 (0.17)
SECTORC4:NB_EMPLYEES_X50 to 249			0.12 (0.08)
SECTORC5:NB_EMPLYEES_X50 to 249			0.10 (0.18)
SECTORCR:NB_EMPLYEES_X50 to 249			0.14 (0.17)
SECTORDE:NB_EMPLYEES_X50 to 249			0.34 (0.25)
SECTORFZ:NB_EMPLYEES_X50 to 249			0.03 (0.17)
SECTORGZ:NB_EMPLYEES_X50 to 249			−0.10 (0.16)
SECTORHZ:NB_EMPLYEES_X50 to 249			−0.09 (0.17)
SECTORIZ:NB_EMPLYEES_X50 to 249			−0.04 (0.21)
SECTORJZ:NB_EMPLYEES_X50 to 249			−0.44 (0.18)*
SECTORKZ:NB_EMPLYEES_X50 to 249			−0.22 (0.25)
SECTORLZ:NB_EMPLYEES_X50 to 249			−0.26 (0.18)
SECTORMN:NB_EMPLYEES_X50 to 249			−0.12 (0.16)
SECTOROQ:NB_EMPLYEES_X50 to 249			−0.18 (0.18)
SECTORRU:NB_EMPLYEES_X50 to 249			−0.37 (0.20)
SECTORC4:NB_EMPLYEES_X250 to 499			−0.05 (0.06)
SECTORC5:NB_EMPLYEES_X250 to 499			0.05 (0.18)
SECTORCR:NB_EMPLYEES_X250 to 499			0.21 (0.16)
SECTORDE:NB_EMPLYEES_X250 to 499			−0.33 (0.24)
SECTORFZ:NB_EMPLYEES_X250 to 499			−0.06 (0.17)
SECTORGZ:NB_EMPLYEES_X250 to 499			−0.12 (0.16)
SECTORHZ:NB_EMPLYEES_X250 to 499			−0.15 (0.17)
SECTORIZ:NB_EMPLYEES_X250 to 499			−0.17 (0.17)
SECTORJZ:NB_EMPLYEES_X250 to 499			−0.33 (0.18)
SECTORKZ:NB_EMPLYEES_X250 to 499			−0.20 (0.24)

	FC (s1)	FC	SAT
SECTORLZ:NB_EMPLYEES_X250 to 499			−0.38 (0.18)*
SECTORMN:NB_EMPLYEES_X250 to 499			−0.35 (0.16)*
SECTOROQ:NB_EMPLYEES_X250 to 499			−0.27 (0.18)
SECTORRU:NB_EMPLYEES_X250 to 499			−0.06 (0.18)
SECTORC4:NB_EMPLYEES_X500 to 999			0.00 (0.07)
SECTORC5:NB_EMPLYEES_X500 to 999			0.07 (0.18)
SECTORCR:NB_EMPLYEES_X500 to 999			0.11 (0.17)
SECTORDE:NB_EMPLYEES_X500 to 999			−0.14 (0.25)
SECTORFZ:NB_EMPLYEES_X500 to 999			−0.02 (0.17)
SECTORGZ:NB_EMPLYEES_X500 to 999			−0.09 (0.16)
SECTORHZ:NB_EMPLYEES_X500 to 999			0.20 (0.18)
SECTORIZ:NB_EMPLYEES_X500 to 999			−0.07 (0.18)
SECTORJZ:NB_EMPLYEES_X500 to 999			−0.42 (0.18)*
SECTORKZ:NB_EMPLYEES_X500 to 999			−0.31 (0.24)
SECTORMN:NB_EMPLYEES_X500 to 999			−0.24 (0.16)
SECTOROQ:NB_EMPLYEES_X500 to 999			−0.50 (0.19)*
SECTORRU:NB_EMPLYEES_X500 to 999			0.25 (0.21)
SECTORC5:NB_EMPLYEES_X1000 plus			−0.01 (0.17)
SECTORCR:NB_EMPLYEES_X1000 plus			0.12 (0.16)
SECTORDE:NB_EMPLYEES_X1000 plus			−0.20 (0.22)
SECTORFZ:NB_EMPLYEES_X1000 plus			−0.13 (0.16)
SECTORGZ:NB_EMPLYEES_X1000 plus			−0.34 (0.16)*
SECTORHZ:NB_EMPLYEES_X1000 plus			−0.29 (0.17)
SECTORIZ:NB_EMPLYEES_X1000 plus			−0.28 (0.16)
SECTORJZ:NB_EMPLYEES_X1000 plus			−0.56 (0.17)**
SECTORKZ:NB_EMPLYEES_X1000 plus			−0.40 (0.24)
SECTORLZ:NB_EMPLYEES_X1000 plus			−0.29 (0.19)
SECTORMN:NB_EMPLYEES_X1000 plus			−0.43 (0.16)**
SECTOROQ:NB_EMPLYEES_X1000 plus			−0.59 (0.18)**
SECTORRU:NB_EMPLYEES_X1000 plus			−0.22 (0.18)
R^2	0.03	−0.10	0.97
Adj. R^2	0.03	−0.10	0.97
Num. obs.	29619	29619	29619
RMSE	0.46	0.49	0.43

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
NC uses the IV information on training with no other covariates. FC includes controls for sector and firm size. (s1) models are the coefficients of a (linear) first stage. The IV estimates for NC and FC were obtained in `R` using command `ivreg()` from the `AER` package. SAT was run manually in two stages. Therefore the standard errors here are incorrect.

References

Alberto Abadie. Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics*, 113(2):231–263, 2003. URL http://www.sciencedirect.com/science/article/pii/S0304407602002014.

Daron Acemoglu. Training and Innovation in an Imperfect Labour Market. *The Review of Economic Studies*, 64(3):445, July 1997. ISSN 00346527. doi: 10.2307/2971723. URL https://academic.oup.com/restud/article-lookup/doi/10.2307/2971723.

Daron Acemoglu. Why do new technologies complement skills? Directed technical change and wage inequality. *Quarterly journal of economics*, pages 1055–1089, 1998. URL http://www.jstor.org/stable/2586974.

Daron Acemoglu and Jorn-Steffen Pischke. Beyond Becker: training in imperfect labour markets. *The Economic Journal*, 109(453):112–142, 1999a. URL http://onlinelibrary.wiley.com/doi/10.1111/1468-0297.00405/full.

Daron Acemoglu and Jorne-Steffen Pischke. Minimum wages and on-the-job training. *NBER Working Paper*, 1999b.

Daron Acemoglu and Jorne-Steffen Pischke. The structure of wages and investment in general training. *NBER Working Paper*, (6357), 1999c. URL http://www.nber.org/papers/w6357.pdf.

Joshua D. Angrist. Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. *Journal of business & economic statistics*, 19(1):2–28, 2001. URL http://amstat.tandfonline.com/doi/abs/10.1198/07350010152472571.

Joshua D. Angrist and Guido W. Imbens. Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity. *Journal of the American*

Statistical Association, 90(430):430–442, 1995. URL https://pdfs.semanticscholar.org/4727/b71f63a7fdfb482124755a118189e0d40d08.pdf.

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2008.

W Arulampalam and Alison L. Booth. Learning and earning: Do multiple training events pay? A decade of evidence from a cohort of young British men. *Economica*, 68:379–400, 2001.

David H. Autor. Why do temporary help firms provide free general skills training. *The Quarterly Journal of Economics*, 116(4):1409–1448, 2001. URL http://qje.oxfordjournals.org/content/116/4/1409.short.

David H. Autor, Frank Levy, and Richard J. Murnane. The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118 (4):1279–1333, 2003. URL http://qje.oxfordjournals.org/content/118/4/1279.short.

Gary S. Becker. Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5, Part 2):9–49, 1962. URL http://www.journals.uchicago.edu/doi/pdfplus/10.1086/258724.

Richard Blundell, Lorraine Dearden, and Costas Meghir. The Determinants and Effects of Work-Related Training in Britain. *IFS Working Paper*, 1996. URL http://www.ifs.org.uk/comms/r50.pdf.

Alison L. Booth. Job-related Formal Training: Who Receives It And What Is It Worth? *Oxford Bulletin of Economics and Statistics*, 53(3):281–294, 1991.

Pedro Carneiro, James J. Heckman, and Edward J. Vytlacil. Estimating marginal returns to education. *The American economic review*, 101(6):2754–2781, 2011. URL http://www.ingentaconnect.com/content/aea/aer/2011/00000101/00000006/art00016.

Lorraine Dearden, Howard Reed, and John Van Reenen. Who gains when workers train? Training and corporate productivity in a panel of British industries. 2000. URL http://discovery.ucl.ac.uk/4072/.

Lorraine Dearden, Howard Reed, and John Van Reenen. The impact of training on productivity and wages: Evidence from British panel data. *Oxford bulletin of economics and statistics*, 68(4):397–421, 2006. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0084.2006.00170.x/full.

de la prospective et de la performance Direction de l'évaluation. Repères et références statistiques: sur les enseignements, la formation et la recherche. Technical report, Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche, 2016. OCLC: 958852908.

Jean-Marie Dubois, Edmund Noack, and Jean-Claude Sigot. Dispositif d'enquêtes sur les formations et itinéraires des salariés - Premiers indicateurs du volet entreprises, April 2016. URL http://www.cereq.fr/content/download/16720/134826/file/netdoc156.pdf.

Christian Dustmann and Uta Schönberg. What Makes Firm-based Vocational Training Schemes Successful? The Role of Commitment. *American Economic Journal: Applied Economics*, 4(2):36–61, April 2012. ISSN 1945-7782, 1945-7790. doi: 10.1257/app.4.2.36. URL http://pubs.aeaweb.org/doi/10.1257/app.4.2.36.

The Economist. Lifelong learning. *The Economist*, page 9, January 2017.

Andrew Gelman. The problems with P-values are not just with P-values. *The American Statistician*, 2016. URL http://www.stat.columbia.edu/~gelman/research/published/asa_pvalues.pdf.

Dominique Goux and Eric Maurin. Returns to firm-provided training: evidence from French worker–firm matched data. *Labour economics*, 7(1):1–19, 2000. URL http://www.sciencedirect.com/science/article/pii/S0927537199000238.

Christine Greenhalgh. Adult vocational training and government policy in France and Britain. *Oxford Review of Economic Policy*, 15(1):97–113, 1999.

James Heckman and Salvador Navarro-Lozano. Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and statistics*, 86(1):30–57, 2004. URL http://www.mitpressjournals.org/doi/abs/10.1162/003465304323023660.

James J. Heckman and Edward Vytlacil. Policy-relevant treatment effects. *The American Economic Review*, 91(2):107–111, 2001. URL http://www.jstor.org/stable/2677742.

James J. Heckman and Edward Vytlacil. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–738, 2005. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0262.2005.00594.x/full.

James J. Heckman and Edward J. Vytlacil. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *edings of the National Academy of Sciences*, 96:4730–4734, 1999. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC16400/pdf/pq004730.pdf.

Guido W. Imbens and Joshua D. Angrist. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467, March 1994. ISSN 00129682. doi: 10.2307/2951620. URL http://www.jstor.org/stable/2951620?origin=crossref.

Anja Kuckulenz and Thomas Zwick. The impact of training on earnings: Differences between participant groups and training forms. Technical report, ZEW Discussion Papers, 2003. URL https://www.econstor.eu/handle/10419/23990.

LegiFrance. Relative à la formation professionnelle, à l'emploi et à la démocratie sociale, March 2014. URL https://www.legifrance.gouv.fr/eli/loi/2014/3/5/2014-288/jo/texte.

Edwin Leuven. A review of the wage returns to private sector training. *Unpublished paper*, 2004. URL http://www.academia.edu/download/42126026/34932279.pdf.

Mark A. Loewenstein and James R. Spletzer. Dividing the Costs and Returns to General Training. *Journal of Labor Economics*, 16(1):142–171, January 1998. ISSN 0734-306X, 1537-5307. doi: 10.1086/209885. URL http://www.journals.uchicago.edu/doi/10.1086/209885.

Lisa M. Lynch. Private-sector training and the earnings of young workers. *The American Economic Review*, 82(1):299–312, 1992. URL http://www.jstor.org/stable/2117617.

Jacob Mincer. Investment in human capital and personal income distribution. *Journal of political economy*, 66(4):281–302, 1958. URL http://www.journals.uchicago.edu/doi/pdfplus/10.1086/258055.

Daniel Parent. Employer-supported training in Canada and its impact on mobility and wages. *Empirical Economics*, 28(3):431–459, 2003. URL http://www.springerlink.com/index/W8M0E9XY7C6WQ02D.pdf.

Alfred C. Pigou. *Wealth and Welfare.* Macmillan and Company, limited, 1912. URL https://ia600201.us.archive.org/28/items/cu31924032613386/cu31924032613386.pdf.

P\a al Schøne. Why is the return to training so high? *Labour*, 18(3):363–378, 2004. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1121-7081.2004.00275.x/full.

Eric Verdier. Training and enterprise in France. *International Journal of Manpower*, 16 (5):38, 1994.