Labour Economics 89 (2024) 102573

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Everybody's got to learn sometime? A causal machine learning evaluation of training programmes for jobseekers in France

Héloïse Burlat[1,2]

Bureau of Theoretical and Applied Economics, University of Strasbourg ; Department of Economics, Ghent University.

ARTICLE INFO

JEL classification:
C21
J68
J08
J24

Keywords:
Policy evaluation
Active labour market policy
Continuing vocational training
Causal machine learning
Causal forest
Conditional average treatment effects

ABSTRACT

This paper estimates the heterogeneous impact of three types of vocational training- preparation, qualifying, and combined – on jobseekers' return to employment using the Modified Causal Forest method. Analysing data from 33,699 individuals over 24 months, it reveals a short-term negative lock-in effect for all programmes, persisting in the medium term for combined training. Only qualifying training shows a positive medium-term effect. Seniors, low-skilled, foreign-born, and those with poor job histories benefit most, while youth and higher education levels benefit less. Targeting foreign-born individuals could significantly enhance programme effectiveness, as indicated by the clustering analysis and optimal policy trees.

1. Introduction

On average in 2019, the unemployment rate in France amounted to 8.4% of the active population with a 20% labour shortage on company side (Grobon et al., 2021). Many in this unemployed population are not graduates and their unemployment rate is 15.5% for those with at middle school graduate level or under, compared to 5.1% for those with a higher education diploma. Moreover, these individuals have difficulty finding a job, because long-term unemployment affects 40.1% of the unemployed and 49.6% of middle school graduates and under (compared to 32.6% of those with a higher education qualification). Although companies are recruiting, a significant proportion of the population with few or no qualifications is unable to find a job and falls into long-term unemployment. This situation is part of a deterioration in the match between labour supply and demand characterized by a shift of the Beveridge curve towards the outside from 2015 to 2019 (Grobon et al., 2021). This can be explained in particular by a difficulty in recruiting in well-identified sectors such as construction, which has reported an inability to find suitable candidates for 75% of its available job offers (Grobon et al., 2021). One of the factors of recruitment difficulties is the lack of skills of jobseekers in relation to what is required by employers. In response to this, continuing vocational training (CVT) for jobseekers is one of the levers used by the public authorities to try to make the labour market match more fluid. The "*Plan d'Investissement dans les Compétences*" (*PIC*) was launched by the French government between 2019 and 2022 following other plans to implement a nationwide CVT policy for jobseekers relying on an investment nearing €15 billion. More details about the *PIC* are given in Section 3.

In France, several studies have been carried out to evaluate specific training programmes, such as the "*Projet personnalisé d'action*" (Fleuret, 2006), they measure entry-into-qualification training after orientation training. The only evaluation of CVT policy on return to work in France was made by Crepon et al. (2012) but they did not estimate the heterogeneity of the effect of CVT on individual characteristics

E-mail address: heloise.burlat@etu.unistra.fr.

[1] Address for correspondence: Héloïse Burlat BETA – Faculté de Sciences Economiques et de Gestion, 61 avenue de la Forêt Noire, 67000 Strasbourg, France

[2] The author thanks Amélie Barbier-Gauchard and Fabrice Gilles for helpful comments and corrections. I am also grateful for the participants in seminars and conferences at European Association of Labour Economists, the journées Louis-André Gérard-Varet, the AFSE annual congress, the Journées de la Microéconomie Appliquée and the Welfare and Policy conference. A special thank for the precious comments of Bart Cockx and Michael Lechner, as well as to the two anonymous reviewers. Access to some confidential data, on which is based this work, has been made possible within a secure environment offered by CASD – Centre d'accèes sécurisé aux données (Ref.10.34724/CASD and https://doi.org/10.34724/CASD.438.4633.V1)). I gratefully acknowledge financial support from the Grand-Est Region and permission to merge their data to the CASD data. The opinions expressed in this paper are those of the author and do not necessarily reflect the views of the Grand-Est Region.

https://doi.org/10.1016/j.labeco.2024.102573
Received 14 November 2023; Received in revised form 23 April 2024; Accepted 14 May 2024
Available online 24 May 2024

or different training types. Most studies use the Rubin model (Rubin, 1974) or the duration model of Abbring and Van Den Berg (2003), which are effective in controlling for selection bias but do not allow for the estimation of the heterogeneity of the treatment effect. However, it is essential to explore this area for policymakers in order to better identify their target audience and to set up training typologies that will allow them to achieve their goals.

To maintain the causal inference of previous methods while being able to estimate the heterogeneity of a treatment effect, Athey et al. (2019) have recently developed the generalized random forest method, modified by Lechner and Mareckova (2022) as the modified causal forest method. This machine learning technique estimates not only the Average Treatment Effect (*ATE*) but also the Individualized Average Treatment Effect (*IATE*), which, aggregated and balanced, constitutes the Balanced Grouped Average Treatment Effect (*BGATE*). The latter two effects constitute the Conditioned Average Treatment Effect (*CATE*). In the only current published application of this method (Cockx et al., 2023), the heterogeneity effect was found to be significant for the type of training, for the variable describing whether the individual is a migrant and for the variable describing the degree of employability. The causal forest method also makes it possible to simulate distributions of treated individuals with individual characteristics that differ from reality in order to optimize the effect of training on return to work and thus to propose new selection criteria at the start of training in order to maximize the desired purpose. The aim of this paper is to measure the effect of CVT policy, particularly with regard to the heterogeneity of individuals and training typologies. Which training is more beneficial to what kind of population? Because this policy is implemented operationally at the local level, we have limited the scope to the French Region "Grand-Est". Indeed, each Region defines the content and the title of its training programmes. In the absence of a nationwide harmonization of the titles and goals of training actions, estimating the national effect of CVT is too challenging. Moreover, the Grand-Est Region funded training for jobseekers increases the skills of jobseekers and by extension their human capital. The Region does not finance orientation services or short job search assistance services as other structures such as *Pôle Emploi* may offer. As such, in addition to studying the heterogeneity of the training effect, this paper will differ from the existing literature by focusing on training that only increases human capital and differentiating types of training by purpose instead of duration in a multiple treatment context.

Here we use a unique database, wherein the FORCE database set up in 2020 is matched with the information system of the CVT of the French Region "Grand-Est" (see Section 4.1 for more information about the database). Due to the merger of the former regions Alsace, Champagne-Ardenne and Lorraine, the regional training programme was only unified in 2018. We will then consider individuals who entered a period of unemployment from 1st January 2018 until 31 December 2019 and, for the treated group, individuals who started a training programme financed by the Region during the first 6 months of their period of unemployment. We will then follow the professional trajectories of 33,699 individuals who lived in the Grand-Est region for 24 months after the beginning of their unemployment spell, covering the period from January 2018 to December 2019. This database contains a large number of individual characteristics (age, gender, level of qualification, industry, employment zone, employment and unemployment history, characteristics of the unemployment spell under study such as duration of compensation or reason for entering unemployment) and a distinction between training typologies. We have chosen to separate them into three main categories within the framework of multi-processing: first, preparation training, aimed at refreshing basic skills and consolidating a professional project; second, qualifying training, aimed at validating the skills necessary to perform a defined job and in some cases to obtain a diploma; third, combined training, a combination of preparation training and qualifying training.

The results shows that there is heterogeneity of the effect between the different types of training. There is a short-term lock-in effect for all types of training, but this effect becomes positive in the medium-term for qualifying training, whereas it remains negative for combined training. The effect of training programmes is also heterogeneous across individual characteristics : older, less-skilled, those with less employment experience and foreign-born individuals experience more significant benefits. Conversely, the advantages are less prominent for youths and those with higher levels of education. Increasing the rate of minimal social income recipients and foreign-born individuals in training entrants would significantly increase the effect of all training programmes on the return to employment. Increasing the proportion of RSA recipients and foreign-born individuals among training entrants would significantly increase the effectiveness of all training programmes. For the average effect of qualifying training, our results are in line with the literature (Card et al., 2018; Crepon et al., 2012) but our paper contributes to the literature by providing a new interpretation regarding general skills training or training that combines general and specific skills. In addition, the results on youths diverge from Crepon et al. (2012), and to our knowledge, no previous investigation into the heterogeneity of the effect of training programmes on all the features we examined had been conducted before.

The paper is structured as follows. Section 2 is the literature review, Section 3 the institutional context of CVT for jobseekers and Section 4 the data and sampling. Section 5 explains the method used and the parameters of interest. Section 6 presents results and Section 7 concludes and discusses the results.

2. Literature review

The history of CVT policy evaluation spans over 50 years, beginning with the paper of Ashenfelter (1978). Indeed, in order to reduce structural unemployment, many OECD countries have a long track record of implementing Active Labour Market Policies (AMLP). Theoretically, AMLP policies aim to encourage the unemployed to return to work and to stimulate wage growth by improving the match between labour demand and supply. One of the founding papers in this literature is by Heckman et al. (1999). It summarizes the various AMLPs, the methods used to evaluate them and their results. Using a model introduced by Layard and Nickell (1986), Calmfors (1995) explained the effects of ALMPs on equilibrium employment. In this model, AMLPs have an effect on employment and wage-setting schedules through the matching function. Firms find better candidates more quickly to fill their vacancies, which costs them less, so they no longer need to offer attractive wages (Layard and Nickell, 1986). Calmfors and Lang (1995) explain that the result is an increase in the number of vacancies, and therefore in labour demand. An expansion of AMLP programmes should increase labour-force participation and as a result have a positive effect on employment. However, contrary to the previous effect on wages, there could be an incentive to raise wages because of a reduced risk of lay-offs and the higher instantaneous utility compared to open unemployment. Targeting people who are less likely to find a job, for example long-term jobseekers, is more likely to reduce wage pressure but implies more competition for the newly laid-off insiders. According to the theory, AMLPs should increase the probability of being in employment, especially if they target long-term jobseekers, but they have an ambiguous effect on wages.

Offering an overview of years of microeconomics studies and assessing the convergence of estimated effects, Card et al. (2018) published a meta-analysis of a 200 international studies on the evaluation of active labour market policies for jobseekers. It shows an overall non-significant or negative effect of programmes on return to employment in the short-term, but a positive overall effect in the medium and long-term. This non-significant or negative effect of programmes in the short-term is called a lock-in effect (Card et al., 2018). It means that trainees, while they are in training, do not have the possibility and the time to find a job because trainings are usually full-time. Meanwhile, jobseekers without training experience find themselves in better conditions on the job market.

However, the majority of these studies included in the meta-analysis of Card et al. (2018) do not specifically measure the effects of CVT policy but the whole range of active labour market policies. The authors reference 42 evaluated programmes for France and 253 programmes for Germany. A key explanation lies in the difficulty of evidencing individual professional trajectories because of the absence of a database matching UI (Unemployment Insurance), employment declarations and continuing vocational training databases. Crepon et al. (2012) evaluated public-sponsored training programmes for jobseekers in France at a national level between July 2001 and December 2005. They use the timing-of-event model of Abbring and Van Den Berg (2003) to estimate the CVT effect on the transition rate from unemployment to employment. Their results indicate that CVT does not reduce the time to return to work but increases the time in employment after training. However, their database only allows for the estimation of the heterogeneity of the effect of CVT on a limited number of individual characteristics and does not take into account different training types by setting up multiple treatments apart from the duration of the training. Examining data on West Germany between January 1992 and June 1994, the study by Lechner et al. (2011) uses a modified propensity score matching estimators for multiple treatments analysis that differentiates between several types of training. The results are significant for training heterogeneity but, for individual heterogeneity, restrictions on the estimation method make the sample too small for the population subsets to yield significant results.

Based on the literature, we propose three hypotheses concerning the time dependence of results, the heterogeneity of trainings and the heterogeneity of individuals.

First, about time dependence, a large majority of studies concur in finding that CVT policy has a negative effect in the short-term due to the lock-in effect but a positive effect in the medium and long-term (Card et al., 2018). However, in the medium-term after completion of the training programme, the effect of CVT policy can be ambiguous. Training may increase the individual's reservation wage and thus slow down the process of finding a job. On the contrary, individuals who trained in an industry characterized by high recruitment pressure may find a job very quickly (Calmfors and Lang, 1995). Our first assumption is that there will be a short-term negative effect linked to the lock-in effect, the duration of which will be proportional to that of the training. The effect of training, particularly industry-specific training, should accordingly be positive in the long-term, as training programmes are based on the local recruitment needs of firms. This point is developed in the following section on the institutional context.

The second aspect is the difference between the types of training. The studies by Crepon et al. (2012) and Lechner et al. (2011) also show that the longer the training courses, and therefore the more qualifying, the more positive the effect on the return to work in the long-term. Becker (1994) defines two categories of training: general training and specific training. Specific training increases the marginal productivity of trainees only for a specific firm, or industry in our case, when general training increases the marginal productivity of trainees for all firms. Because it is difficult for firms to find employees with specific skills related to their activity, they offer them a better salary and have fewer incentives to lay them off than employees with general skills. Our second hypothesis is that training courses might usefully be differentiated by type rather than by duration. We therefore consider three types of training based on Becker (1994): preparatory or general training, qualifying or specific training and training that is a combination of preparatory and qualifying training. According to Becker (1994), trainees in qualifying training should return to work more quickly and have a greater probability of being on a permanent contract than trainees in preparatory training. The total duration of training for those taking a combination of the two courses is much higher than for the other two, especially given the waiting time between the two training sessions. As a result, it is possible that the lock-in effect is higher for them than for the others and that, over a long period, it overshadows the effect of the increase in specific and general human capital.

The third and most important focus concerns the heterogeneity of the effect of CVT on return to work according to individual characteristics. Previous studies only took into account a few general variables such as age or level of qualification essentially by making sub-groups (Lechner et al., 2011; Crepon et al., 2012; Card et al., 2018). Usual estimation methods cannot incorporate a vector of high dimensional individual characteristics or conclude a causal inference from the training on a particular characteristic, which makes it difficult to study heterogeneity. Crepon et al. (2012) have integrated individual characteristics covariates in estimated parameters of transition rate and have made subgroups to estimate the training effect on transition rate. They concluded that CVT is particularly beneficial for young people and people with lower education levels. Having a rich database, particularly in terms of labour market history, is crucial to obtaining the most accurate, unbiased estimators possible (Biewen et al., 2014). However, taking into account unobservable individual characteristics to reduce selection bias or improve estimation does not seem to have a significant impact on the result. Indeed, using experimental or administrative data leads to the same results (Card et al., 2018); the inclusion of psycho-social issues likewise does not affect the results (Caliendo et al., 2017). Our third assumption is that training has a greater effect on the probability of being in employment for young people and those with a low level of qualification. Proxies for the individual's labour market history should also have a significant impact on the effect of training.

3. The institutional context

Introduced in France during a period of full employment by the law of 16 July 1971, continuing vocational training is to be distinguished from initial training. Continuing vocational training (CVT) targets people who have completed their initial studies and trains them in skills specific to an occupation. However, this definition needs to be qualified in practice, since the Region also offers non-degree training, aimed at consolidating the individual's professional project or at acquiring basic skills.

The French employment agency, *Pôle Emploi*, and the Regions are the main actors in the local implementation of CVT for all jobseekers, regardless of their status. Its goals are to re-mobilize jobseekers, raise their level of qualification and integrate them into the labour market. The target group of these CVT funders is in particular those who are furthest from employment, i.e. those with a low level of qualification (middle school level or below) and young people aged 15 to 29 who are NEET (neither in employment, nor in education, nor in training).

As such, the French Region Grand-Est is an important local actor in this policy and uses it as a lever against structural unemployment in its territory. It has financed the training of an average of 30,000 jobseekers per year since the beginning of the *PIC Plan d'Investissement dans les Compétences in french, Skills Investment Plan in English*). The structural offer is a biannual call for tenders based on feedback on local labour needs. The cyclical offer is set up on an ad-hoc basis on the request of one or more companies, for specific programmes (illiteracy for example) or to support major projects.

The French Region Grand-Est finances three main types of training with different purpose. The first one is qualifying training, geared towards the acquisition of technical skills and their validation with a diploma and professional integration. The second one is preparation training, which aims to re-mobilize those who are the furthest from employment and enable them to continue in downstream training (with no obligation). The last one is language training, designed to teach language skills that are linked to the individual's professional project.

In financial terms, the CVT policy is a major item of expenditure for the Region, intensified since the arrival of the *PIC* and its regional version the *PACTE*. Indeed, the dedicated regional envelope amount to 1.2 billion euros between 2019 and 2022, shared between its own funds and the *PACTE*. This is one of the Region's largest expenditure items. The *PIC* was launched by the French government between 2019 and 2022 to complement the CVT policies for jobseekers already implemented by the Regions and the French employment agency for the purpose of ramping up the response to structural unemployment. The *PIC* is made of national programmes and regional *PACTEs* negotiated between the French government and the Regions, in association with the social partners, for a budget of almost 15 billion euros. It aims to train one million jobseekers with few or no qualifications and one million young people who are isolated from the labour market. It is the first national plan with such a budgetary envelope on this subject.

4. Data and sampling

4.1. A unique dataset

In this study we use a unique and particularly rich database: the FORCE system produced by the DARES (Ministry of Labour, Full Employment and Inclusion in France), which creates a unique identifier per individual for the databases of the Historical Job Search Files from the French employment agency (FH) and the Job Contracts Base (MMO). We have matched this system with the CVT database of the Region (Athena), which gives us more exhaustive information on the type and sector of training, but and the precise dates of entry and exit of the individual in training. This information is detailed in Online Appendix A. We have used the latest wave of matching available at the time of our study, i.e. wave 10.

This database makes it possible to trace histories of unemployment over 10 years and employment over five years. We have individual trajectories with periods of unemployment, employment and training at national level and more detailed information on individuals trained in programmes financed by the Region. The combined data from these three sources provide exhaustive individual characteristics, particularly concerning unemployment and employment history, which are both control variables for selection for training entry, and outcome variables (employment status, continuation in training, salary, etc.). The matching with the Athena database makes it possible to distinguish between the different types of training actions and thus to measure the heterogeneity of the programmes thanks to the large number of observations.

However, the MMO database does not include self-employed, cross-border workers and only began including public employees in 2022. As a result, the return to work of treated and controlled individuals might be slightly underestimated. Cross-border workers are located in employment areas identified by borders. A cross-border feature was assigned to individuals living in the relevant labour market areas, i.e., less than 100 kilometres from Basel (Switzerland), Luxembourg (Luxembourg), Freiburg im Breisgau (Germany), Karlsruhe (Germany) or Saarbrucken (Germany).

4.2. Population under study

Our population of interest is that of individuals who live in the Grand-Est region registered as category A jobseekers (immediately available for employment and without a part-time job) with the French employment agency between 1st January 2018 and 31 December 2019. We chose this period in order to be able to follow as many individuals as possible for 24 months, which is a sufficient timeframe to measure short and medium-term effects. Furthermore, the Regional Training Plans of the three former regions Alsace, Lorraine and Champagne-Ardenne were only unified in an operational manner in the form of a French "Grand-Est" Regional Training Plan in 2018. To avoid including individuals nearing retirement or fresh out of school, we restricted the scope of the study to individuals aged 20–55. The set of individual characteristics is fixed at the beginning of the unemployment period under consideration and does not change over time.

Similarly, we also removed individuals who attended a training programme during the nine months before the beginning of the unemployment period under consideration. We consider the individuals treated as having started a training programme financed by the Region during the first 6 months of the unemployment spell considered. Trainees who did not complete the training were removed. We created six cohorts based on the number of months elapsed between the start of the unemployment spell and the start of the training.

Concerning the individuals in the control group, we removed those who followed a training programme during the 24 months after the beginning of their considered unemployment spell. The treatment group is composed of 7,527 individuals and the control group of more than 100,000 individuals. However, to prevent excessive computation time as well as to form cohorts, we reduced the control group to 26,172 individuals by selecting the 5th-nearest neighbours to each treated individual according to their propensity score as well as the year and month in which their unemployment spell began. The number of months elapsed between the beginning of the unemployment spell and the start of the training is then assigned to the control individual who was matched with the trainee. Creating six cohorts based on month of entry into training allowed us to conduct a clustered analysis. Studying these six groups simultaneously may have led to a diffuse effect over time, potentially resulting in an overestimation or underestimation of the effect due to temporal differences among treated individuals. By doing this, we separated the estimates for different cohorts so that the t_0 of our study is the start date of training for all individuals. This was the method used by Sianesi (2004) and Biewen et al. (2014) to take dynamic matching into account. However, we excluded individuals who subsequently attended a training from the control group, so that it would only include individuals who did not enter in a training programme at all.

We also exclude controls who found a job before the individual matched with them entered training. In the French system, there is no selection criteria or obligation to undergo training, to continue receiving benefits for example. Should certain unemployed individuals choose to abstain from enrolment in training programmes, their decision may stem from a scepticism regarding the efficacy of such training or, alternatively, from obtaining employment prior to the commencement of the training period. It is imperative to underscore that the structure and implementation of the CVT policy do not impart a selection bias in the enrolment process for individuals participating in training initiatives.

4.3. Training types

We have chosen to carry out a multiple-treatments modelling, i.e. the treatment variable is not binary but has four different values. Indeed, individuals can either not be treated, or have received a preparation training, a qualifying training or a combination of preparation and qualifying. For the last two types, we only consider the first training programme for each individual. We named the third type of training "combined". A total of 7,527 individuals in our sample entered training and, taking all programmes together, training programmes lasted an average of 179 days, with a standard deviation of 117 days.

The preparation type includes the training typologies Basic Skills, Goal, Orientation and Refresher Courses. These programmes are aimed to provide basic skills such as French, mathematics and interpersonal skills required for the individual to carry out their professional project, but also to build an action plan that includes further training, with work placements in companies and immersions in training centres leading to a qualification. The first purpose of this preparation type of training is therefore not to find a job but to pursue training leading

Table 1
Descriptive statistics for outcome variables and features.

Features	Without training (N = 26,172)	Preparation (N = 2,323)	Qualifying (N = 4,452)	Combined (N = 752)
Average number of months in employment after the start of the unemployment spell				
First 6 months	0.44 (1.32)	0.27 (0.86)	0.27 (0.87)	0.03 (0.23)
First 12 months	1.32 (3.02)	1.12 (2.50)	1.54 (2.68)	0.30 (1.08)
First 24 months	4.28 (7.54)	4.04 (6.68)	6.75 (7.83)	3.48 (5.22)
Individualized characteristics				
Women	48%	56%	40%	48%
Education level				
Below high school diploma	49%	59%	41%	51%
High school diploma	27%	27%	31%	32%
BTEC	12%	7%	15%	9%
Bachelor's degree	7%	4%	8%	6%
Master's degree	5%	3%	5%	2%
RSA recipients	22%	26%	17%	27%
Born outside France	15%	24%	11%	21%
Age				
− 25 years	34%	44%	38%	34%
25–35 years	33%	25%	34%	31%
36–45 years	19%	16%	18%	21%
+ 45 years	14%	15%	10%	14%
Disability	5%	9%	4%	8%
Number of UI benefit months to claim				
0	74%	83%	70%	77%
1–3	5%	2%	3%	2%
4–6	5%	4%	7%	3%
7–12	7%	5%	11%	8%
12 +	9%	6%	9%	10%
Living in a priority neighbourhood	14%	19%	13%	18%
Living less than 100km from a border	62%	62%	58%	57%

Standard deviation are in brackets for continuous features.

to a qualification. However, there is no obligation for the individual to enrol in a qualifying training programme at the end of a preparation training programme. It is similar to the general training introduced by Becker (1994) and used in theory to differentiate between types of skills, general versus specific. These training programmes are short, lasting on average 108 days with a standard deviation of 63 days. 2,323 individuals entered a preparation training, i.e. 30% of the trainees.

The qualifying type includes the training typologies Professionalization and Qualification. These programmes are aimed at teaching technical skills, to be applied, if internships in companies are planned, in order to find a job in the activity sector concerned. It is similar to the specific training introduced by Becker (1994). These training courses are longer, lasting on average 185 days with a standard deviation of 94 days. 4,452 individuals entered a qualifying training, i.e. 60% of the trainees' population.

The combined type includes individuals who first signed in for a preparation training financed by the Region and continued with a qualifying training financed either by the Region or by the French employment agency less than 12 months after the start of the preparation training. This is supposed to be a comprehensive programme with general and basic skills. However, beyond the training periods, the waiting period between the two programmes can be long (Fleuret, 2006) even if the transition is supposed to be short. Indeed, this type of training lasts on average 364 days with a standard deviation of 154 days. 10% of our treated group entered a combined training, i.e. 752 individuals, which is our smallest treated group.

Types of foreign language training are not considered, since these are specific to a professional project to access employment in a cross-border country or to have a more international professional profile. Similarly, we did not include training courses to prepare for competitive examinations in the health and social sector, since they are designed to give entry into a school or a long training course, whereas our aim here is to measure the effect of CVT on return to work.

Table 1 shows the distribution of the main individual features for the three treatment types. "Features" is the name used in the causal machine learning literature for the variables.

Individuals entering preparation training have lower education levels and job experiences than those in the other groups. There is also a high proportion of young people under the age of 25 and people with a low education level in all types of training, which is unsurprising, as the CVT specifically focuses on them, although it is not a selection criterion. We also find a high proportion of RSA recipients, because their social minimum income can be conditioned on taking action, such as signing up for training, towards finding a job. All variables, as well as balance tests between the various treated groups and the control group, are listed in Online Appendix A.

5. Econometric method

5.1. The model

For the methodology, we rely on the work of Lechner and Mareckova (2022) and more particularly his modified causal forest estimator to estimate the effect of the treatments and its heterogeneity according to individuals' characteristics.

Only an individual's outcomes with the treatment they received were observed (not with all three possible treatments). To estimate the effect of a treatment, it was therefore necessary to define a counterfactual who did not receive the treatment. To do this, we defined two vectors of variables $\tilde{X}$ and Z, which contain the vector of individual characteristics for each individual. The first vector, Z, was used to correct for selection effects. Z allowed us to group individuals to estimate the heterogeneity of the effect by making subgroups. Here we choose to work only with discrete, time-fixed variables. The variables can be present in $\tilde{X}$ and Z and their set X is noted with $X = \{\tilde{X}, Z\}, \dim(X) = p$. W denotes a subgroup of features of X excluding Z.

We then sought to estimate three average causal effects :

– Individualized Average Treatment Effect:

$$IATE(d, d_0; x) = E(Y^d - Y^{d_0} | X = x) \tag{1}$$

– Balanced Grouped Average Treatment Effect:

$$BGATE(d, d_0; x) = E[E[Y^d - Y^{d_0} | Z = z, W = w]] \tag{2}$$

– Average Treatment Effect:

$$ATE(d, d_0) = E(Y^d - Y^{d_0}) = \int IATE(d, d_0; x) f_X(x) dx \tag{3}$$

The parameter d represents the treatment variable with $t \in 1, 2, 3$ where 1 equals preparation training, 2 qualifying training, 3 combined training and d_0 represents the non-treatment variable with $d_0 = 0$. z is a possible value of the variable Z.

The *IATE* represents the effect between a treatment d and non treatment d_0 for individuals with the same characteristics x whereas the *ATE* represents the overall effect between treatment d and d_0, without distinction of individual characteristics.

The *BGATE* represents the *ATE* for a particular group of individuals and measures the heterogeneity of the effect. The *BGATE* is the effect between treatment d and d_0 for a group of individuals with individual characteristics Z, balanced by individual characteristics W (X without Z, the feature of interest). *BGATEs* are aggregated *IATEs* for a value z of the selected feature Z and are estimated in subsamples for which balancing tests have been performed beforehand on the vector of variables W. Here, *BGATEs* are balanced for all features, except for having experienced past layoffs and having a subsidized job contract in the past. More details about *BGATE* are detailed in Bearth and Lechner (2024).

However, because we are in a context of dynamic matching where trainees can enter a training programme at any time in their unemployment spell, we clustered them according to the number of months between the start of their unemployment spell and the beginning of their training. This results in six clusters. Each tree of the causal forest only features observations for the same cluster including treated and control group. After estimating the *ATE* for each cluster, these estimations were aggregated to obtain a global estimate of the *ATE* for the entire dataset. The goal of this aggregation was to obtain an overall measure of the average treatment effect that considers variations specific to each cluster. The first observation date is common to all individuals and is the date on which they started their training. *BGATEs* and *IATEs* are also aggregated over clusters. All standards errors are computed by bootstrap with 199 replications.

The equation for the Clustered Average Treatment Effect is:

$$(C)ATE(d, d_0) = E\left(\frac{1}{K}\sum_{k=1}^{K} Y_k^d - \frac{1}{K}\sum_{k=1}^{K} Y_k^{d_0}\right) \tag{4}$$

The parameter k is the number of the cluster, K is the total number of clusters.

5.2. Identification

The usual Rubin identification assumptions apply in this context:

– Conditional Independence Assumption (CIA):

$$\left\{Y^0, Y^1, Y^2, Y^3\right\} \coprod D | X = x, \forall x \in \chi \tag{5}$$

This implies that X contains the set of variables that influence both the choice of treatment and the potential outcomes.

– Common Support Assumption (CSA):

$$0 < P(D = d, X = x) = p_d(x), \forall x \in \chi, \forall d \in \{0, 1, 2, 3\} \tag{6}$$

This implies that each individual has a counterfactual that has the same individual characteristics.

– Stable Unit Treatment Value Assumption (SUTVA):

$$Y_i^{(D_1, \ldots, D_i, \ldots, D_P)} = Y_i^{D_i} \tag{7}$$

With P the size of the population where $Y_i^{(D_1, \ldots, D_i, \ldots, D_P)}$ denotes the potential outcome of individual $i \in \{1, \ldots, i, \ldots, P\}$ individuals in this population receive treatments $(D_1, \ldots, D_i, \ldots, D_P)$. This implies that there is no spillover effect, i.e. one individual's treatment should not influence another's outcome.

To justify the CIA hypothesis, the database here is very rich in terms of individual characteristics, particularly employment and unemployment history. Indeed, since we are not in an experimental setting, there could be a selection bias, which is a distortion in study results caused by the non-random selection of individuals or samples, introducing differences between the selected group and the larger population. The richness of the database is crucial to ensure that there are no selection bias and many studies have already identified the most important covariates to control for (Heckman et al., 1997; Lechner and Wunsch, 2013). Although we lack qualitative information from the caseworker, psycho-social covariates do not appear to be an important source of bias reduction according to Caliendo et al. (2017). In the first place, it is imperative that the treated and control groups have identical covariates from the same sources to closely resemble an experimental evaluation (Heckman et al., 1997). Moreover, Lechner and Wunsch (2013) and Biewen et al. (2014) demonstrated the importance of variables pertaining to an individual's labour market history, notably by delineating the short-term circumstances in the labour market immediately prior to entering training. In our study, we have access to information about the duration and number of unemployment spells 5 years before the training and about employment spell at least 1 year before the training. We also control for UI benefit months to claim as a valuable proxy for labour market history, as used by Crepon et al. (2012), because individuals can benefit from it only if they have worked at least 4 months in the last 2 years. We include the industry of the last job, as in many other studies (Sianesi, 2004; Biewen et al., 2014; Lechner et al., 2011; Cockx et al., 2023). Regional information is also important (Lechner and Wunsch, 2013) and, in addition to having the local unemployment rate like Biewen et al. (2014), we include the local evolution of employment rate. This covariate is a valuable indicator related to the lockdown phases resulting from the Covid-19 pandemic. Cross-border variable is also included because of the specificity of the Grand-Est Region. As Cockx et al. (2023), we do not have data on previous earnings. However, since this is not an outcome variable and we have other proxy variables such as duration of previous employment, industry, age, and education level, this does not seem to be an important issue. Furthermore, the placebo test, which verifies the plausibility of the CIA hypothesis in addition to the richness of our database, does not indicate a violation of the CIA hypothesis (see Online Appendix B). We conclude that the CIA hypothesis is plausible.

The *CSA* is verified by the selection rules implemented for both the treated and control groups, pertaining to age, reason for entering unemployment, not having taken training in the 9 months preceding the unemployment spell, region of residence and at most 6 months to enter training from the start date of unemployment. We also applied the min/max common support rule and tested the sensitivity of the results of different common support rules (see Online Appendix B). The common support rule dropped 5% of the observations for the cumulative months of employment after 24 months. Finally, for the *SUTVA*, the number of individuals leaving training each year is not high enough to have an effect on the whole unemployed population.

5.3. Estimation

The causal machine learning estimation literature has recently progressed, especially with the paper of Athey et al. (2019). By combining the predictive capability of machine learning with microeconometric causal effect estimation techniques, it is possible to obtain more accurate estimators than before (Knaus et al., 2021). Furthermore, by using the causal forest method, it is possible to estimate *BGATEs* with a very large vector of individual characteristics X. The estimation of a finer-grained effect is more robust than if we had calculated the effect for subgroups ourselves because it is the aggregate of the *IATEs* and it is balanced by the vector X. In particular, we used the Modified Causal Forests estimator of Lechner and Mareckova (2022), which allows for multiple treatments. We use here the 0.4.1 version of the mcf package on Python. The causal forest estimator is based on dividing a sample whose individual characteristics become more and more homogeneous until a so-called "final leaf" is reached. The difference in outcome between the control group and the treated group on this final leaf is the effect of the treatment for a specific population. In addition, we set the algorithm to perform a balancing test at the start to obtain a sample that is as homogeneous as possible in terms of individual characteristics. To estimate this difference and thus measure *IATEs*, the authors propose to minimize the Mean Squared Error (*MSE*) and the Mean Correlated Error (*MCE*), i.e.:

$$\widehat{MSE}_{S_x}[\hat{\mu_d}(x)] = \frac{1}{N^d_{S_x}} \sum_{i=1}^{N} \underline{1}(x_i \in S_x)\underline{1}(d_i = d)[\hat{\mu_d}(x_i) - y_i]^2 \tag{8}$$

$$\begin{aligned}\widehat{MCE}(d, d_0, S_x) = {}& \frac{1}{N^{d_0}_{S_x} + N^d_{S_x}} \sum_{i=1}^{N} \underline{1}(x_i \in S_x)\left[\underline{1}(d_i = d) + \underline{1}(d_i = d_0)\right] \\ & \times \left[\hat{\mu}_d(x_i) - \widetilde{y}_{(i,d)}\right]\left[\hat{\mu}_{d_0}(x_i) - \widetilde{y}_{(i,d_0)}\right],\end{aligned} \tag{9}$$

$$\widetilde{y}_{(i,d)} = \begin{cases} y_i & \text{if } d_i = d, \\ y_{(i,d)} & \text{if } d_i \neq d. \end{cases}$$

With $N^d_{S_x}$ the number of observations having received treatment t or $N^{d_0}_{S_x}$ not having received any treatment in a certain leaf S_x which is defined by the values of the features x. $\hat{\mu_d}(x)$ is the estimator of the $IATE(d, d_0, x)$. If there are no observations with the same value of x for all treatments, we use the nearest neighbour technique.

The criterion defined to stop the splitting is crucial because of its potential instability since the estimator will also be different if the splits are not similar between two trees. To solve this, the algorithm generates a multitude of trees from different subsamples and conditions the splits on a random but defined number of features. Moreover, the starting sample is itself divided into two parts to allow for a "honest" approach: a first part which is only used for training to define the selection criterion for each feature and a second part which will calculate the value of the estimator from the parameters defined by the training sample. This random approach and the repetition make it possible to obtain a stable final leaf and to avoid an over-fit model. Furthermore, to avoid selection bias, the authors propose to integrate the propensity score in the penalty term. This will penalize divisions where the probabilities of entering training in the daughter leaves are similar and therefore favour divisions with a high heterogeneity of propensity scores.

The authors propose as penalty criteria:

$$\begin{aligned} penalty(x', x'') = \lambda \Bigg\{ 1 - \frac{1}{t} \sum_{d=0}^{D-1} [P(D = d|X \in leaf(x')) \\ - \; P(D = d|X \in leaf(x''))]^2 \Bigg\} \end{aligned} \tag{10}$$

With $leaf(x')$ and $leaf(x'')$ the value of the features in the daughter leaves that result from the division of a parent leaf, $P(D = d|X = x)$ the propensity score and λ the penalty function. The choice of the exact form of the penalty function λ is to be made among eight possible forms as indicated in the authors' paper. Here we choose to set λ as random switches between outcome and MSE-MCE criterion in combination with the penalty function. In addition, to avoid a bias linked to variations in labour market conditions over time, we programmed the algorithm to keep the feature concerning the year of entry into the unemployment spell in all splits. In general, all technical explanations of the modified causal forest method are given in their paper (Lechner and Mareckova, 2022).

6. Results

6.1. Average treatment effects with programmes heterogeneity

This section reports the main results. Because of the selection of the control group based on observed individual characteristics, the results on the overall population could be biased. To ensure the validity of our results, we compute the *ATE* and the policy tree selection criteria for a selection of the control group based on the simulation of their entry date, as Lechner et al. (2011) and Cockx et al. (2023), but with a random forest model. We also compute these results with a random selection of 20% of the control group. We randomly assign individuals from the control subgroup to each cohort and only retain those who were not employed before the cohort's entry month into the training. Thus, the selection process of control individuals and their assignment to a cohort is entirely random, and their distribution of individual characteristics is similar to that of the entire untreated population. Results are in Online Appendix B. These tests yield similar estimates of *ATEs* and their standard errors for the three ways to select the control group.

We start by considering the average population effects of the three types of training on the number of employment months after 6, 12 and 24 months after the start of the unemployment spell for 33,699 jobseekers. Then, we investigate *BGATEs* for previously selected features we compared to *ATEs*. More details about the *BGATEs* in the paper of Bearth and Lechner (2024). *BGATE* allows us to interpret the grouped effect all other things being equal, which informs us about the overall effectiveness of the different programmes and the dynamics of the effects.

Our study is situated within a timeline where the COVID-19 crisis occurred. Card et al. (2018) suggest that the impact of training programmes is countercyclical, and therefore more likely to be positive during recessions, although the explanation of this effect are not clearly identified. To mitigate the bias resulting from this shock, we first used the local rate of change in labour demand as a control variable, which is expected to be low during lockdown periods. Additionally, in estimating the results, we ensured that variables indicating the month and year of entry into unemployment remain consistent across all final leaves, ensuring that individuals being compared entered unemployment at the same time. Finally, during lockdown periods, most training sessions were postponed either because the courses did not offer an online option (requiring practical skills in real-life situations) or because trainees lacked access to online training (lack of computer, internet connection, or insufficient computer literacy). Therefore, we estimate that the effects of the COVID-19 crisis on our results are limited.

Table 2
Effect of the different training programmes on the number of employment months after the start of the unemployment spell compared to no training (*ATE*).

	Estimate of the *ATE*
After 6 months	
Preparation training	−0.171*** (0.049)
Qualifying training	−0.165*** (0.038)
Combined training	−0.430*** (0.026)
After 12 months	
Preparation training	−0.190 (0.139)
Qualifying training	0.127 (0.108)
Combined training	−1.053*** (0.130)
After 24 months	
Preparation training	0.183 (0.358)
Qualifying training	1.92*** (0.307)
Combined training	−0.669 (0.541)

All effects represent average treatment effect (*ATE*). Outcomes are in months. Standard deviation are in brackets. Symbols *, **, and *** denote significance tests, indicating p-values below 10%, 5%, and 1%, respectively.

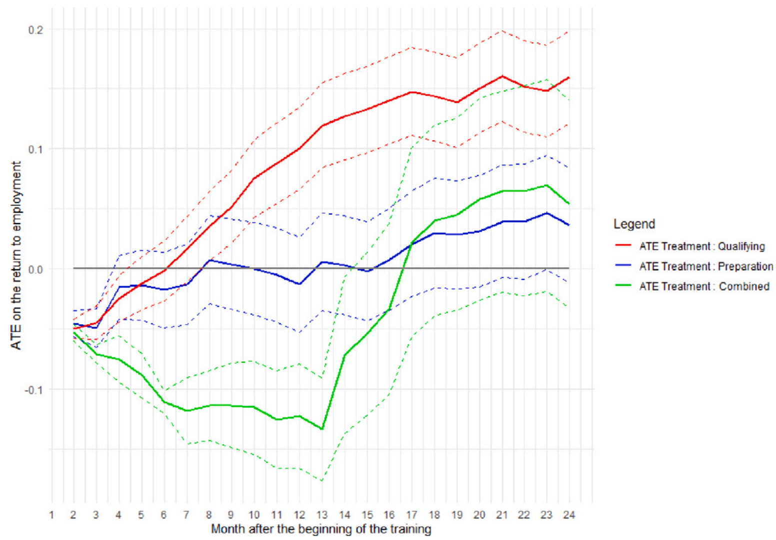

Fig. 1. *ATE* on the probability of being employed of preparation, qualifying and combined training.[3]

The results in Table 2 show that these three types of training do not have the same effect. Qualifying and preparation training both have a negative short-term effect, while combined training has a short- and medium-term lock-in effect. For qualifying training, the effect is positive on the number of months in employment 24 months after the start of the unemployment spell. If a jobseeker starts a qualifying training, they will have increased their number of employment months by 1.92 months at 24 months after the beginning of the training. To get an indication of the magnitude of the effect, according to our results at 24 months, qualifying training programme would increase the number of cumulative months of employment by 73% (4.58 - (4.58–1.92)/(4.58–1.92), with the average number of cumulative months in employment for the overall population as the final value and this value without the estimated effect for qualifying training as the initial value). This result is consistent with the literature (Card et al., 2018). The effect for preparation training and combined training is not significant 24 months after the start of the training. For the preparation training, it means that general skills do not improve return to employment. For combined training, this result can be explained by the fact that this type of training is longer than the others. The effect might be positive three or four years later. However, these results indicate that preparation training, despite being the shortest and thus having the smallest lock-in effect, has no positive impact on return to employment. As for combined training, even though it lasts the longest, the effect remains significantly negative even 24 months after the start of the first training. To put this result into context, the French employment agency considers an individual as a very long-term unemployed person after two years of inactivity.

These results in Fig. 1 are consistent with previous findings. Indeed, qualifying training has a negative effect until the 4th month and becomes positive at the 8th month. Following this, the effect remains positive and increases progressively over time, resulting in a 16 percentage point higher the probability of being employed for those who underwent qualifying training compared to individuals in the control group at the 24th month. For individuals in the preparation training, there is a negative effect until the third month; it then becomes non-significant. It is briefly positive at month 23, then returns to non-significance. For combined training, the effect is negative until month 14 and later non-significant. Given that preparation training lasts an average of 3 months and qualifying training is 6 months, such a long lock-in effect suggests a significant waiting period between the two training sessions.

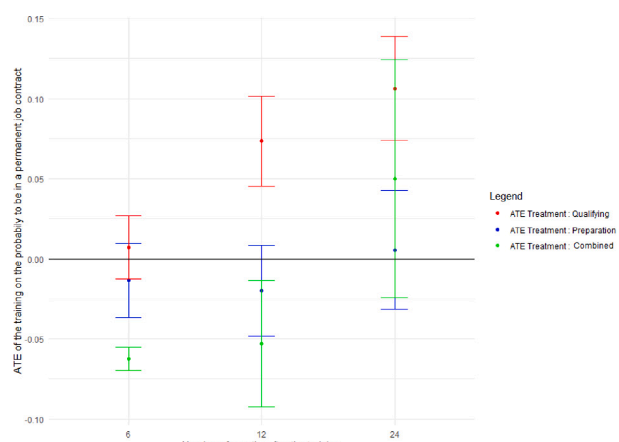

Fig. 2. *ATE* on the probability of having a permanent contract of preparation, qualifying and combined training at the 6th, 12th and 24th month after the beginning of the training with confidence intervals.[4]

The effects are the same when looking at the probability of having a permanent contract at the 6th, 12th and 24th month after the start of the training in Fig. 2, except that there is no short-term negative effect for qualifying and preparation training. This suggests that individuals who did not undergo training were more likely to find a job in the six months following the start of training than those who were trained, but that these jobs were not durable. Qualifying training increases by 11 percentage point higher the probability of being in a permanent job

[3] All effects represent average treatment effect (*ATE*). Outcomes are the probability to be employed each month after the beginning of the training. The dashed lines represent the confidence intervals, computed with a p-value of 5%.

[4] All effects represent average treatment effect (*ATE*). Outcomes are the probability to be in a permanent job contract at 6, 12 or 24 months after the beginning of the training. Confidence intervals are computed with a p-value of 5%.

Table 3
BGATEs of selected features on the cumulative number of employment months after 6, 12 and 24 months.

	Preparation training		Qualifying training		Combined training	
Features	*BGATE*	*BGATE-ATE*	*BGATE*	*BGATE-ATE*	*BGATE*	*BGATE-ATE*
Cumulative number of months of employment at 6 months						
Youth	−0.214*** (0.051)	−0.055*** (0.016)	−0.176*** (0.044)	−0.021 (0.014)	−0.473*** (0.026)	−0.059*** (0.007)
Senior	−0.127*** (0.047)	0.032*** (0.010)	−0.135*** (0.036)	0.0198** (0.008)	−0.376*** (0.025)	0.038*** (0.004)
Level of education below high school	−0.158*** (0.048)	0.031 (0.022)	−0.125*** (0.040)	0.073*** (0.017)	−0.403*** (0.027)	0.056*** (0.011)
Master's degree	−0.212*** (0.056)	−0.022** (0.010)	−0.233*** (0.040)	−0.035*** (0.008)	−0.482*** (0.025)	−0.023*** (0.005)
RSA recipient	−0.134*** (0.047)	0.023*** (0.007)	−0.141*** (0.035)	0.013*** (0.005)	−0.373*** (0.034)	0.035*** (0.006)
No UI	−0.185 *** (0.048)	−0.017 (0.025)	−0.165*** (0.039)	0.010 (0.016)	−0.435*** (0.026)	0.015 (0.009)
Highest number of UI months	−0.120* (0.063)	0.047** (0.021)	−0.149*** (0.037)	0.026*** (0.009)	−0.384*** (0.021)	0.066*** (0.005)
Born outside France	−0.155*** (0.047)	0.010** (0.004)	−0.142*** (0.038)	0.012*** (0.003)	−0.403*** (0.027)	0.015*** (0.002)
Women	−0.172*** (0.047)	0.001 (0.009)	−0.168*** (0.037)	−0.004 (0.007)	−0.423*** (0.023)	0.009* (0.005)
Cumulative number of months of employment at 12 months						
Youth	−0.254* (0.137)	−0.062* (0.033)	0.104 (0.114)	−0.030 (0.028)	−1.142*** (0.142)	−0.111** (0.044)
Senior	−0.144 (0.140)	0.047 (0.030)	0.176 (0.109)	0.041* (0.022)	−0.938*** (0.111)	0.093*** (0.029)
Level of education below high school	−0.191 (0.133)	0.028 (0.054)	0.205* (0.112)	0.163*** (0.038)	−0.980*** (0.127)	0.154*** (0.054)
Master's degree	−0.249 (0.157)	−0.029 (0.028)	−0.027 (0.113)	−0.070*** (0.018)	−1.189*** (0.137)	−0.054* (0.029)
RSA recipient	−0.150 (0.132)	0.028 (0.018)	0.153 (0.105)	0.017 (0.012)	−0.952*** (0.126)	0.065*** (0.018)
No UI	−0.243* (0.134)	−0.105 (0.09361)	0.116 (0.111)	0.016 (0.055)	−1.068*** (0.138)	0.031 (0.082)
Highest number of UI months	−0.061 (0.202)	0.076 (0.053)	0.121 (0.123)	0.021 (0.025)	−1.016*** (0.149)	0.084* (0.045)
Born outside France	−0.16 (0.141)	0.018 (0.026)	0.193* (0.115)	0.0411 (0.025)	−0.919*** (0.132)	0.081*** (0.027)
Women	−0.197 (0.135)	0.004 (0.022)	0.133 (0.108)	0.011 (0.017)	−1.014*** (0.137)	0.046* (0.024)
Cumulative number of months of employment at 24 months						
Youth	−0.054 (0.367)	−0.285** (0.142)	1.902*** (0.326)	−0.046 (0.131)	−0.946 (0.632)	−0.347 (0.235)
Senior	0.386 (0.389)	0.155 (0.095)	1.985*** (0.327)	0.036 (0.081)	−0.363 (0.515)	0.235* (0.140)
Level of education below high school	0.224 (0.354)	0.141 (0.093)	2.026*** (0.309)	0.221*** (0.071)	−0.515 (0.561)	0.341 (0.114)
Master's degree	−0.044 (0.388)	−0.127* (0.068)	1.644*** (0.321)	−0.160*** (0.053)	−1.014* (0.530)	−0.157* (0.083)
RSA recipient	0.311 (0.356)	0.080 (0.083)	1.925*** (0.309)	−0.003 (0.069)	−0.514 (0.514)	0.104 (0.101)
No UI	0.105 (0.349)	−0.214 (0.202)	1.949*** (0.309)	0.075 (0.134)	−0.687 (0.563)	0.074 (0.232)
Highest number of UI months	0.472 (0.479)	0.151* (0.092)	1.896*** (0.347)	0.022 (0.059)	−0.693 (0.615)	0.068 (0.112)
Born outside France	0.262 (0.359)	0.044 (0.069)	2.077*** (0.332)	0.089 (0.078)	−0.252 (0.617)	0.250** (0.115)
Women	0.156 (0.352)	−0.015 (0.041)	1.993*** (0.312)	0.062 (0.038)	−0.615 (0.570)	0.069 (0.066)

All effects represent average treatment effect (*ATE*) or balanced grouped average treatment effect (*BGATE*). Outcomes are in months. Standard errors are presented in brackets and symbols *, **, and *** denote significance tests, indicating p-values below 10%, 5%, and 1%, respectively. Clustered standard errors are computed by bootstrap with 199 replications.

contract compare to those who did not enter a training at the 24th month after the start of the training.

6.2. *Balanced group average treatment effects with population heterogeneity*

Examining the comprehensive impact of training offers an initial insight into employment outcomes, but does not provide insights into the mechanisms that drive this effect in either an upward or downward direction. Additionally, public policymakers specifically target the youth and individuals with lower qualifications. Given the tailored design of training programmes for this demographic, we expect observing more favourable outcomes them.

We consider results for *BGATEs* we selected as relevant to this policy. Age, level of education and gender are the most common subpopulation studied by the literature (Crepon et al., 2012) but also RSA recipient, being born outside France and number of UI months as a proxy for employment history are selected here to explore other sources of heterogeneity in the effect. Indeed, a minimum number of months in employment is required for eligibility to UI and the number of UI months is positively correlated with the duration of the previous employment. Therefore, RSA recipients do not receive UI.

Less and less heterogeneity can be observed as time goes by, and this is consistent with the results of Cockx et al. (2023). Only six months after the start of training do most features show statistical significance in Table 3. The significance of the features and their sign is the same for all the training programmes, except for age and level of education.

Indeed, the three types of training are particularly beneficial for RSA recipients, but with an opposite effect on the number of UI months, where the training seems to be more beneficial for those who have the larger number of UI months, to the detriment of those with the smallest number of UI months. Preparatory training and combined training have a more beneficial effect for older people and a less beneficial effect for young people. The effect is also weaker for individuals with a high level of education, but qualifying training and combined training are particularly beneficial for those with a low level of qualification. The effect of training is also more significant for individuals born outside of France.

In the medium-term (Table 3), preparatory training no longer shows heterogeneity in its effect concerning the features we have selected. Qualifying training also eliminates heterogeneity in the medium-term, but the effect on education levels persists in the medium-term, with a particularly positive effect for lower education levels and a negative effect for higher education levels. Both types of training maintain significance and the sign of the same variables in the medium-term before no longer having significant variables at the 24th month.

Apart from the medium-term heterogeneity in the effect of qualifying training based on education levels, our results suggest that differentiated effects based on the selected features only appear shortly

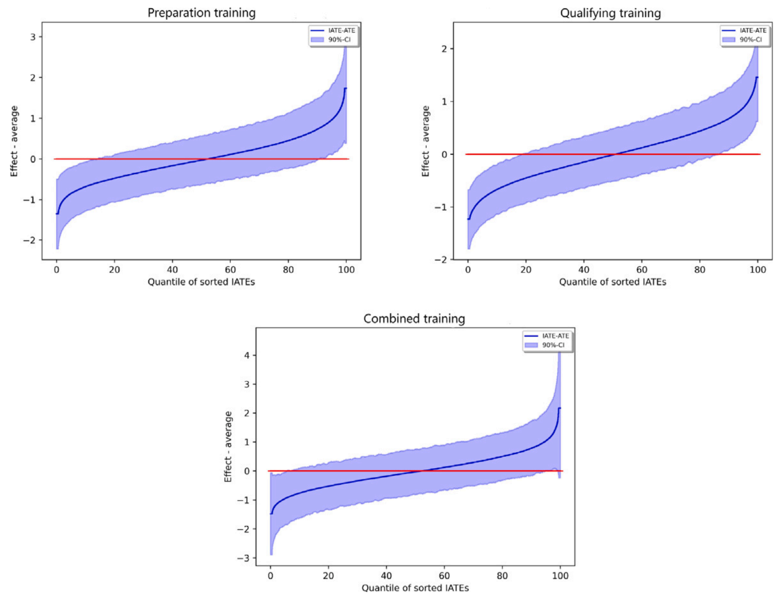

Fig. 3. *IATEs* on the cumulative number of employment months 24 months after the training for respectively, from top to bottom: preparation training vs no training, qualifying training vs no training, combined training vs no training.

after exiting training, especially since both types of training have a longer duration than the others. When differentiation occurs, it favours individuals facing more challenges in finding employment, such as older and less qualified people, RSA recipients, and those born outside France. Conversely, the effect of training is less beneficial for those with a low number of months on UI, although it is not significantly different for those receiving no UI at all, as for RSA recipients. Preparation and combined types of training are also less beneficial for the younger population, even though they are among the target populations.

6.3. Individualized average treatment effects and clustering

The previous section outlined the variations in the impact of training programmes on employment based on specific individual characteristics. Following these results, we now analyse the Individualized Average Treatment Effect (*IATE*) to examine the nuanced impact of our intervention on the cumulative number of months in employment. *IATEs* afford us the opportunity to explore how each type of training uniquely shapes the trajectory of each individual, accounting for their specific characteristics. This methodological approach facilitates a granular understanding of variations in short and medium-term effects, offering an individualized perspective on post-training employment. Following a comprehensive analysis of the *IATEs*, the inquiry turns to a detailed examination of the composition in terms of individual characteristics within three discrete clusters. These clusters span from those exhibiting the fewest benefits to those experiencing the greatest gains from the training interventions. This investigation aims to provide preliminary insights into the optimal profile of trainees, for the purpose of maximizing the impact of training programmes on employment outcomes. By discerning the characteristics associated with the most pronounced effects, this study seeks to provide valuable guidance for the refinement and customization of training initiatives to enhance their overall effectiveness.

Fig. 3 shows the value of the difference between the *IATEs* and *ATEs* as well as the confidence interval of this difference for each training programme on the number of cumulative months of employment 24 months after the start of the training. The significance of this difference indicates the heterogeneity of the effect. We observe heterogeneity of the effect for the preparation and qualifying training programmes since, respectively, 23% and 31% of their *IATEs* are significantly different from their *ATE* at the 10% threshold. This is less the case for the combined training programmes, where 9% of the *IATEs* are statistically different from the *ATE*. These differences are rather unfavourable to the overall effect, as 18% of the *IATEs* for the qualifying training are significantly lower than the overall effect. The percentage is 14% for the preparation training and 6% for the combined training. These results show a significant heterogeneity of the effect in the medium term for the preparation and qualifying training programmes, but low for the combined training programmes. The fact that a large part of this heterogeneity of the effect is negative suggests that, to maximize the effect of training programmes, certain individuals should not undergo training. It seems important, therefore, to determine the characteristics of the groups of individuals benefiting most and least from the different training programmes and to establish selection criteria for entry into training.

Moreover, we use *IATEs* to compute k-means++ clustering in order to characterize the relationship between the effects and the features. The k-means++ clustering procedure enhances the standard k-means algorithm by improving the initial selection of centroids. In traditional k-means, initial centroids are chosen randomly, which can lead to suboptimal clustering results. K-means++ addresses this by employing a probabilistic method to select initial centroids that are more likely to be well-distributed across the dataset. It begins by choosing one centroid randomly and then selects subsequent centroids with a probability proportional to their squared distance from the nearest existing centroid. Once centroids are initialized, the standard k-means clustering process proceeds, iteratively assigning data points to the nearest centroid and updating centroids until convergence is achieved.

Results in Table 4 show the importance of the heterogeneity in the training effect 24 months after the beginning of the training, especially for the effects of preparation and combined training, which can be positive and negative, depending on the cluster. Number of past months employed, number of past months unemployed, age and number of months on UI are ordered categorical variables, with categories encompassing multiple values (refer to Online Appendix A for further details). Trainings are the most effective for people with less work experience, a low level of education (below high school graduation), a majority of people born outside France, a many RSA recipients and those close to zero months on UI. This profile delineates individuals facing the most difficulties in attaining employment. However, the situation is more

Table 4
Distribution of individual characteristics according to the three clusters.

Cluster number	1	2	3
Individualized average treatment effects (*IATEs*) for the comparison to no training (NT)			
Preparation training vs NT	−0.17	0.32	0.41
Qualifying training vs NT	1.70	1.95	2.19
Combined training vs NT	−1.29	−0.65	0.20
Features			
Number of past months employed	2.26	2.40	1.54
Number of past months unemployed and registered to UI	1.97	2.44	2.50
Local unemployment rate	8.56	8.62	8.46
Local evolution of employment rate	−0.01	−0.01	0.03
Education level	2.36	1.86	1.41
Gender	0.41	0.46	0.63
RSA recipient	0.09	0.25	0.36
Age	1.51	2.30	2.60
Born in France	0.96	0.87	0.62
Number of months on UI	1.55	1.95	1.57

nuanced for age, which is centred around the 25–35 bracket. Regarding the local unemployment rate and the evolution of the local employment rate, logically, the most favourable situation for trainees occurs when the unemployment rate is low and the employment rate is increasing. The difference in results between the clusters is quite significant: for qualifying training, the *IATE* for cluster 3 is 28% ((2,19–1,70)/1,70) higher than that for cluster 1.

However, even in the most favourable situation of the cluster 3, the effect of the preparation and combined training is very small as it represents not even an additional month of employment compared to those who did not undergo training. Hence, the composition of these two training programmes is not the primary cause of their unfavourable outcome, but rather their structure and the instructional content they offer.

6.4. Optimal policy tree

In this section, we focus on entry selection criteria that would maximize the cumulative number of months employed two years after entering the training. We use optimal policy trees to compute the selection criteria that would maximizes the outcome variable with capacity constraints on the maximum number of trainees per programme. The small sample size of individuals who underwent combined training prevents us from incorporating this type of training into these results. As potential selection criteria, we include age, education level, RSA recipient, number of UI months, born in France, women, and disability. We use the optimal policy tree algorithm implemented in the modified causal forest by Lechner and Mareckova (2022), which is very similar to the method used by Zhou et al. (2023). This method uses shallow trees with a defined number of nodes. Results are easier to interpret and useful for public decision-makers.

We consider three scenarios. First, we impose no constraints on the maximum number of individuals per training programme. This allows us to assess the theoretical potential gain if all unemployed individuals had the opportunity to enter a qualifying or preparation training. Next, we impose a constraint on the maximum size of training programmes that is similar to what we observe in our sample, i.e. 7% of the total sample for preparation training and 13% for qualifying training. Finally, we double the observed size of each training programme, i.e. 14% of the total sample for preparation training and 26% for qualifying training. This scenario is a simulation of the increase in the CVT policy budget.

Table 5 reports the performance gains of these simulated hypothetical allocations. Performance gains are the deviation between the allocated performance and the observed performance. The second set

Table 5
Policy trees share of individuals into programmes and performance gain.

Tree depth	Share of individuals allocated to programmes[a]		Performance gain for the overall sample
	Preparation	Qualifying	
Observed	7.10%	11.84%	
Random	20.04%	32.21%	−0.01
No constraints			
2	0.00%	100%	1.52
3	0.00%	100%	1.52
Constraints : observed size of each training programme			
2	0.00%	12.63%	0.09
3	0.00%	10.32%	0.04
Constraints : double the size of the observed sample for each training programme			
2	0%	13.55%	0.11
3	0%	24.20%	0.32

[a] Individuals who are not allocated to a training programme are allocated to the no training group.

of columns reports the share of individuals allocated in the two programmes by the algorithm. Individuals who are not allocated to a programme are allocated to the no training group. We then compute the performance gain of these allocations for the whole sample and for two level of depth. It means that for a depth value of two, the algorithm considers potentially different allocations of four strata of individual characteristics. Here we consider only trees of depth 2 and 3 to keep the results clear and useable for public decision-makers.

These results suggest that increasing the number of trainees would have a highly significant effect on the cumulative number of months of employment 2 years after starting training for the whole sample. However, according to the algorithm, it is more efficient to allocate all individuals in the sample to qualifying training and none to preparation training with or without capacity constraints. It means that preparation training programme is not effective in improving the return to employment compared to qualifying training programme. If policymakers were to double the number of entrants into the training programmes, it would significantly increase the performance gain, especially in the case of the tree of depth 3. This result suggests that allocating a higher budget to the CVT policy would require particular attention in establishing selection criteria to achieve the strongest possible effect on return to employment. The performance gain for the random allocation is very small and demonstrates that caseworkers do not take into account our results on heterogeneity of effect for selecting entrants to training programmes. With more information, they could significantly improve the training effect on the employment by selecting individuals who benefit most from training programmes.

Table 6 is consistent with our previous results. When constraining the maximum number of trainees in the training programmes to those observed in our sample, the policy tree assigns individuals only to the qualifying training. It prioritizes individuals born abroad, with a lower level of education and those with limited recent work experience. This suggests that the current targeting of individuals with lower education level is effective, but considering the employment history and individuals born abroad would enhance this effectiveness. If return to employment is the only outcome variable considered by policymakers, qualifying training should also be favoured over preparation training. If the CVT policy budget is to be increased, our results argue in favour of selecting either less-educated young people born abroad, or more educated foreign-born individuals with more recent substantial employment history, or less educated individuals over 30 with no recent employment history. Once again, these results show that youths do not benefit more from training programmes than others, and targeting them is not effective with the current design of training programmes.

Table 6
Assignment rules of policy trees[a].

Tree depth	Preparation training	Qualifying training
Constraints of the observed size of each training programme		
2	Nobody	Foreign-born High school education level or below
3	Nobody	Foreign-born High school education level or below No UI
Constraints of the double size of the observed sample for each training programme		
2	Nobody	3 months of UI or less Foreign-born
3	Nobody	Below high school education level 35 or younger Foreign-born
	Nobody	Below high school education level Older than 35 No UI
	Nobody	Above high school education level More than 3 UI months Foreign-born

[a] Individuals who are not allocated to a training programme are allocated to the no training group.

Regardless of the maximum capacity of the programmes, individuals born abroad, those with a low level of education, and those with limited recent work experience still appear to be the ones who benefit the most from the CVT policy.

These recommendations are not intended to be taken literally, but rather to give a general idea of the current effectiveness of selection on entry to training, the characteristics to be taken into account to maximize the effect of training programmes on the return to employment, the optimal distribution of training programmes between them and the impact of an increase or reduction in the budget on the effectiveness of the policy and on any changes to be made to the selection criteria. Moreover, results of the policy tree for two other selection methods of the control group, including the one that selects 20% of individuals from the control group randomly and also assigns them to a cohort randomly, in Online Appendix B are similar regarding the assignment rules of directing individuals exclusively towards qualifying training, with the selection criterion being foreign-born individuals but they do not include the level of education as a selection criterion. Therefore, it is evident that caseworkers should prioritize the selection of individuals born abroad to enter qualification training in order to maximize their effects on the return to employment. However, regarding the second selection criterion, given that it may vary depending on the method of selecting control individuals, it becomes challenging to conclude on the characteristic to consider.

7. Sensitivity analysis

7.1. Tuning parameters

To ensure that the results are not sensitive to the parameters assigned in the modelling, we have conducted a sensitivity analysis by modifying key aspects.

First, we varied the value of λ, which dictates how the parent leaf is split into two daughter leaves. We compare results when we applied the MSE-MCE criterion or the rule maximizing effect heterogeneity, as proposed by Wager and Athey (2018). Following this, we assess the sensitivity of results to changes in the common support rule. We first test the absence of a common support rule and subsequently assign different quantiles to it.

Based on these results, there is no evidence of result sensitivity to variations in these parameters. The corresponding results are provided in Online Appendix B.

7.2. Placebo test

To ensure accurate measurement of the treatment effect in accordance to the CIA, we have conducted a placebo test as proposed by Imbens and Wooldridge (2009). We considered a period of unemployment preceding the one under consideration, allowing us to observe the professional situation of individuals for at least 12 months without interference from the specified unemployment period. Also, our database containing information on employment contracts starts in January 2017. As a result, we could only consider 1951 observations, which represent 6% of the total sample. We did not observe any significant effect of different training programmes during this time and concluded that there is no placebo effect, and that the CIA is plausible. The results of this test can be found in Online Appendix B.

7.3. Comparison to nearest neighbour matching

The Modified Causal Forest is a brand-new method and comparison with other methods is necessary to ensure the accuracy of our results. We provide here a comparison with the nearest neighbour matching method (NNM) with the 10 nearest neighbours for the number of cumulative employed months after 6, 12 and 24 months after the training. These results are in Online Appendix B and yield similar estimates of the *ATEs* and standard errors.

8. Conclusion

In this paper we used causal machine learning to investigate the average and heterogeneous effects of training programmes supported by the Grand-Est Region, using administrative individual data from the French Ministry of Labour, Full Employment and Inclusion and the Grand-Est Region. This was based on a database detailing the professional trajectories of 33,699 jobseekers living in the Grand-Est region during the 24 months after their unemployment spell. We found that on average all training programmes have a lock-in effect on the short term. Qualifying training is the only training programme with a positive effect in the medium-term. Two months after the beginning of the qualifying training, trainees are employed 1.92 months longer than non-trainees. Preparation training has no significant effect in the medium-term, and the lock-in effect of combined training persists in the medium-term, then becomes non-significant in the longer-term. General training does not seem to have a significant effect on employment. Moreover, the lock-in effect of long programmes such as combined training has a significant negative effect that endures over time.

No specific effect for women was found, as expected according to the literature, and a negative effect was found for youth, at odds with the findings of Crepon et al. (2012). Surprisingly, training most benefits seniors. These differences with previous papers show the importance of heterogeneity studies and the proper segmentation of active labour market policies, as was done here by considering only types of training which increase human capital. However, individuals with low education level are those who most benefit from training, contrary to people with high education levels. This result indicates that increasing human capital, especially in a specific profession, enhances an individual's employability by narrowing the gap between the skills expected by the employer and those possessed by the jobseeker. This is an argument in favour of using training to reduce frictions in the labour market in terms of human capital. However, if the training is very long, the opposite effect occurs as is shown here by the results on combined training. By significantly extending the unemployment period, even if this is for training purposes, it decreases the individual's employability. Therefore, both a minimum and a maximum training duration should be taken into consideration. Furthermore, for the estimation of *BGATEs*, we selected variables that are less explored in the current literature. The number of months on UI as a proxy for an individual's employment history indicates that individuals with fewer past employment periods experience a stronger positive effect of training. This effect is also observed for those further away from the labour market RSA recipients. Training thus appears to be a lever to assist those facing the greatest challenges in returning to the workforce. Individuals who were not born in France also derive more benefits from training than others, even though they are not targeted by the policy.

This is summarized by the cluster analysis, which indicates that to maximize the number of months in employment, training should target less qualified individuals, seniors, RSA recipients, and those not born in France. For policymakers aiming to enhance the effectiveness of training programmes in employment, it is essential to take into account the heterogeneity among jobseekers and training programmes. The current design of qualifying training appears to be particularly effective for individuals facing the greatest challenges in employment. However, targeting the youth not only fails to yield a specific positive impact on their employment but also falls short of providing effective training on general skills and engaging those extremely distant from the job market. It might be worthwhile to explore a training design similar to combined training but significantly shorter, incorporating a combination of general and specific training within a single training programme. Moreover, observing longer professional trajectories could lead to different results at long term, especially for the training programmes that last the longest.

CRediT authorship contribution statement

Héloïse Burlat: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Appendix. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.labeco.2024.102573.

References

Abbring, J.H., Van Den Berg, G.J., 2003. The nonparametric identification of treatment effects in duration models. Econometrica 71 (5), 1491–1517. http://dx.doi.org/10.1111/1468-0262.00456.

Ashenfelter, O., 1978. Estimating the effect of training programs on earnings. Rev. Econ. Stat. 60 (1), 47–57, Publisher: MIT Press.

Athey, S., Tibshirani, J., Wager, S., 2019. Generalized random forests. Ann. Statist. 47 (2), 1148–1178. http://dx.doi.org/10.1214/18-AOS1709.

Bearth, N., Lechner, M., 2024. Causal machine learning for moderation effects. http://dx.doi.org/10.48550/arXiv.2401.08290, arXiv preprint arXiv:2401.08290.

Becker, G.S., 1994. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, 3rd ed. University of Chicago Press, Chicago.

Biewen, M., Fitzenberger, B., Osikominu, A., Paul, M., 2014. The effectiveness of public-sponsored training revisited: The importance of data and methodological choices. J. Labor Econ. 32 (4), 837–897. http://dx.doi.org/10.1086/677233.

Caliendo, M., Mahlstedt, R., Mitnik, O.A., 2017. Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labor market policies. Lab. Econ. 46, 14–25. http://dx.doi.org/10.1016/j.labeco.2017.02.001.

Calmfors, L., 1995. Labour market policy and unemployment. Eur. Econ. Rev. 39 (3), 583–592. http://dx.doi.org/10.1016/0014-2921(94)00065-8.

Calmfors, L., Lang, H., 1995. Macroeconomic effects of active labour market programmes in a union wage- setting model. Econ. J. 105 (430), 601–619. http://dx.doi.org/10.2307/2235020.

Card, D., Kluve, J., Weber, A., 2018. What works? A meta analysis of recent active labor market program evaluations. J. Eur. Econom. Assoc. 16 (3), 894–931. http://dx.doi.org/10.1093/jeea/jvx028.

Cockx, B., Lechner, M., Bollens, J., 2023. Priority to unemployed immigrants? A causal machine learning evaluation of training in Belgium. Lab. Econ. 80, 102306. http://dx.doi.org/10.1016/j.labeco.2022.102306.

Crepon, B., Ferracci, M., Fougère, D., 2012. Training the unemployed in France: How does it affect unemployment duration and recurrence? Ann. Econ. Stat. (107/108), 175–199. http://dx.doi.org/10.2307/23646576.

Fleuret, A., 2006. L'orientation des chômeurs vers la formation de 2002 à 2004. Premières synthèses DARES (29.2), 1–6.

Grobon, S., Ramajo, I., Roucher, D., 2021. Quelle relation entre difficultés de recrutement et taux de chômage ? DARES.

Heckman, J.J., Ichimura, H., Todd, P.E., 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. Rev. Econ. Stud. 64 (4), 605–654. http://dx.doi.org/10.2307/2971733.

Heckman, J.J., Lalonde, R.J., Smith, J.A., 1999. Chapter 31 - the economics and econometrics of active labor market programs. In: Ashenfelter, O.C., Card, D. (Eds.), In: Handbook of Labor Economics, vol. 3, Elsevier, pp. 1865–2097. http://dx.doi.org/10.1016/S1573-4463(99)03012-6.

Imbens, G.W., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. J. Econ. Literat. 47 (1), 5–86.

Knaus, M.C., Lechner, M., Strittmatter, A., 2021. Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence. Econom. J. 24 (1), 134–161. http://dx.doi.org/10.1093/ectj/utaa014.

Layard, R., Nickell, S., 1986. Unemployment in Britain. Economica 53 (210), S121–S169. http://dx.doi.org/10.2307/2554377.

Lechner, M., Mareckova, J., 2022. Modified causal forest. http://dx.doi.org/10.48550/arXiv.2209.03744, arXiv preprint arXiv:2209.03744.

Lechner, M., Wunsch, C., 2013. Sensitivity of matching-based program evaluations to the availability of control variables. Lab. Econ. 21 (C), 111–121. http://dx.doi.org/10.1016/j.labeco.2013.01.004.

Lechner, M., Wunsch, C., Miquel, R., 2011. Long-run effects of public sector sponsored training in west Germany. J. Eur. Econom. Assoc. 9 (4), 742–784. http://dx.doi.org/10.1111/j.1542-4774.2011.01029.x.

Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and non-randomized studies. J. Educ. Psychol. 66, 688–701. http://dx.doi.org/10.1037/h0037350.

Sianesi, B., 2004. An evaluation of the Swedish system of active labor market programs in the 1990s. Rev. Econ. Stat. 86 (1), 133–155. http://dx.doi.org/10.1162/003465304323023723.

Wager, S., Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. J. Amer. Statist. Assoc. 113 (523), 1228–1242.

Zhou, Z., Athey, S., Wager, S., 2023. Offline multi-action policy learning: Generalization and optimization. Oper. Res. 71 (1), 148–183. http://dx.doi.org/10.1287/opre.2022.2271.